

UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE CIÊNCIAS AGRÁRIAS  
DEPARTAMENTO DE FITOTECNIA  
PROGRAMA DE PÓS GRADUAÇÃO EM RECURSOS GENÉTICOS  
VEGETAIS

GUSTAVO HENRIQUE FERRERO KLABUNDE

**ANÁLISE TRANSCRIPTÔMICA NAS CONÍFERAS  
BRASILEIRAS *Araucaria angustifolia* (Bert.) O. Kuntze  
(Araucariaceae) e *Podocarpus lambertii* Klotzsch ex Eichler  
(Podocarpaceae): ANOTAÇÃO FUNCIONAL E MINERAÇÃO DE  
MARCADORES MOLECULARES SSRs E SNPs**

Tese submetida ao Programa de Pós-Graduação em Recursos Genéticos Vegetais da Universidade Federal de Santa Catarina, linha de pesquisa **Genética e Melhoramento de Plantas**, como requisito parcial para a obtenção do título de Doutor em Ciências.

Orientador: Dr. Rubens Onofre Nodari

Florianópolis  
2016

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Klabunde, Gustavo Henrique Ferrero

Análise transcriptômica nas coníferas brasileiras  
*Araucaria angustifolia* (Bert.) O. Kuntze (Araucariaceae) e  
*Podocarpus lambertii* (Podocarpaceae) Klotzsch ex Eichler:  
Anotação funcional e mineração de marcadores moleculares  
SSRs e SNPs / Gustavo Henrique Ferrero Klabunde ;  
orientador, Rubens Onofre Nodari - Florianópolis, SC, 2016.  
142 p.

Tese (doutorado) - Universidade Federal de Santa  
Catarina, Centro de Ciências Agrárias. Programa de Pós  
Graduação em Recursos Genéticos Vegetais.

Inclui referências

1. Recursos Genéticos Vegetais. 2. Melhoramento  
genético vegetal. 3. Análise transcriptômica. 4. *Araucaria  
angustifolia*. 5. *Podocarpus lambertii*. I. Nodari, Rubens  
Onofre. II. Universidade Federal de Santa Catarina.  
Programa de Pós-Graduação em Recursos Genéticos Vegetais.  
III. Título.

**Análise transcriptômica nas coníferas brasileiras  
*Araucaria angustifolia* (Bert.) O. Kuntze  
(Araucariaceae) e *Podocarpus lambertii* Klotzsch ex  
Eichler (Podocarpaceae): anotação funcional e  
mineração de marcadores moleculares SSRs E SNPs**


por


**Gustavo Henrique Ferrero Klabunde**

Tese julgada e aprovada em 25/02/2016, em sua forma final, pelo Orientador e membros da Banca Examinadora, para obtenção do título de Doutor em Ciências. Área de Concentração Recursos Genéticos Vegetais, no Programa de Pós-Graduação em Recursos Genéticos Vegetais, CCA/UFSC.

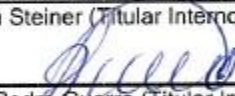
Banca Examinadora:

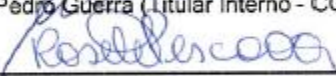
  
\_\_\_\_\_  
Prof. Dr. Rubens Onoffe Nodari (Presidente - CCA/UFSC)

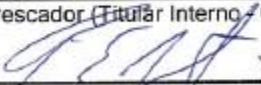
  
\_\_\_\_\_  
Prof. Dr. Valdir Marcos Stefanon (Titular Externo-UNIPAMPA/RS)

  
\_\_\_\_\_  
Profª. Drª. Roseli Wassermann (Titular Externo - UFPR)

  
\_\_\_\_\_  
Profª. Drª. Neusa Steiner (Titular Interno - CCB/UFSC)

  
\_\_\_\_\_  
Prof. Dr. Miguel Pedro Guerra (Titular Interno - CCA/UFSC)

  
\_\_\_\_\_  
Profª. Drª. Rosete Pescador (Titular Interno - CCA/UFSC)

  
\_\_\_\_\_  
Prof. Dr. Paulo Emilio Lovato (Coordenador do Programa)

Florianópolis, fevereiro de 2016

## IDENTIFICAÇÃO

**Título:** ANÁLISE TRANSCRIPTÔMICA NAS CONÍFERAS BRASILEIRAS *Araucaria angustifolia* (Bert.) O. Kuntze (Araucariaceae) e *Podocarpus lambertii* Klotzsch ex Eichler (Podocarpaceae): ANOTAÇÃO FUNCIONAL E MINERAÇÃO DE MARCADORES MOLECULARES SSRs E SNPs

**Período de Execução:** Março de 2012 a Fevereiro de 2016.

**Instituição Executora:** Universidade Federal de Santa Catarina – UFSC.

**Responsável:**

Gustavo Henrique Ferrero Klabunde, Engº Agrônomo (UFSC - 2010), Mestre em Recursos Genéticos Vegetais (PPGRGV - UFSC - 2012).

**Orientador:**

Prof. Dr. Rubens Onofre Nodari, Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal, Departamento de Fitotecnia, Centro de Ciências Agrárias, Universidade Federal de Santa Catarina.

**Colaboradores:**

Prof. Dr. Miguel Pedro Guerra, Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal, Departamento de Fitotecnia, Centro de Ciências Agrárias, Universidade Federal de Santa Catarina.

Prof. Dr. Fábio, de Oliveira Pedrosa, Núcleo de Fixação de Nitrogênio (NFN), Departamento de Bioquímica, Centro Politécnico, Universidade Federal do Paraná.

Prof. Dr. Emanuel Maltempi, de Souza, Núcleo de Fixação de Nitrogênio (NFN), Departamento de Bioquímica, Centro Politécnico, Universidade Federal do Paraná.

Dra. Michelle Zibetti Tadra Sfeir, Pós-doutoranda, Núcleo de Fixação de Nitrogênio (NFN), Departamento de Bioquímica, Centro Politécnico, Universidade Federal do Paraná.

Prof. Dr. Marcelo Rogalski, Departamento de Biologia Vegetal, Universidade Federal de Viçosa.

**Instituição financiadora:**

**Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina**

**Projeto FAPESC, termo de outorga: 14.848/2011-2**

**"Análises genômicas e transcriptômicas nas coníferas brasileiras *Araucaria angustifolia*, *Podocarpus sellowii*, *Podocarpus lambertii* e *Retrophyllum piresii* visando uso, conservação, estudos evolutivos, moleculares e biotecnológicos"**

## AGRADECIMENTOS

À Universidade Federal de Santa Catarina pelos meus 11 anos de formação acadêmica, pessoal e profissional.

Ao Programa de Pós-Graduação em Recursos Genéticos Vegetais, professores e Bernardete Ribas.

Ao Professor Rubens Onofre Nodari pelos anos de orientação, paciência, liberdade e confiança.

À minha noiva Tatiane Bratti pelo amor, compreensão, parceria e paciência! Amo você!

À toda minha família, em especial a meus pais, Marlene e Ronaldo, que abdicaram de inúmeras coisas para proporcionar uma educação de qualidade para mim e para minha irmã Elisa.

Aos amigos da *Core Collection*: Denise Olkoski, Lilian Machado, Vinícius Vilperte, Rafael Benevenuto, Tiago Montagna, Raissa Guse, Daniela Werner, Daniel Holderbaum. “Aqui estão os loucos. Os desajustados. Os rebeldes. Os criadores de caso. Os pinos redondos nos buracos quadrados”.

Aos demais colegas que passaram pelo LFDGV, NPFT e RGV nesses meus nove de laboratório. São tantas as pessoas que tive a oportunidade de conviver, aprender e ensinar. Com certeza, colegas, vocês contribuíram e muito para a minha formação acadêmica e pessoal.

Aos amigos sempre presentes Júlio Graeff Erpen, Diego Jacob Kurtz, Alan Fabrício Malinski, Ricardo Cipriani, Ruã Darós, Rafael Moreira de Fúcio, Dr. João Henrique Cardoso Costa, Lino Jorge Moysés, Leonardo Zimmermann e Viane Kist.

Aos demais amigos *nerds* do *Trading Card Game* pelas horas de descontração.

Ao Núcleo de Fixação de Nitrogênio da UFPR, em especial à Dra. Michelle Zibetti Tadra-Sfeir pelo tempo e esforço na construção das bibliotecas e realização das corridas de sequenciamento.

Aos doutorandos Alison Bernardi e Newton Costa por cederem gentilmente suas imagens para a construção das pranchas.

Aos relatores e membros da banca avaliadora deste trabalho de tese.

Ao CNPq pela concessão da bolsa de estudos e à FAPESC pelos recursos financeiros do projeto.

A todos que direta ou indiretamente me ajudaram no desenvolvimento desta tese, muito obrigado!

Tempo virá em que uma pesquisa diligente e contínua esclarecerá aspectos que agora permanecem escondidos. O espaço de tempo de uma vida, mesmo se inteiramente devotada ao estudo do céu, não seria suficiente para investigar um objetivo tão vasto... este conhecimento será conseguido somente através de gerações sucessivas. Tempo virá em que os nossos descendentes ficarão admirados de que não soubéssemos particularidades tão óbvias a eles... Muitas descobertas estão reservadas para os que virão, quando a lembrança de nós estará apagada. O nosso universo será um assunto sem importância, a menos que haja alguma coisa nele a ser investigada a cada geração... A natureza não revela seus mistérios de uma só vez

. — Sêneca, Problemas Naturais  
Livro 7, século I

## RESUMO

*Araucaria angustifolia* e *Podocarpus lambertii* são coníferas brasileiras criticamente ameaçadas de extinção. Várias lacunas em diversas áreas de pesquisa como a fisiologia do desenvolvimento e biologia reprodutiva são obstáculos para o conhecimento, uso sustentável e proposição de medidas de conservação destes recursos genéticos. Entretanto, poucos estudos genômicos e transcriptômicos foram realizados nestas espécies. Este trabalho teve como principal objetivo o sequenciamento e a caracterização de transcriptomas de amostras de *A. angustifolia* (folhas, pólen e hastes) e *P. lambertii* (folhas). Um total de 341.278.253 (45.6 Gb) de leituras brutas foram obtidas da plataforma de sequenciamento de segunda geração Ion Proton, sendo utilizadas para a montagem referenciada e *de novo* dos transcriptomas, gerando 60.043 sequências contíguas. Em *A. angustifolia*, foram anotados, via *gene ontology* (GO) 4.354 unigenes distintos a partir do transcriptoma de folhas, 3.643 unigenes do transcriptoma de hastes e 9.354 unigenes do transcriptoma de pólen. Em *P. lambertii* foram identificados 4.649 unigenes no transcriptoma de folhas. A identificação e caracterização *in silico* e desenho de iniciadores de 56 marcadores moleculares microssatélites (SSR) e 259 marcadores de polimorfismo de base única (SNP) para as duas espécies foi realizada a partir dos dados disponíveis. Os motivos SSRs mais abundantes foram mononucleotídeos (39,7 %), seguidos por tetranucleotídeos (22,22 %) e trinucleotídeos (20,63 %). A relação entre o número de transversões e transições (Ti/Tv) foi de 2,15 para *A. angustifolia* e de 1,05 para *P. lambertii*. Os resultados obtidos para os marcadores SNPs apontam diferentes padrões na substituição de nucleotídeos entre as espécies, no entanto, os padrões dos marcadores SSRs foram muito semelhantes entre *Araucaria* e *Podocarpus*. Este trabalho descreve os primeiros transcriptomas gerados de amostras coletadas a campo para ambas espécies e os resultados gerados fornecem subsídios que auxiliarão no entendimento sobre funções específicas de estruturas, evolução e servirão de base para futuros estudos funcionais e populacionais.

**Palavras-chave:** Bioinformática; Detecção de variação; Expressão gênica de tecido específico; Melhoramento genético; Marcadores genéticos; Montagem *de novo*; Pinheiro-Brasileiro; Pinho-bravo; RNA; RNA-Seq; Transcrição.

## ABSTRACT

*Araucaria angustifolia* and *Podocarpus lambertii* are critically endangered Brazilian conifers. Several gaps in many research areas like developmental physiology and reproductive biology are obstacles to its knowledge, breeding, sustainable use of these genetic resources and proposition of conservation measures. However, few genomic and transcriptomic studies have been conducted on these species. The main objective of this work was the sequencing and characterization of transcriptomes from *A. angustifolia* (leaf, pollen and stem) and *P. lambertii* (leaf). A total of 341.278.253 (45.6 Gb) of raw reads were obtained from the Ion Proton second generation DNA sequencing platform and being used for reference based and de novo transcriptome assembly, generating 60.043 contigs. In *A. angustifolia* were annotated with the use of gene ontology (GO) 4.354 distinct unigenes from leaves transcriptome, 3.643 unigenes from stem and 9.354 unigenes from the pollen transcriptome. In *P. lambertii* were identified 4.649 unigenes from leaves transcriptome. In addition, sequence data allowed the in silico identification, characterization and primer design of 56 simple sequence repeats (SSR) and 259 single nucleotide polymorphisms (SNP) molecular markers for both species were performed from the available data. The most abundant SSR repeat motif was mononucleotide (39,7 %), followed by tetranucleotide (22,22 %) and trinucleotide (20,63 %). The ratio between transitions and transversions was 2,15 for *A. angustifolia* and 1,05 for *P. lambertii*. SNP results show distinct patterns for nucleotide substitutions between species, however, SSRs patterns were very similar between *Araucaria* and *Podocarpus*. This work therefore describes the first transcriptomes from field samples for both species and the results provide valuable resources for plant breeding, will further contribute on the understanding about tissue specific functions, conifers evolution and will be the base for future functional and populational studies.

**Keywords:** Bioinformatics; Brave pine; Brazilian pine; De novo assembly; Genetic markers; Plant breeding; RNA; RNA-Seq; Tissue-specific gene expression; Transcription; Variant detection.



## LISTA DE FIGURAS

- Figura 1.** Árvore filogenética de espécies representantes da Ordem Coniferales, baseada nas sequências dos genes organelares *matK* e *rbcL* ..... 25
- Figura 2. 1:** Grão de pólen de *A. angustifolia* visualizado em estereomicroscópio ampliado 100 x. Escala: 25 µm. **2:** Grão de pólen visualizado por meio de SEM com ampliação de 400 x. Escala: 25 µm. **3:** Detalhe da superfície do grão de pólen (exina) visualizado por meio de SEM com ampliação de 8.000 x. Escala: 2 µm ..... 29
- Figura 3. A:** Área de estudo, Coxilha Rica, Lages - SC. **B:** Pinha madura de *A. angustifolia*. **C:** Cones masculinos (androstróbilos) de *A. angustifolia* ..... 30
- Figura 4.** Áreas de ocorrência natural “histórica” e atual de *A. angustifolia* (Bert.) O. Kuntze (em verde escuro), dentro do domínio da Mata Atlântica (área verde clara) ..... 31
- Figura 5. A:** Indivíduo adulto de *Podocarpus lambertii*. **B:** Regenerantes de *P. lambertii* identificados com etiqueta metálica. **C:** Ramo de planta masculina contendo cones (ginostróbilos). **D:** Ramo de planta feminina contendo sementes imaturas e pedúnculo carnoso (epimácio) ..... 33
- Figura 6.** Áreas de ocorrência atual de *P. lambertii* Klotzsch ex Eichler (em vermelho) ..... 34
- Figura 7.** Regenerantes de *A. angustifolia* (esquerda) e *P. lambertii* (direita), na área de estudo. Coxilha Rica, Lages - SC..... 35
- Figura 8.** Representação esquemática do fluxo de protocolos para a obtenção dos transcriptomas, anotação funcional e mineração de marcadores moleculares em *A. angustifolia* e *P. lambertii* ..... 49
- Figura 9.** Imagem de satélite da área de coleta. Mata ciliar nas margens do rio Pelotinhas. Propriedade do Dr. José Antônio Ribas Ribeiro, localidade da Coxilha Rica / Lages / SC ..... 50

**Figura 10.** Folhas de *P. lambertii* (A); Haste de *A. angustifolia* (B); Acículas de *A. angustifolia* (C); Estróbilo masculino de *A. angustifolia* (esquerda), pólen (centro) e tubo de coleta (15 ml) (direita) (D) ....53

**Figura 11.** Eletroferogramas das amostras pré-depleção ribossomal utilizando o kit RNA 6000 Nano. Picos não marcados na base correspondem a rRNA adicionais. Amostras 1 a 3: RNA total isolado de folhas de *P. lambertii*; Amostras 4 a 6: RNA total isolado de folhas de *A. angustifolia*; Amostras 7 a 9: RNA total isolado de hastes de *A. angustifolia*; Amostras 10 a 12: RNA total isolado de pólen de *A. angustifolia* .....69

**Figura 12.** Eletroferogramas das amostras pós-depleção ribossomal utilizando o kit RNA 6000 Nano. Amostras 1 a 3: RNA total isolado de folhas de *P. lambertii*; Amostras 4 a 6: RNA total isolado de folhas de *A. angustifolia*; Amostras 7 a 9: RNA total isolado de hastes de *A. angustifolia*; Amostras 11 e 12: RNA total isolado de pólen de *A. angustifolia* .....71

**Figura 13.** *Box plot* Mapeamento contra o genoma completo de *Arabidopsis thaliana*. As cores representam amostras distintas; vermelho - folhas de *P. lambertii*; verde - folhas de *A. angustifolia*; azul - hastes de *A. angustifolia*; amarelo - pólen de *A. angustifolia* 75

**Figura 14.** *Heat map* mostrando o padrão de expressão das amostras e suas replicatas biológicas quando do mapeamento referenciado contra *Arabidopsis thaliana*. .....76

**Figura 15.** Análise dos componentes principais (PCA) quando da utilização de *Arabidopsis thaliana* como referência. A cor em vermelho representa as replicatas das amostras de folhas de *P. lambertii*; verde representa folhas de *A. angustifolia*; azul representa hastes de *A. angustifolia*; amarelo representa pólen de *A. angustifolia*. .....78

**Figura 16.** *Volcano plot* das amostras comparadas par-a-par e mapeadas contra o genoma completo de *Arabidopsis thaliana*. FAA: Folhas de *A. angustifolia*; HAA: Hastes de *A. angustifolia*; PAA: Pólen de *A. angustifolia*; FPL: Folhas de *P. lambertii*. .....79

<b>Figura 17.</b> Distribuição dos top-hits por espécie na análise de BLASTX para os transcriptomas de folhas, hastes e pólen de <i>A. angustifolia</i> e folhas de <i>P. lambertii</i> .....	<b>81</b>
<b>Figura 18.</b> Distribuições das anotações funcionais com utilização dos termos GO no transcriptoma de folhas de <i>A. angustifolia</i> . As barras do histograma representam o número de unigenes identificados nas mais frequentes categorias funcionais de processos biológicos (BP), funções moleculares (MF) e componentes celulares (CC) .....	<b>84</b>
<b>Figura 19.</b> Distribuições das anotações funcionais com utilização dos termos GO no transcriptoma de hastes de <i>A. angustifolia</i> . As barras do histograma representam o número de unigenes identificados nas mais frequentes categorias funcionais de processos biológicos (BP), funções moleculares (MF) e componentes celulares (CC) .....	<b>85</b>
<b>Figura 20.</b> Distribuições das anotações funcionais com utilização dos termos GO no transcriptoma de pólen de <i>A. angustifolia</i> . As barras do histograma representam o número de unigenes identificados nas mais frequentes categorias funcionais de processos biológicos (BP), funções moleculares (MF) e componentes celulares (CC) .....	<b>86</b>
<b>Figura 21.</b> Distribuições das anotações funcionais com utilização dos termos GO no transcriptoma de folhas de <i>P. lambertii</i> . As barras do histograma representam o número de unigenes identificados nas mais frequentes categorias funcionais de processos biológicos (BP), funções moleculares (MF) e componentes celulares (CC) .....	<b>87</b>
<b>Figura 22.</b> Distribuição dos termos GO (level 2) mais abundantes nos transcriptomas de folhas (azul), hastes (vermelho) e pólen (verde) de <i>A. angustifolia</i> . .....	<b>88</b>
<b>Figura 23.</b> Distribuição dos termos GO (level 2) mais abundantes nos transcriptomas de folhas de <i>A. angustifolia</i> (vermelho) e <i>P. lambertii</i> (azul) .....	<b>88</b>

## LISTA DE TABELAS

<b>Tabela 1.</b> Coordenadas geográficas das coletas de amostras para RNA-Seq .....	<b>51</b>
<b>Tabela 2.</b> Indicadores de integridade, pureza e quantidade das amostras de RNA total pré-depleção de rRNAs .....	<b>68</b>
<b>Tabela 3.</b> Quantificação de mRNA pós-depleção ribossomal .....	<b>70</b>
<b>Tabela 4.</b> Resumo das leituras, bases e sequências contíguas obtidas por meio de RNA-Seq das amostras analisadas .....	<b>73</b>
<b>Tabela 5.</b> Proporção de nucleotídeos (%) e conteúdo CG% após o controle de qualidade das bibliotecas sequenciadas .....	<b>74</b>
<b>Tabela 6.</b> Resumo dos SNPs putativos identificados em <i>A. angustifolia</i> e <i>P. lambertii</i> .....	<b>89</b>
<b>Tabela 7.</b> Distribuição dos SNPs putativos em <i>A. angustifolia</i> e <i>P. lambertii</i> por tipo de mutação .....	<b>90</b>
<b>Tabela 8.</b> Resumo dos SSRs putativos identificados em <i>A. angustifolia</i> e <i>P. lambertii</i> .....	<b>91</b>
<b>Tabela 9.</b> Distribuição dos motivos dos SSRs putativos identificados em <i>A. angustifolia</i> e <i>P. lambertii</i> de acordo com o número de repetições...	<b>92</b>

## LISTA DE ABREVIATURAS, SIGLAS E EXPRESSÕES

- .fastq – Formato baseado em texto para representar sequências de nucleotídeos e peptídeos
- mer - Tamanho de oligonucleotídeo, sempre precedido do número de bases
- °C - Graus Celcius
- 260/230 – Relação entre as absorbâncias 260 nm e 230 nm, que é usada para estimar a pureza de RNA e DNA, indicando contaminação por compostos fenólicos, carboidratos e tiocianato de guanidina
- 260/280 – Relação entre as absorbâncias 260 nm e 280 nm, que é usada para estimar a pureza de RNA e DNA, indicando contaminação por proteínas, compostos fenólicos e outras substâncias que absorvem fortemente a 280 nm
- 28S/18S – Relação entre as quantidades dos rRNAs 28S e 18S, que é usada para estimar a integridade do mRNA
- 2C - Quantidade de DNA, em picogramas, contida em um núcleo celular diploide
- 2n - Número cromossômico em célula diploide
- 454 - Plataforma de sequenciamento de DNA de segunda geração, baseada em pirosequenciamento
- µg – Micrograma
- µl – Microlitro
- µm - Micrômetro
- µM – Micromolar
- A – Adenina (nucleobase)
- Bert. - Giuseppe Bertolini, botânico italiano que descreveu, em 1819, o pinheiro-brasileiro como *Columbea angustifolia*
- BLAST – *Basic Local Alignment Search Tool* – Ferramenta para realização de alinhamento local
- BLASTX – Algoritmo de alinhamento de sequências traduzidas de nucleotídeos contra uma base de dados de proteínas
- bp - Base pairs - Pares de base
- BP – Biological process – Categoria GO de processos biológicos
- C – Citosina (nucleobase)
- CC- Cellular component – Categoria GO de componentes celulares
- cDNA - DNA complementar, obtido pela polimerização do mRNA pela enzima transcriptase reversa
- CesA - Família gênica da enzima celulose sintase
- Cfb – Clima mesotérmico subtropical úmido sem estação seca
- CNCFlora – Centro Nacional de Conservação da Flora

CONSEMA – Conselho Estadual do Meio Ambiente  
contig - Sequência contígua, formada pela sobreposição de leituras em determinado alinhamento  
cpDNA - DNA plastidial  
cpSSR - Marcador SSR localizado no genoma plastidial  
cv – Cultivar  
CTAB – Brometo de cetiltrimetilamônio  
dATP - desoxiadenosina trifosfatada  
dCTP - desoxicitosina trifosfatada  
dGTP- desoxiguanosina trifosfatada  
dTTP - desoxitimidina trifosfatada  
DEPC – Dicarbonato de dietila  
*de novo* - Método de alinhamento e montagem de genoma ou transcriptoma sem a utilização de genoma de referência  
DNA – Ácido desoxirribonucleico  
DNase - Nuclease que catalisa a degradação do ácido desoxirribonucleico  
dNTP – Desoxirribonucleotídeos trifosfatados  
dsRNA – Double stranded RNA – RNA dupla fita  
EC – *Enzyme Commission numbers* – Sistema de nomenclatura e classificação de enzimas  
EDTA – Ácido Etilenodiaminotetracético  
e-Value – Valor de confiabilidade de alinhamento de sequências, sendo 100% confiável quando do valor 0.  
EST – *Expressed Sequence Tags* – Marcador de sequência expressa  
EST-SSR - SSRs derivados de sequências expressas  
FASTA – Formato baseado em texto para representar sequências de nucleotídeos e peptídeos  
FASTQ - Formato FASTA, com respectivos valores de qualidade de cada base sequenciada (.fastq ou .fq)  
FL-cDNA - Full length cDNA - cDNAs possuindo todas as estruturas de um mRNA, desde sítio CAP até cauda poli-A  
FOM - Floresta Ombrófila Mista  
G – Guanina (nucleobase)  
g – Gramas  
gap – *Indel* presente em determinado alinhamento de sequências nucleotídicas ou proteicas  
GC - Porcentagem das nucleobases Guanina e Citosina em determinada sequência ou grupos de sequências de DNA/RNA  
GB – Giga Bases  
GO - *Gene Ontology* – Ontologia do gene

GOSlims – Versão resumida e representativa dos GO's  
*in silico* - Caracterização, modelagem computacional  
indel - Inserções ou deleções de bases nitrogenadas em determinada sequência de DNA/RNA  
intron - INTRagenic regiON - região intra-gênica, excisada durante a etapa de *splicing*  
IUCN - *International Union for Conservation of Nature*  
ISSR - Marcador molecular baseado entre sequências repetidas simples  
JCR - *Journal Citation Reports*  
kb – kpb – Kilo Bases = 1.000 pb  
kd - Kilo Daltons  
l - Litro  
LiCl – Cloreto de lítio  
M - Molar  
MA - Microarranjo - *Microarray*  
MAF - *Minimum Allele frequency* – Frequência mínima de alelo  
MAS - *Marker Assisted Selection* - Seleção assistida por marcadores moleculares  
matK – Gene cloroplastidial Maturase K  
Mb - Mpb - Mega Bases = 1.000.000 pb  
MDB - *Membrane Dessalting Buffer* - Tampão de dessalinização de membrana de sílica  
MF – *Molecular function* – Categoria GO de funções moleculares  
mg – Miligrama  
miRNA - MicroRNA  
ml - Mililitro  
mm – Milímetro  
MMA – Ministério do Meio Ambiente  
MOPS - Ácido 3-(N-morpholino)propanosulfônico  
mtDNA - DNA mitocondrial  
mRNA - RNA mensageiro  
N50 - Medida de continuidade em uma montagem de leituras. Por exemplo, um N25 de 5 Mb significa que pelo menos um quarto das bases estão em uma sequência contígua de pelo menos 5 Mb  
NaCl – Cloreto de sódio  
NCBI – *National Center for Biotechnology Information* - Centro Nacional para Informações Biotecnológicas, EUA  
ncRNA - *Non-coding RNAs* - RNAs não codificantes  
NGS - *Next Generation Sequencing* – Sequenciamento de DNA de Nova Geração  
ng/μl – Nanogramas/microlitro

nm - Nanômetros

nM - Nanomolar

O. Kuntze – Otto Kuntze, botânico alemão que descreveu, em 1898, a araucária como *Araucaria angustifolia*

ORF - *Open Reading Frame* - Fase aberta de leitura

LFDGV – Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal

P10 - Micropipeta para manipulação de volumes entre 2  $\mu$ l e 10  $\mu$ l

P100 - Micropipeta para manipulação de volumes entre 10  $\mu$ l e 100  $\mu$ l

P200 - Micropipeta para manipulação de volumes entre 20  $\mu$ l e 200  $\mu$ l

P1000 - Micropipeta para manipulação de volumes entre 200  $\mu$ l e 1000  $\mu$ l

pb – Pares de base

PCA – *Principal Component Analysis* - Análise dos componentes principais

PCR – *Polymerase Chain Reaction* - Reação em cadeia da polimerase

PGM - *Personal Genome Machine* - Sequenciador de DNA de segunda geração, da família Ion Torrent

pg – Picograma. Equivalente a 10-12 grama. Representa cerca de 109 pares de base

pg/ $\mu$ l - Picogramas/microlitro

pH – Potencial hidrogeniônico

Phred score – Valores de qualidade associados a acurácia de nucleotídeos em sequência de DNA

PI - *Chip* de sequenciamento das plataformas

IonTorrent<sup>TM</sup>/IonProton<sup>TM</sup>

PIC - *Polymorphism Information Content* - Conteúdo de informação polimórfica

pM - Picomolar

PVP – Polivinilpirrolidona

poli-A - Estrutura poliadenilada na extremidade 3' de moléculas de mRNA

Q20 - Acurácia de 99% de uma base sequenciada

QTL - *Quantitative Trait Loci* - Locus de características quantitativas

q.s.p. - Quantidade suficiente para

RA3 - Tampão de lavagem, baseado em etanol

rbcL – Gene plastidial que codifica para a Sub-unidade grande da enzima Enzima ribulose-1,5-bisfostato carboxilase oxigenase (RuBisCo)

RAM - *Random Access Memory* - Memória de acesso aleatório

RAP - Tampão de lise celular, baseado em cloridato de guanidina



RAW2 - Tampão de lavagem, baseado em cloridato de guanidina e etanol

reads - Leituras de sequenciamento com tamanho de sequências dependendo da plataforma e química de NGS

RIN - RNA *Integrity Number* - Valor entre 0 e 10 que representa a integridade de determinada amostra de RNA

RNA-40 - Constante utilizada para o cálculo de concentração de RNA

RNA – Ácido Ribonucleico

RNase – Nucleases que catalisam a degradação do ácido ribonucleico

RNase III – Nuclease que cliva dsRNA em fragmentos de dsRNA de 12 a 15 pb com 2 a 3 bases de *overhang* 3'

RNA-Seq - Sequenciamento de RNA, técnicas baseadas em NGS para a obtenção de transcriptomas completos

RPM – Rotações por min

RPKM – *Reads Per Kilobase per Million mapped reads* – Método de quantificação da expressão gênica

rRNA - RNA ribossômico ou ribossomal

RT - Reação de transcrição reversa mediada pela enzima Transcriptase reversa

RuBisCo - Enzima ribulose-1,5-bisfosfato carboxilase oxigenase

S - Sub-unidades do rRNA, composto pelo rRNA citoplasmático, mitocondrial e cloroplastidial

s – Segundos

S/I – Relação entre o número de substituições (SNPs) e *indels*

SDS – Dodecil sulfato de sódio

SEM - *Scanning Electron Microscope* - Microscópio eletrônico de varredura

SNP – *Single Nucleotide Polymorphism* - Polimorfismo de base única

SNP/kb – Frequência de SNPs por kilobases

SNV – *Single Nucleotide Variation* – Variação de base única

SOLiD - *Sequencing by oligonucleotide ligation and detection* - Sequenciamento por ligação e detecção de nucleotídeos

SRA – *Sequence Read Archive* – Base de dados do NCBI para depósito de sequências brutas provenientes de NGS

SSR - Marcador molecular baseado em sequências repetidas simples (Microsatélite)

SSTE – Tampão composto por cloreto de sódio, dodecil sulfato de sódio, tris e EDTA

STR - *Short Tandem Repeat* - Repetições curtas em tandem. Outra denominação das marcas SSR

T – Timina (nucleobase)

TE - Tampão contendo Tris-base e EDTA

Ti/Tv – “*Transitions / Transversions*” - Relação entre o número de transições e transversões de bases nitrogenadas em SNPs

tRNA – RNA transportador

U – Uracil (nucleobase)

U – Unidade enzimática

UTR - *UnTranslated Region* - região não traduzida, localizada na extremidade 5' ou 3' dos mRNA

VNTR - *Variable Number of Tandem Repeats* - Repetições em tandem de número variado

X - Cobertura de sequenciamento, número de vezes em que a mesma base nitrogenada é sequenciada

xg – Força centrífuga relativa

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	23
<b>2 REVISÃO DE LITERATURA</b> .....	25
<b>2.1 Coníferas</b> .....	25
<b>2.2. <i>Araucaria angustifolia</i> (Bert.) O. Kuntze</b> .....	27
<b>2.3 Podocarpus lambertii Klotzsch ex Eichler</b> .....	32
<b>2.4 Sequenciamento de Ácidos Nucleicos</b> .....	36
<b>2.5 Análise transcriptômica (RNA-Seq)</b> .....	37
<b>2.6 Mineração e desenvolvimento de marcadores     moleculares SSRs e SNPs</b> .....	43
<b>3 QUESTÕES ORIENTADORAS</b> .....	47
<b>4 OBJETIVOS</b> .....	48
<b>4.1 Objetivos gerais</b> .....	48
<b>4.2 Objetivos específicos</b> .....	48
<b>5 MATERIAL E MÉTODOS</b> .....	49
<b>5.1 Área de estudo e coleta das amostras</b> .....	50
<b>5.2 Amostras</b> .....	52
<b>5.3 Isolamento de RNA total</b> .....	54
<b>5.3.1 CTAB e Cloreto de lítio</b> .....	54
<b>5.3.2 Kit Nucleospin RNA Plant®</b> .....	56
<b>5.4 Quantificação e verificação da qualidade do RNA</b> ..	56
<b>5.5 Precipitação de RNA total</b> .....	57
<b>5.6 Depleção de RNA</b> .....	58
<b>5.7 Construção das bibliotecas de cDNA</b> .....	58
<b>5.7.1 Fragmentação do RNA</b> .....	59
<b>5.7.2 Purificação do RNA fragmentado</b> .....	59
<b>5.7.3 Avaliação da qualidade e distribuição de tamanho</b>	

do RNA fragmentado.....	60
5.7.4 Hibridização e ligação do RNA .....	60
5.7.5 Transcrição reversa (RT) .....	60
5.7.6 Purificação do cDNA .....	61
5.7.7 Amplificação e purificação do cDNA .....	62
5.7.8 Avaliação da quantidade e distribuição de tamanho do cDNA amplificado.....	63
5.8 Sequenciamento de cDNA por <i>chips</i> semi-condutores .....	63
5.9 <i>Trimming</i> e montagem <i>de novo</i> .....	64
5.10 Mapeamento e análise referenciada ( <i>Reference based</i> ).....	64
5.11 Análise funcional e caracterização dos transcriptomas .....	65
5.12 Identificação de SNPs e SSRs putativos.....	66
6 RESULTADOS .....	68
6.1 Qualidade e quantidade de RNA total .....	68
6.2 Depleção de RNA .....	70
6.3 Sequenciamento, <i>trimming</i> e montagem dos transcriptomas .....	72
6.4 Análise funcional e caracterização dos transcriptomas .....	74
6.5 Identificação de SNPs e SSRs putativos.....	89
7 DISCUSSÃO.....	93
7.1 Isolamento, qualidade e quantidade de RNA total ...	93
7.2 Sequenciamento das bibliotecas de cDNA .....	94
7.3 Análise funcional e caracterização dos transcriptomas .....	95

7.4 Identificação de SNPs e SSRs putativos .....	97
8 CONCLUSÕES .....	102
9 PERSPECTIVAS .....	103
10 BIBLIOGRAFIA .....	104
APÊNDICE A – Demais soluções utilizadas na rotina laboratorial envolvendo análise de RNA. ....	126
APÊNDICE B – Composição dos tampões e soluções para o isolamento de RNA. ....	126
APÊNDICE C – Composição dos tampões e soluções para eletroforese nativa e desnaturante de RNA. ....	127
APÊNDICE D – Distribuição das classes enzimáticas identificadas nos transcriptomas sequenciados. ....	129
APÊNDICE E – <i>Contigs</i> , posição, base referência e base polimórfica dos SNPs putativos identificados em <i>P. lambertii</i> . ....	130
APÊNDICE F - <i>Contigs</i> , posição, base referência e base polimórfica dos SNPs putativos identificados em <i>A. angustifolia</i> . ....	131
APÊNDICE G – Listagem de <i>primers</i> EST-SSR obtidos para <i>A. angustifolia</i> , sequências, temperaturas de hibridização, tamanho de <i>amplicon</i> , motivo SSR e início e fim dos motivos nas sequências dos <i>contigs</i> . ....	134
APÊNDICE H - Listagem de <i>primers</i> EST-SSR obtidos para <i>P. lambertii</i> , sequências, temperaturas de hibridização, tamanho de <i>amplicon</i> , motivo SSR e início e fim dos motivos nas sequências dos <i>contigs</i> . ....	136
APÊNDICE I – Resultados do BLASTX dos <i>contigs</i> de <i>P. lambertii</i> contendo motivos SSRs, <i>E-value</i> , nº de acesso, espécie referente, proteína e origem dos transcritos. ....	137
APÊNDICE J – Resultados do BLASTX dos <i>contigs</i> de <i>A. angustifolia</i> contendo motivos SSRs, <i>E-value</i> , nº de acesso, espécie referente, proteína e origem dos transcritos. ....	138
ANEXO A – Protocolo de isolamento de RNA total de pólen utilizando o kit comercial Nucleo Spin Plant®, com modificações. ....	139
ANEXO B – Protocolo de depleção de rRNA em amostras de	

<b>plantas. Ribominus Plant Kit for RNA-Seq (Invitrogen, Carlsbad, EUA), com modificações.....</b>	<b>140</b>
<b>ANEXO C – Tipos de RNAs ribossomais e coeficientes de sedimentação encontrados em plantas. ....</b>	<b>142</b>

# 1 INTRODUÇÃO

Este trabalho faz parte dos múltiplos esforços que estão sendo realizados, no âmbito do Programa de PG em Recursos Genéticos Vegetais, visando a caracterização, conservação e melhoramento das coníferas brasileiras *Araucaria angustifolia* (Araucariaceae) e *Podocarpus lambertii* (Podocarpaceae). As atividades deste projeto atentaram para o saneamento de lacunas substanciais, ao se realizar os primeiros estudos do transcriptoma de diversas estruturas, a partir de espécimes que vivem em um dos habitats naturais das duas espécies. Tratam-se de novos estudos que são adicionais aos estudos que já foram feitos ou que estão em andamento, notadamente nas áreas de genética, ecologia e fisiologia.

Tanto *A. angustifolia* quanto *P. lambertii* possuem grande relevância ecológica e socioeconômica. Embora já demonstraram sua importância social, econômica e ambiental para o sul do país, seu uso predatório as levou para diversas listas nacionais e internacionais de espécies ameaçadas e próximas da ameaça de extinção (CONSEMA, 2014; MMA, 2014; THOMAS, 2013).

Diversos genomas vegetais já foram sequenciados e caracterizados para espécies de angiospermas (GOLD DATABASE, 2015). Entretanto, a escassez de informações para coníferas reduz o entendimento em evolução de plantas e restringem os impactos potenciais na pesquisa genômica deste grupo taxonômico. Muitas lacunas em diversas áreas de pesquisa destas espécies, como a fisiologia do desenvolvimento e a biologia reprodutiva, são entraves para o avanço do entendimento, utilização destes recursos genéticos para novas finalidades e proposição de medidas de conservação.

Em coníferas, o tamanho dos genomas passa de 30 gigabases, fazendo com que o processo de sequenciamento completo seja difícil e demorado em função do grande tamanho e complexidade de suas sequências não codificantes. O sequenciamento de todas as sequências transcritas, denominado de transcriptoma, é rápido e informativo, revelando quais genes estão atuando nas amostras estudadas. Transcriptoma é o conjunto de transcritos (moléculas de RNA) expressos em certas células ou tecidos em determinado momento e condição (ambiental ou controlada). Estes conjuntos representam porções transcritas do genoma da espécie em estudo, sendo muito dinâmicos e dependentes de fatores ambientais como temperatura, umidade relativa

do ar, regime hídrico do solo, luminosidade, interação com micro-organismos, entre outros fatores. Os transcritos conectam o genoma da espécie com funções biológicas e fenótipos, sendo que sua caracterização total é atualmente uma das melhores alternativas para a obtenção de dados genômicos em espécies com grandes genomas, como as coníferas em geral.

O primeiro trabalho caracterizando transcriptomas, via RNA-Seq, em *A. angustifolia* foi realizado por Elbl et al. (2015a). Foram sequenciadas distintas culturas celulares e embriões somáticos e zigóticos, fornecendo mais informações acerca dos mecanismos subjacentes à embriogênese da espécie. No entanto, estes dados não representam os reais perfis de expressão gênica da espécie em seus distintos habitats naturais. Neste contexto, o presente trabalho teve como principal proposta o sequenciamento, com uso de uma plataforma de sequenciamento de DNA de segunda geração, e a caracterização dos transcriptomas de amostras foliares, de hastes e de pólen de *Araucaria angustifolia* e amostras foliares de *Podocarpus lambertii*, coletadas de plantas de ocorrência natural em Lages/SC. Desta forma, os transcritos identificados neste trabalho refletem a interação dos genomas das espécies com seu ambiente natural, fornecendo subsídios valiosos para o melhoramento genético, uso e conservação.

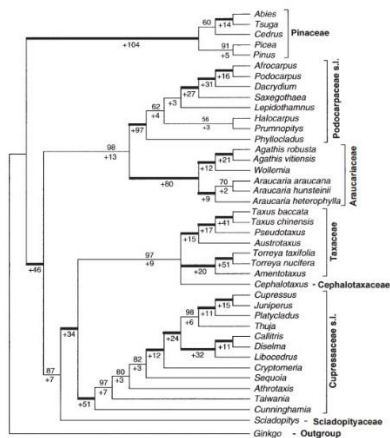


## 2 REVISÃO DE LITERATURA

### 2.1 Coníferas

As coníferas, o grupo mais importante das gimnospermas, representam cerca de 650 espécies divididas em sete famílias: Araucariaceae, Cephalotaxaceae, Cupressaceae, Pinaceae, Podocarpaceae, Sciadopityaceae e Taxaceae. As espécies desse grupo são os maiores e mais antigos organismos terrestres presentes no planeta (AHUJA; NEALE, 2005; DOYLE, 1998; HENRY, 2005). Até o momento foram reconhecidas as sete famílias existentes e mais 12 famílias extintas, que seriam Thucydieaceae, Emporiaceae, Utrechtiaceae, Ferrugliocladaceae, Majoniaceae, Ulmanniaceae, Voltziaceae, Cheirolepidiaceae, Palissyaceae, Pararaucariaceae, Geinitziaceae e Doliostrubaceae. A filogenia das famílias existentes da ordem Coniferales (QUINN et al. 2002; Figura 1), baseada nas sequências dos genes organelares *matK* e *rbcL*, coloca lado a lado as famílias Araucariaceae e Podocarpaceae, estas possuindo representantes no Brasil, como a *A. angustifolia* e *P. lambertii*, entre outras espécies.

Figura 1. Árvore filogenética de espécies representantes da Ordem Coniferales, baseada nas sequências dos genes organelares *matK* e *rbcL*



Fonte: Quinn et al. 2002

As coníferas apresentam imensa importância ecológica, pois

dominam muitas paisagens terrestres e são as plantas vasculares com o maior potencial de fixação de carbono atmosférico. Além disso elas têm uma forte interação com a fauna, em particular por meio de suas sementes, que são alimentos para uma fauna variada. Atualmente, elas estão presentes em um grande número de ecossistemas nos dois hemisférios devido à aquisição de um eficiente sistema fisiológico de adaptação durante o processo de evolução. (DOYLE, 1998; FARJON, 2008; FARJON, 2010; HENRY, 2005; MAGALLÓN; SANDERSON, 2005).

Durante o Holoceno Superior, a expansão florestal sobre as pastagens (paisagens com dominância de gramíneas) ocorreu sob certas condições ambientais, tais como o clima úmido e baixo nível de perturbação, mas os mecanismos relevantes que envolveram este processo ecológico espacial ainda não são bem compreendidos (ARCHER et al. 1988; BEHLING et al. 2004; BEHLING et al. 2005; PILLAR, 2003). Essa expansão das coníferas sobre outros ambientes demonstra uma grande capacidade adaptativa para tolerar condições ambientais diferentes daquelas comumente encontradas nas florestas, como a alta variação de temperatura, níveis de umidade no solo, altas taxas de transpiração e alta competição com as gramíneas (SANKARAN et al. 2004).

No Hemisfério Sul, as floras da América do Sul, Antártida e Austrália continuam, assim como hoje em dia, descendentes da flora de angiospermas de Gondwana meridional, com próteas, murtas, *Nothofagus* e coníferas do Sul, tais como a *Araucaria*, *Podocarpus* e *Dacrydium*. Essas florestas cobriam, pelo menos, a região periférica da Antártida. As florestas de coníferas sul-americanas são caracterizadas por coníferas de afinidade neotropical e subantártida (COX; MOORE, 2009). Atualmente, elas incluem numerosas espécies de Podocarpaceae (gêneros *Podocarpus* e *Retrophyllum*), presentes desde a região tropical, passando pelos subtrópicos até latitudes temperadas; duas espécies da família Araucariaceae, uma no sudeste subtropical da América do Sul (*Araucaria angustifolia*) e outra em latitudes temperadas da região dos Andes (*Araucaria araucana*), além de três espécies (*Austrocedrus chilensis*, *Fitzroya cupressoides*, *Pilgerodendron uviferum*) da família Cupressaceae que ocorrem na região dos Andes em médias e altas latitudes (FARJON, 2008; FARJON, 2010; VEBLEN et al. 2005).

No Brasil são encontradas nove espécies da família Podocarpaceae: oito do gênero *Podocarpus* (*P. lambertii* Klotzch ex Endl., *P. sellowii* Klotzch ex Endl., *P. brasiliensis* Laubenfels, *P. barretoii* Laubenfels & Silba, *P. roraimae* Pilger, *P. acuminatus* Laubenfels, *P.*

*aracensis* Laubenfels & Silba e *P. celatus* Laubenfels) e uma do gênero *Retrophyllum* (*Retrophyllum piresii*) (LEDRU et al. 2007; SOUZA, 2013; STEHMANN et al. 2009). *P. lambertii* e *P. sellowii* são as espécies com maior distribuição do gênero no Brasil e são encontradas na Floresta Atlântica e conhecidas popularmente como pinheiros-bravos (JUDD et al. 2009). *P. lambertii* acompanha a araucária na floresta ombrófila mista, mas sua distribuição geográfica se estende também para a região pampeana até o Uruguai, com populações isoladas na parte central do estado da Bahia. Por sua vez *P. sellowii* se distribui ao longo da floresta atlântica desde o Rio Grande do Sul até as pequenas serras e brejos de altitude no Nordeste, especialmente junto às florestas ombrófilas densa e estacional semidecidual, além das florestas de galerias serranas no domínio do cerrado (OLIVEIRA; ROTTA, 1982; SILVA; MARCONI, 1990; STANNARD, 1995).

*Retrophyllum piresii* (Silba) C.N.Page (Pinheiro-da-Amazônia) é outra espécie da família Podocarpaceae, endêmica do Brasil e considerada extremamente rara. A espécie é encontrada apenas no estado de Rondônia no município de Guajará-Mirin, integrante do Parque Estadual do Guarujá-Mirin e do Parque Estadual da Serra dos Pacaás Novos. No entanto os dados sobre esta espécie ainda são deficientes, principalmente em função de sua difícil localização (GARDNER, 2013). Em função disso a espécie pode estar altamente ameaçada de extinção devido ao avanço das fronteiras agrícolas na região norte do país. É admitido que *R. piresii* possa ocorrer também no Peru e Bolívia, mas ainda não existe tal comprovação (GARDNER, 2013). Atualmente a espécie é cultivada e propagada no horto do Museu Paraense Emílio-Goeldi, em Belém, Pará (PAGE, 1989; SILBA, 1983). O plastoma completo da espécie foi recentemente sequenciado e anotado funcionalmente, representando um importante recurso genômico para o entendimento de relações filogenéticas das coníferas nativas do Brasil (VIEIRA et al. 2016).

## **2.2. *Araucaria angustifolia* (Bert.) O. Kuntze**

O gênero *Araucaria* de Jussieu possui 19 espécies e é o mais diversificado da família Araucariaceae (HENKEL; HOCHSTETTER, 1865), com distribuição restrita ao hemisfério sul (Chile, Argentina, sul do Brasil, Nova Caledônia, Ilha Norfolk, Austrália, Nova Guiné e Nova Zelândia) (GOLTE, 1993; SETOGUCHI et al. 1998). Sua presença e distribuição atual é considerada como relíquia da evolução (FARJON, 2008; FARJON, 2010; SETOGUCHI et al. 1998), uma vez que seus

fósseis estão entre os mais antigos existentes dos gêneros de coníferas, sendo encontrados desde os períodos Cretáceo, Jurássico e Triássico (FARJON, 2008; FARJON, 2010; MILLER, 1988; STOCKEY, 1982), as quais estão presentes na terra a aproximadamente 300 milhões de anos. Apesar de sua atual importância em áreas naturais e do seu aspecto evolutivo, poucos estudos têm sido realizados focando aspectos filogenéticos e evolutivos (ESCAPA et al. 2013; KRANITZ et al. 2014; SETOGUCHI et al. 1998; STEFANOVIC et al. 1998; STEFENON et al. 2006).

*Araucaria angustifolia* (Bert.) O. Kuntze, ou Pinheiro do Paraná, ou Pinheiro Brasileiro, é uma conífera dioica e perenifólia, que ocorre exclusivamente no Brasil, Argentina e Paraguai, na região da Floresta Ombrófila Mista (FOM). Esta formação florestal também recebe o nome de Floresta com Araucária, devido à abundância e ao grande porte da espécie, que imprime a fisionomia característica da floresta. Esta espécie é endêmica das regiões sul (Estados de Santa Catarina, Paraná e Rio Grande do Sul) e sudeste do Brasil (Estados de São Paulo, Minas Gerais e Rio de Janeiro), com extensões em pequenas manchas ao noroeste da Argentina e Paraguai, em áreas próximas às fronteiras brasileiras, onde encontra condições ideais para o desenvolvimento em altitudes entre 500 m e 1800 m (GUERRA et al. 2008; JUDD et al. 2009; KLEIN, 1960; MATTOS, 2011; REITZ; KLEIN, 1966; VELOSO et al. 1991; VELOSO; GOES FILHO, 1982) (Figura 4).

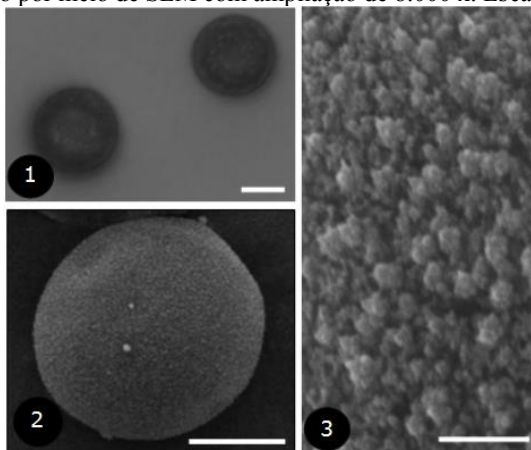
*A. angustifolia* é polinizada principalmente pelo vento e no sul do Brasil a deiscência dos grãos de pólen ocorre de agosto a novembro, sendo que o pico de polinização ocorre no mês de setembro. O estróbilo masculino (cone, mingote) é visível de agosto a janeiro e o estróbilo feminino (pinha) é visível durante o ano todo (Figura 3). O desenvolvimento do gametófito masculino é controlado por dois programas de desenvolvimento sucessivos, sendo o primeiro logo após a meiose (transcriptoma inicial) e o segundo logo após a primeira divisão mitótica (transcriptoma tardio). Os mRNAs do transcriptoma inicial são efêmeros, logo desaparecendo antes da completa maturação dos grãos de pólen. Os transcritos tardios são estáveis, resistindo a desidratação durante a maturação e posterior reidratação no momento da polinização. Estes transcritos são os responsáveis pela síntese proteica durante a germinação do tubo polínico, sendo que a transcrição após a maturação dos grãos é quase inexistente (RAFINSKA et al. 2010; TWELL et al. 2006; ZIENKIEWICZ et al. 2006).

O transcriptoma tardio dos grãos de pólen pode ser dividido em

quatro grupos, considerando apenas os mRNAs: (1) transcritos responsáveis pela estrutura da parede celular; (2) transcritos responsáveis pela regulação da transcrição e tradução (Incluindo tRNAs e rRNAs); (3) transcritos responsáveis pela regulação do transporte via membranas; (4) transcritos responsáveis pela regulação do ciclo celular (RAFINSKA et al. 2010).

Os grãos de pólen de *A. angustifolia* possuem aproximadamente 50  $\mu\text{m}$  de diâmetro, sendo similar em tamanho ao *Pinus taeda* e maiores em relação a outras coníferas como *Pinus banksiana* (35  $\mu\text{m}$ ). A forma é esférica e não existem estruturas para auxiliar o vôo. Estes fatores contribuem para uma baixa velocidade terminal dos grãos de pólen, de aproximadamente 3  $\text{cm.s}^{-1}$  (WILLIAMS, 2009). A superfície possui aspecto granulado em função da deposição irregular de exina (Figura 2, STEFENON; CAPRESTANO, 2009).

Figura 2. 1: Grão de pólen de *A. angustifolia* visualizado em estereomicroscópio ampliado 100 x. Escala: 25  $\mu\text{m}$ . 2: Grão de pólen visualizado por meio de SEM com ampliação de 400 x. Escala: 25  $\mu\text{m}$ . 3: Detalhe da superfície do grão de pólen (exina) visualizado por meio de SEM com ampliação de 8.000 x. Escala: 2  $\mu\text{m}$



Fonte: Adaptado de Stefenon e Caprestano, 2009

As pinhas amadurecem de fevereiro a dezembro, de acordo com a variedade botânica (Figura 3) (MATTOS, 2011; REITZ; KLEIN, 1966). A monoícia é rara e tem sido creditada a infecções patogênicas, injúrias em árvores adultas e a estresses fisiológicos (REITZ; KLEIN, 1966). Contudo, Stefenon e Caprestano (2009) descreveram uma planta monóica sem evidências de patógenos ou injúrias. Em sua região de ocorrência

natural, *A. angustifolia* tem grande relevância ecológica, econômica e social. Trata-se de uma espécie secundária longeva dominante de crescimento inicial lento, mas de temperamento pioneiro. Este temperamento pioneiro é de grande importância, visto que a *A. angustifolia* tem um papel ecológico como espécie berçário no avanço de espécies florestais sobre os campos adjacentes, pois cria um ambiente ideal para outras espécies tolerantes à sombra, como bromélias, orquídeas e outras espécies epífitas, que encontram ambiente adequado para o desenvolvimento (DUARTE et al. 2006; DUARTE; DILLENBURG, 2000; FRANCO; DILLENBURG, 2007).

Figura 3. A: Área de estudo, Coxilha Rica, Lages - SC. B: Ginostróbilo maduro de *A. angustifolia*. C: Cones masculinos (androstróbilo) de *A. angustifolia*

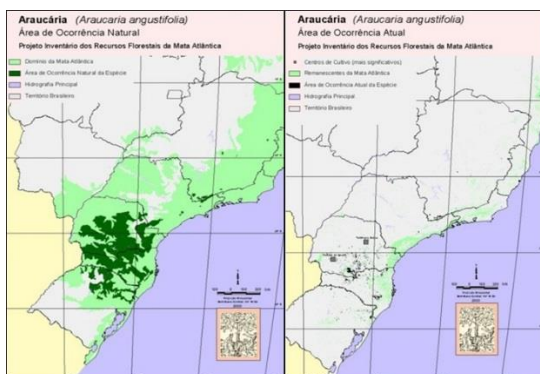


Fonte: A: Núcleo de Pesquisas em Florestas Tropicais, 2012; B: Rubens Onofre Nodari, 2010; C: CiFlorestas, 2015

Ainda em relação ao aspecto ecológico, suas sementes (pinhões) servem de alimento para a fauna silvestre (principalmente aves e roedores, os quais são os principais dispersores das suas sementes), para a qual é considerada espécie chave, pois fornece alimento rico em carboidratos durante o início das estações mais frias, enquanto as outras espécies da floresta não estão produzindo frutos. Muitos roedores enterram as sementes para a utilização deste recurso no inverno, fato este

reportado por Schüssler (2014). Do ponto de vista econômico e social, suas sementes servem de alimento humano com grande valor nutricional, possuindo 54% de amido e 5,1% de proteínas (ROSADO et al. 1994), sendo que a coleta e a venda destas sementes se constituem em relevante atividade econômica para um considerável número de famílias que vivem nas regiões de ocorrência (SCHÜSSLER, 2014). Esta espécie também possui madeira de alta qualidade o que determinou sua importância econômica histórica, sendo empregada especialmente para construção civil, de móveis, como também, para a produção de celulose (GUERRA et al. 2008; FARJON, 2008; FARJON, 2010; MATTOS, 2011; REITZ et al. 1978). Devido à drástica redução em sua área de ocorrência natural (Figura 4), o pinheiro-brasileiro atualmente consta em diversas listas nacionais e internacionais de espécies ameaçadas de extinção. Na lista do Ministério do Meio Ambiente, a espécie se encontra em perigo de extinção (MMA, 2014). O Conselho Estadual do Meio Ambiente de Santa Catarina classifica a espécie como criticamente em perigo de extinção (CONSEMA, 2014), já na lista da IUCN, a araucária é classificada como criticamente ameaçada de extinção (THOMAS, 2013).

Figura 4. Áreas de ocorrência natural “histórica” e atual de *A. angustifolia* (Bert.) O. Kuntze (em verde escuro), dentro do domínio da Mata Atlântica (área verde clara)



Fonte: Guerra et al. 2002

Ainda restam apenas alguns remanescentes isolados de *A. angustifolia* (Figura 4), os quais representam menos de 2% da área total de sua ocorrência histórica que abrangia uma área correspondente a 40%, 31%, 25%, 3% e 1% do território dos estados do Paraná, Santa Catarina, Rio Grande do Sul, São Paulo e Minas Gerais, respectivamente

(GUERRA et al. 2000; GUERRA et al. 2002; GUERRA et al. 2008; REITZ et al. 1978; VELOSO et al. 1991). Este tipo de exploração, sem critérios de seleção ou manejo, vem causando erosão genética nesta espécie em função da extrema fragmentação e isolamento de fragmentos florestais (AULER et al. 2002, CRISTOFOLINI, 2013). Em função disso, diversos trabalhos foram realizados visando caracterizar a fenologia, ciclo reprodutivo, variabilidade genética, estrutura e distância genética em populações de *A. angustifolia* utilizando ferramentas moleculares em nível de DNA nuclear e plastidial, RNA e proteína (AULER et al. 2002; BITTENCOURT; SEBBENN, 2007; CRISTOFOLINI, 2013; ELBL et al. 2015a; KLABUNDE, 2012; MANTOVANI et al. 2004; PATREZE; TSAI, 2010; ROGGE-RENNER et al. 2013; SANT'ANNA et al. 2013; SEDREZ DOS REIS et al. 2012; SEDREZ DOS REIS et al. 2014; SCHLÖLG et al. 2012; STEFENON et al. 2007; STEFENON et al. 2008; SCHÜSSLER, 2014; ZECHINI, 2012). No entanto, o baixo número de informações genômicas, como sequências de genes, transcritos e isoformas, para *A. angustifolia* restringe os impactos potenciais na pesquisa genômica, melhoramento genético e conservação desta espécie.

### 2.3 *Podocarpus lambertii* Klotzsch ex Eichler

A família Podocarpaceae (ENDLICHER, 1847), (divisão Pinophyta, ordem Pinales), contém 18 gêneros e aproximadamente 173 espécies, as quais são representadas por árvores e arbustos, utilizados para fins madeireiros e ornamentais, e distribuídos principalmente no Hemisfério Sul e sudeste asiático (VEBLEN et al. 2005).

O gênero *Podocarpus* é um dos mais diversos entre todas as coníferas e ocorre em todas as porções de terra originadas a partir da Gondwana, exceto na Antártida. Engloba mais de 100 espécies, muitas das quais apresentam grande interesse florestal e apresenta ampla distribuição geográfica mundial, sendo registrada sua presença na Nova Caledônia, Sudeste da Ásia, China, Japão, Malásia, Austrália, Nova Zelândia, Bornéu, Nova Guiné, Ilhas do Pacífico, Ilhas Fiji, Antilhas, Américas Central e do Sul (KELCH, 1997; KELCH, 1998). Este gênero foi utilizado recentemente, por intermédio da palinologia, para reconstruções climáticas da região amazônica (LEDRU et al. 2007), pois alguns gêneros da família Podocarpaceae tem um registro fóssil que data por volta do Cretáceo (KRASSILOV, 1974) e podem gerar informações importantes sobre o efeito das mudanças climáticas ocorridas durante a evolução destas espécies e sua distribuição atual (LEDRU et al. 2007).



*P. lambertii*, ou Pinho Bravo, ou Pinheiro Bravo, é uma espécie dióica, apresentando o desenvolvimento das estruturas reprodutivas masculinas de abril a junho e o estróbilo feminino de setembro a maio (Figura 5). Os principais vetores de polinização são insetos da ordem hymenoptera, principalmente as abelhas. Na região sul do país, as sementes amadurecem de dezembro a fevereiro, sendo que a principal dispersão é ornitocórica (CARVALHO et al. 2003).

Figura 5. A: Indivíduo adulto de *Podocarpus lambertii* B: Regenerantes de *P. lambertii* identificados com etiqueta metálica. C: Ramo de planta masculina contendo cones (estróbilos). D: Ramo de planta feminina contendo sementes imaturas e pedúnculo carnoso (epimácio)



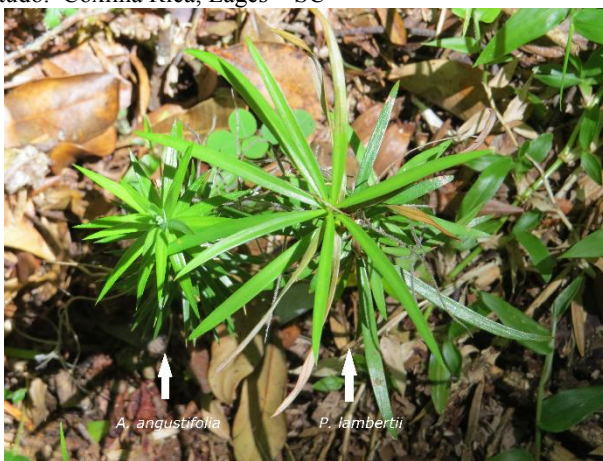
Fonte: A: Martin Molz 2008; B-C-D: Alison Bernardi 2015

*P. lambertii* ocorre tanto em áreas sombreadas quanto expostas ao sol. A definição quanto ao grupo sucessional ao qual pertence é difícil, pois as informações na literatura são bastante controversas, constando como uma espécie secundária (CHIAMOLERA et al. 2010) ou pioneira (FONTOURA et al. 2006; INOUE et al. 1984). As matas de *P. lambertii* ocorrem na região de origem da *A. angustifolia* e da Floresta Ombrófila Mista (FOM), no Sul do Brasil, intimamente ligadas a essas matas, nos vales e depressões planas do planalto (LONGHI et al. 2010). A espécie também é encontrada nos estados do Sudeste, com exceção do Espírito Santo, e na região central estado da Bahia, na Floresta Estacional Decidual Montana (STANNARD, 1995) (Figura 6).



fauna, sendo um recurso que culmina na dispersão das sementes, principalmente por pássaros. Isto pode permitir que os regenerantes da espécie (Figura 7) se desenvolvam longe de plantas aparentadas, promovendo altos índices de diversidade genética. Além disso, a coloração roxa da estrutura é um indicador de maturidade fisiológica das sementes (RAGAGNIN; COSTA; HOPPE, 1994).

Figura 7. Regenerantes de *A. angustifolia* (esquerda) e *P. lambertii* (direita), na área de estudo. Coxilha Rica, Lages – SC



Fonte: Alison Bernardi, 2015

Embora estas espécies da família Podocarpaceae demonstrem um grande potencial ornamental, na indústria madeireira, bioquímica e farmacêutica, pouco do que se sabe se relaciona com o gênero *Podocarpus*, o qual sofreu com o desmatamento, onde o corte seletivo eliminou os melhores ou maiores exemplares, sem qualquer tipo posterior de estratégia visando conservação e renovação da espécie na floresta. Pouco são os estudos relacionados à estrutura reprodutiva, maturidade, crescimento, variabilidade genética, ecofisiologia e tratos silviculturais (CHIAMOLERA et al. 2010; BERNARDI, 2014; MARIN-RAGAGNIN et al. 1994; ZIMMER et al. 2010).

## 2.4 Sequenciamento de Ácidos Nucleicos

A química dideoxi, ou método de Sanger (SANGER et al. 1977), permaneceu durante quase três décadas como padrão ouro de sequenciamento de ácidos nucleicos. Este período, denominado de primeira geração de sequenciamento foi responsável por uma revolução em todas as áreas da genética. Mas em contrapartida possuiu grandes limitações em termos de alto valor por base sequenciada, elevado tempo de projetos e número limitado de sequências analisadas (DAVIES, 2010). A partir de 2001, após a conclusão do Projeto Genoma Humano (PGH) (LANDER et al. 2001) e do sequenciamento do genoma humano pelo grupo do Dr. John Craig Venter (VENTER et al. 2001), foi lançada uma corrida para o desenvolvimento de químicas de sequenciamento e instrumentos que em pouco tempo fossem capazes de sequenciar qualquer genoma a um custo reduzido. Esta necessidade surgiu do após o PGH levar 10 anos e custar aproximadamente US\$3 bilhões para ser concluído, com a utilização de plataformas de sequenciamento de primeira geração, como a ABI 3730, MegaBACE 1000 e 5000. Esta corrida foi denominada de “genoma de mil dólares” (*The \$ 1,000 genome*), sendo responsável por diversos avanços na genômica (DAVIES, 2010).

A técnica de sequenciamento total de RNA, conhecida como RNA-Seq, foi concebida no início da segunda geração de sequenciamento de DNA. A partir de 2006, esta segunda geração foi chamada de *Next Generation Sequencing* (NGS), em função do desenvolvimento de sequenciadores de DNA com maior capacidade operacional e custo reduzido por base sequenciada em comparação aos sequenciadores de DNA de primeira geração, baseados na química dideoxi (DAVIES, 2010; SANGER et al. 1977).

As primeiras plataformas de segunda geração, como a 454, SOLiD, Polonator e Illumina foram todas baseadas em químicas distintas, mas iguais no princípio de sequenciamento por detecção de luz, onde cada base nitrogenada incorporada na leitura emite um comprimento de onda distinto (METZKER, 2010). Já as três últimas plataformas de segunda geração, como a *Ion Torrent Personal Genome Machine* (PGM), a Ion Pronton e a IonS5 não foram baseadas na detecção de luz, mas sim na variação de pH. Em toda formação de ligação fosfodiéster, a enzima DNA polimerase libera como subproduto um íon H<sup>+</sup>. O sistema realiza injeções sucessivas dos quatro deoxinucleotídeos (dATP, dCTP, dGTP e dTTP), limpando o sistema após cada base injetada. A construção da leitura é realizada pela quantidade de variação de pH detectada em cada injeção,

gerando até 80 GB de seqüências por chip de sequenciamento (ROTHBERG et al. 2011).

Independente do sequenciador e da química de segunda geração utilizada, as corridas de sequenciamento geram GB de dados com tamanhos e qualidades variadas. A bioinformática é a disciplina responsável por transformar estes dados oriundos de luz ou de variação de pH em informações biológicas que nos permitam compreender uma infinidade de vias metabólicas e processos celulares em tão pouco tempo (METZKER, 2010; NEKRUTENKO; TAYLOR, 2012).

Atualmente, a comunidade científica internacional admite o início da terceira geração de sequenciamento de DNA, com a química SMRT (Single Molecule Real Time) pertencente a plataforma PACBIO RSII (Pacific Biosciences). Esta química permite o sequenciamento, por exemplo, do genoma humano com análise em tempo real a partir de uma única cópia de cada cromossomo, com leituras apresentando mais de 20 kb, contra os quase 0,5 kb das plataformas de segunda geração. Esta nova tecnologia revolucionará ainda mais as análises genômicas, transcriptômicas, epitranscriptômicas e epigenômicas, mas atualmente o alto custo por base sequenciada ainda é um entrave à popularização da técnica (PACIFIC BIOSCIENCES, 2015). Uma segunda plataforma de terceira geração está atualmente em desenvolvimento pela empresa Oxford Nanopore. Ela promete o sequenciamento de DNA e RNA em pequenos dispositivos do tamanho de pendrives (MinION™). O sequenciamento acontece em nanoporos e sem a necessidade de preparo de bibliotecas, amplificação e marcação de bases com fluoróforos. As leituras obtidas possuem tamanho semelhante aos da química SMRT, encurtando o tempo de obtenção dos genomas e transcriptomas devido a simplificação das etapas de alinhamento e montagem (OXFORD NANOPORE, 2015). A tendência para os próximos anos é a da utilização conjunta de plataformas de terceira e segunda geração de sequenciamento, objetivando a obtenção de leituras extremamente longas (> 20 kb) e fechamento de gaps, respectivamente.

## **2.5 Análise transcriptômica (RNA-Seq)**

Transcriptoma é o conjunto de transcritos (moléculas de RNA) expressos em certas células, tecidos ou estruturas, em determinado momento e condição (ambiental ou controlada). Estes conjuntos representam porções transcritas do genoma da espécie em estudo. Os transcriptomas são extremamente dinâmicos e dependentes de fatores

ambientais como temperatura, umidade relativa do ar, regime hídrico do solo, luminosidade, interação com micro-organismos, entre outros fatores (MARTIN; WANG, 2011; WANG et al. 2009b). Os transcritos conectam o genoma da espécie com funções biológicas e fenótipos. A caracterização total de determinado conjunto de transcritos é atualmente uma das melhores alternativas para a obtenção de dados genômicos em espécies com grandes genomas, como as coníferas em geral (MACKAY e DEAN, 2011).

Considerando um conjunto hipotético de transcritos de determinada célula vegetal, a grande maioria das moléculas de RNA serão do tipo ribossomal (mais de 80% de rRNA), seguidos de moléculas de RNA transportador (cerca de 15%) e por fim as moléculas de RNA mensageiro e outros pequenos RNAs (aproximadamente 5%) (LODISH et al. 2000). Quando o interesse é caracterizar a expressão gênica total em nível de mRNA e microRNAs, as sequências dos RNAs ribossomais e transportadores se tornam redundantes e dificultam a detecção de transcritos com baixa expressão. Em função disso se aplicam técnicas de enriquecimento de mRNA (ex. via cauda poli-A) ou de depleção de rRNA visando a drástica redução ou eliminação das moléculas ribossomais (O'NEIL et al. 2013).

A detecção simultânea dos milhares de transcritos que compõe determinado transcriptoma exige o emprego de técnicas de sequenciamento de DNA de nova geração (NGS). A técnica de RNA-Seq (Sequenciamento de RNA) é utilizada em diversas plataformas de sequenciamento de RNA de segunda geração, como Illumina HiSeq, Ion Torrent PGM, Ion Proton e 454 FLX. A técnica é caracterizada pela construção de bibliotecas a partir da fragmentação do mRNA e posterior síntese de cDNA. Este cDNA é amplificado, purificado e sequenciado da mesma maneira que o DNA genômico. Como resultado, são obtidos milhões de leituras curtas, com tamanho variando em função da química de sequenciamento e estratégia de sequenciamento utilizada (WANG et al. 2009b). Os primeiros trabalhos publicados que utilizaram a técnica de RNA-Seq para o estudo de transcriptomas surgiram em 2008. Mortazavi et al. (2008) sequenciaram os transcritos de tecidos de camundongos como cérebro, fígado e músculo estriado esquelético. Nagalakshmi et al. (2008) sequenciaram os transcritos de *Saccharomyces cerevisiae*, revelando os primeiros padrões completos de expressão gênica da espécie. A primeira grande revisão publicada sobre a técnica de RNA-Seq em análise transcriptômica surgiu no ano de 2009 (WANG et al. 2009b). Na revisão, a ferramenta é classificada como revolucionária, pois

forneceu novos subsídios para estudos de expressão gênica em larga escala.

Após isso, são efetuadas análises de bioinformática visando a retirada das sequências adaptadoras, *barcodes*, leituras de baixa qualidade e de sequências contaminantes. As leituras restantes são então alinhadas, montadas e mapeadas contra uma referência (método referenciado - *reference based*), que de preferência deve ser o próprio genoma da espécie em estudo ou outro genoma de uma espécie do mesmo gênero. Caso a espécie em estudo não possua genoma sequenciado, como *A. angustifolia* e *P. lambertii*, pode ser aplicado o método *de novo*, que realiza o alinhamento e montagem apenas levando em consideração as próprias leituras sequenciadas. Neste caso, quanto menor o tamanho da leitura sequenciada, mais difícil será a montagem do transcriptoma em questão. Porém, os algoritmos de montagem de novo estão sendo cada vez mais aperfeiçoados e as químicas de sequenciamento estão sendo atualizadas para um aumento no tamanho de leitura sequenciada. Assim, a falta de informações genéticas para determinada espécie incipientemente estudada (*non-model organisms*) não é mais considerado um entrave para a realização de experimentos (SURGET-GROBA; MONTOYA-BURGOS, 2010).

A partir do sequenciamento completo dos transcritos de determinada amostra, é possível o agrupamento destes mRNAs em diferentes categorias, tais como processos biológicos, componentes celulares e funções moleculares. Desta forma é possível caracterizar quais processos e vias metabólicas estão em atuação em determinadas células ou tecidos (ASHBURNER et al. 2000).

Anteriormente à técnica de RNA-Seq, estudos de expressão gênica total em larga escala requeriam elevado volume de trabalho para a obtenção de poucas centenas de *Expressed Sequence Tags* (EST). Entre as diversas etapas envolvidas, estavam presentes a construção de vetores, transformação, clonagem bacteriana e sequenciamento Sanger (primeira geração de sequenciamento). As desvantagens estavam no altíssimo volume de trabalho e custo para a obtenção de poucas centenas de EST, sendo que existia o risco da não obtenção dos transcritos de real interesse (MACKAY; DEAN, 2011). Em meados dos anos 2000, com o aumento exponencial do número de sequências de DNA, RNA e proteínas e depósitos em bases de dados, foi criada a técnica de Microarranjo (MA - *Microarray*). Esta técnica consiste em um arranjo pré-definido de milhares de sequências de oligonucleotídeos quimicamente ligadas a um chip. Basicamente, o RNA isolado de determinada amostra é colocado em

contato com os oligos do *chip* e assim que duas sequências complementares se ligam, ocorre emissão de luz que é captada por um scanner específico. Desta forma, se tornou possível monitorar o nível de expressão gênica de milhares de genes e de suas variantes de *splicing*, com mais precisão e menor tempo do que em relação a obtenção de EST (HOHEISEL, 2006). Entretanto, o número de estudos transcriptômicos publicados, com a utilização da técnica de microarranjo, a partir de 2008 diminuiu significativamente muito em função da robustez e alto rendimento (*Throughput*) da técnica de RNA-Seq (SHENDURE, 2008).

Em coníferas, uma das primeiras análises transcriptômicas utilizando microarray foi realizada em *Picea sitchensis* (RALPH et al. 2006), com o objetivo de mensurar a dinâmica da expressão gênica durante a herbivoria causada pelas principais pragas. Com o aumento do número de dados de sequências de DNA, RNA e de *microarray* para coníferas, foram criadas diversas bases de dados visando o armazenamento e a disseminação destas informações entre diversos grupos de pesquisa. As bases de dados de sequências transcritas, para grupos de espécies, são extremamente valiosas por concentrarem informações em nível de RNA, especialmente para espécies com escassez de informações, como as coníferas.

Uma das mais importantes bases de dados para coníferas é a EuroPineDB. Esta base de dados possui aproximadamente 1 milhão de leituras RNA-Seq de xilema, floema, acículas, embriões, raízes, hastes e culturas embriogênicas. Na EuroPineDB foram caracterizados mais de 55.000 Unigenes, em sua maioria pertencendo a *Pinus pinaster* (FERNÁNDEZ-POZO et al. 2011). A SustainPineDB (<http://www.scbi.uma.es/sustainpine/>) é outra base de dados de sequências transcritas de *Pinus pinaster*. Esta base de dados é composta por 11 instituições de 4 países europeus que trabalham exclusivamente com as diversas "ômicas" do Pinheiro-marítimo. Estão disponíveis sequências de mais de 200.000 unigenes caracterizados por meio de RNA-Seq. (CANALES et al. 2014). A base de dados PiceaGenExpress contém 23.543 sequências de genes expressos de *Picea abies* Karst. L. e *Picea sitchensis*. Um diferencial desta base de dados é a construção de um *Microarray*, baseado nas sequências de *Picea sp.*, para ser utilizado em outras coníferas (RAHERISON et al. 2012). Existem também outras bases de dados com menor volume de dados brutos e anotados, como a Treenomix database, Arborea database, TreeGenes database, Accelerating Pine genomics database e ForestTreeDB (PAVY et al. 2007).

Mackay e Dean (2011) e MacKay et al. (2012) propõem uma



série de iniciativas para caracterizar os gigas genomas das coníferas. Entre estas iniciativas, a análise transcriptômica (via RNA-Seq) tem sido fundamental para o aumento da compreensão de diversos processos biológicos e mecanismos evolutivos em coníferas. Uma informação essencial que antecede qualquer análise transcriptômica é o conhecimento prévio do tamanho do genoma da espécie em estudo. O tamanho do genoma dá uma estimativa de tamanho dos transcriptomas. Assim, as bibliotecas de cDNA são preparadas para apresentarem alta cobertura de sequenciamento, aumentando a chance de identificação de transcritos raros ou de transcritos diferencialmente expressos. Durante a evolução das coníferas, tem se observado o aumento significativo dos genomas das espécies, desde as espécies basais até as mais recentes (AHUJA; NEALE, 2005; MORGANTE; DE PAOLI, 2011). Essa expansão se deve basicamente aos elementos transponíveis (transposons e retrotransposons) e a sequências de natureza repetitiva, como as regiões satélites, mini-satélites e microssatélites. Estas sequências não possuem natureza auto-duplicativa, como a dos elementos transponíveis. O aumento no número de cópias das sequências repetitivas se deve a erros cometidos pelas DNA polimerases (adição equívoca de motivos repetitivos - *Slippage*) durante a replicação do DNA (AHUJA; NEALE, 2005; CHOUDHARY, TRIVEDI, 2010; NYSTEDT et al. 2013).

Atualmente, o número de genomas completamente sequenciados de coníferas é baixo devido ao seu grande tamanho, em torno de 15-30 gigabases (AHUJA et al. 2001; MORGANTE; DE PAOLI, 2011; MURRAY et al. 2010). O genoma de coníferas até agora estudado está repleto de sequências não codificadoras altamente repetitivas (exemplificando, os genomas de *Pinus sp.* são aproximadamente 160 vezes maior do que o genoma de *Arabidopsis thaliana* e sete vezes maior que o genoma humano), sendo maior do que qualquer outro genoma sequenciado até o momento (AHUJA et al. 2001; AHUJA; NEALE, 2005; FRIESEN et al. 2001; RALPH et al. 2008). Somente os genomas de três coníferas estão completamente sequenciados. O genoma completo do Espruce-europeu (*Picea abies*) foi sequenciado e caracterizado por Nystedt et al. (2013). O genoma possui 19,6 Gb de tamanho (1n), cariótipo 2n=24 cromossomos e 28.354 genes. O número de genes é próximo ao de *Arabidopsis thaliana* (27.407 genes), que possui genoma 100 vezes menor do que *P. abies*. O genoma completo do Pinheiro-do-Canadá (*Picea glauca*) foi sequenciado e caracterizado por Birol et al. (2013), apresentando 20,8 Gb de tamanho. O genoma de *Pinus taeda* (ZIMIN et al. 2014) apresenta características similares ao genoma de

*Picea glauca*, como  $n=12$  e tamanho de 20,15 Gb.

Outras espécies de coníferas estão com o sequenciamento do genoma nuclear em fase final, entretanto, ainda é baixa a cobertura de sequenciamento das seguintes espécies: *Pinus taeda*, *Abies sibirica*, *Juniperus communis*, *Taxus baccata* e *Gnetum gnemon* (GOLD DATABASE, 2015).

Na família Araucariaceae foi determinado o tamanho dos genomas (valor 2C) de todas as 19 espécies, com a utilização da técnica de citometria de fluxo Zonneveld et al. (2012a). Todas as araucariaceas apresentam o mesmo número cromossômico,  $2n=26$ . *Araucaria araucana*, nossa vizinha sul-americana com distribuição no Chile e Argentina, apresenta o maior genoma da família, com valor 2C de 45,5 pg de DNA. Um picograma representa cerca de 109 pares de base. *A. angustifolia*, apresenta o segundo maior genoma, com valor 2C de 44,7 pg de DNA. Os autores identificaram uma forte correlação entre o tamanho dos genomas e a distribuição espacial das espécies desta família. Todas as 13 espécies com distribuição restrita à Nova Caledônia possuem valor médio de  $2C = 31,9$  pg de DNA e espécies como *Agathis dammara* e *Wollemia nobilis* apresentam valor 2C de 26,9 e 28,4 pg de DNA, respectivamente. Estes distintos tamanhos de genomas por região fornecem hipóteses sobre a adaptação destas espécies em seus respectivos ambientes.

Já os genomas das espécies da família Podocarpaceae são menores em relação às espécies da família Araucariaceae. Zonneveld (2012b) avaliou cinco espécies do gênero *Podocarpus*, sendo que o valor 2C variou entre 15,9 e 19,4 pg.

Chen et al. (2012) caracterizaram o transcriptoma e avaliaram os níveis de expressão gênica diferencial, via RNA-Seq, de acículas de *Picea abies* Karst. L. coletadas em condições de ampla luminosidade e total escuro. Além disso, quase 15 mil marcadores SNPs foram identificados nos dados de RNA-Seq. Em *Pinus monticola*, a análise transcriptômica das acículas, via RNA-Seq, revelou uma série de interações moleculares entre *P. monticola* e seu principal patógeno, um fungo causador de uma severa ferrugem. Foram identificados os principais genes do patossistema e possíveis mecanismos de resistência (LIU et al. 2013). Carvalho et al. (2013) caracterizaram transcriptomas de xilema e floema de diversas coníferas e identificaram genes responsáveis pelo crescimento secundário e formação da madeira. Foram sequenciados e caracterizados diversos genes da família gênica Cesa, responsável pela enzima celulose sintase. O transcriptoma do pinheiro marítimo (*Pinus pinaster*) foi caracterizado

por Canales et al. (2014) com a utilização das plataformas de segunda geração Illumina e 454. Diversas amostras de plantas adultas foram sequenciadas, incluindo cones e estróbilos masculinos e femininos, xilema, floema, gemas, folhas, raízes e embriões somáticos. A partir da montagem *de novo* foram identificados 26.020 unigenes possuindo ortólogos, 9.799 unigenes sem ortólogos, 9.642 FL-cDNA e 176 ncRNAs. Também foram identificados marcadores SSRs e SNPs.

Outros transcriptomas de *Pinus pinaster* foram sequenciados e caracterizados por Cañas et al. (2014). Foi utilizada a técnica de microdissecção e captura a laser para a separação e coleta de pequenos tecidos de plântulas recém germinadas. Estes tecidos foram sequenciados com a plataforma de sequenciamento 454 e os resultados forneceram informações importantes relacionadas a germinação e ao desenvolvimento inicial dos tecidos.

O primeiro trabalho caracterizando transcriptomas, via RNA-Seq, em *A. angustifolia* foi realizado por Elbl et al. (2015a). Foram sequenciadas distintas culturas celulares e embriões somáticos e zigóticos, fornecendo mais informações acerca do controle da embriogênese da espécie. Foram preditos 112.772 unigenes a partir do sequenciamento de 24 bibliotecas de cDNA.

O presente estudo é o primeiro trabalho que visa caracterizar transcriptomas, via RNA-Seq e sem genoma nuclear de referência, de *A. angustifolia* e *P. lambertii* a partir de amostras obtidas diretamente de um fragmento de floresta ciliar, dentro de campos naturais (Estepe Ombrófila), sendo esta uma das áreas de ocorrência natural destas espécies.

## **2.6 Mineração e desenvolvimento de marcadores moleculares SSRs e SNPs**

Após a montagem dos transcritos, é possível identificar a presença de diversos tipos de padrões de sequência nos unigenes. Entre estes padrões, é possível e simples a identificação de marcadores moleculares SNP (*Single Nucleotide Polymorphism*) e SSR (*Simple Sequence Repeats*). Estes marcadores vêm sendo extensivamente utilizados há mais de 20 anos em genética populacional, evolutiva, forense e melhoramento genético (BLANCA et al. 2011). É importante salientar que a identificação de marcadores moleculares neste estudo não é uma simples consequência da grande quantidade de dados obtidos, mas é realizada em função destas marcas serem transcritas e baseadas em

genes. Portanto, é elevada a probabilidade de se encontrarem marcas associadas a genes de interesse (GAO et al. 2012).

Os marcadores SNPs, também conhecidos como SNV (*Single Nucleotide Variation*), são extremamente abundantes nos genomas eucariotos, sendo a marca molecular mais abundante em qualquer genoma. A marca é caracterizada pela mudança de uma única base em uma sequência de DNA. As mutações podem ser de dois tipos, as transições e as transversões de bases. As transições acontecem quando ocorre a mudança de bases nitrogenadas do mesmo tipo (base púrica  $\Leftrightarrow$  base púrica ou base pirimidica  $\Leftrightarrow$  base pirimidica). Já as transversões são compostas pela mudança de tipos diferentes de bases nitrogenadas (base púrica  $\Leftrightarrow$  base pirimidica) (JEHAN; LAKHANPAUL, 2006).

A herança co-dominante é uma importante característica responsável pela alta empregabilidade deste tipo de marcador molecular. Porém, a maioria das marcas são bi-alélicas, variando apenas em uma transversão ou transição de bases nitrogenadas. A taxa de mutação em sítios neutros varia de  $1 \times 10^{-1}$  a  $5 \times 10^{-9}$  por nucleotídeo por ano, em eucariotos. A probabilidade de dois eventos mutacionais independentes acontecerem no mesmo sítio é muito inferior em relação as taxas de mutação citadas anteriormente. A característica bi-alélica facilita a genotipagem, que dependendo do caso pode ser efetuada com apenas com uma simples amplificação por PCR seguida de restrição enzimática e eletroforese em gel de agarose (VIGNAL et al. 2002).

Alguns autores consideram que a presença de inserções ou deleções de apenas uma base nitrogenada também pode ser classificada como um SNP. Porém a base genética geradora deste tipo de polimorfismo é totalmente diferente das transições e transversões de bases (JEHAN; LAKHANPAUL, 2006). Em sequências transcritas, os SNP podem ocorrer nas regiões não traduzidas denominadas de UTR (*UnTranslated Region*). As UTRs ocorrem nas duas extremidades dos transcritos, a extremidade 5' e extremidade 3'. Os SNPs também podem ocorrer dentro da fase de leitura (ORF - *Open Reading Frame*), sendo que funcionalmente estas mutações podem ou não traduzir para o mesmo aminoácido, dependendo da redundância do código genético. A frequência dos SNPs é maior nas regiões não traduzidas, pois estas apresentam menor conservação (VIGNAL et al. 2002).

Os marcadores SSR são também popularmente conhecidos como marcadores microssatélites, em referência aos experimentos clássicos envolvendo a separação do DNA repetitivo (DNA satélite) em gradientes de centrifugação. *Variable number of tandem repeats* (VNTR) e *Short*

*tandem repeats* (STR) são outras nomenclaturas utilizadas para o mesmo tipo de marcador e as que mais se aproximam da base genética destas sequências. Nos SSR, sequências curtas (motivos possuindo geralmente de 1 a 10 bases) se repetem diversas vezes em tandem, sendo flanqueadas por regiões conservadas (LI et al. 2002).

O polimorfismo genético neste marcador (diferentes formas alélicas para o mesmo loco) é baseado nos diferentes números de repetições de determinado motivo. O principal mecanismo gerador deste polimorfismo é o processo de *Slippage*, caracterizado por erros da enzima DNA polimerase durante a replicação do DNA (CHOUDHARY; TRIVEDI, 2010; WANG et al. 2009a). Os SSR possuem alto polimorfismo e conteúdo informativo (PIC - Polymorphism Information Content) (WANG et al. 2009a). Em *A. angustifolia* já foram reportados a presença de 14 alelos diferentes, para o mesmo marcador, em apenas 16 indivíduos amostrados (SCHMIDT et al. 2007). As marcas SSR, assim com as SNP, também possuem herança co-dominante. As utilizações são diversas, como mapeamento genético, identificação de QTLs, seleção assistida por marcadores (MAS), DNA *fingerprinting*, teste de paternidade, análise de diversidade genética, fluxo gênico, filogeografia, entre outros (WANG et al. 2009a).

Os marcadores microssatélites possuem ampla distribuição nos genomas de eucariotos. Eles estão presentes tanto em sequências codificantes como em regiões não codificantes, tais como *introns* e espaçadores inter-gênicos. Porém, o polimorfismo encontrado nos SSR das regiões codificantes é menor do que em relação ao polimorfismo de marcas não codificantes, em função da baixa conservação e pressão de seleção sofrida pelas regiões não-codificantes. SSRs encontrados em sequências codificantes são denominados de SSR gênicos, ou EST-SSR por serem derivados de sequências expressas. Os motivos presentes em sequências não codificantes são denominados de SSR genômicos (WANG et al. 2009a). SSRs gênicos são facilmente transferidos para outras espécies do mesmo gênero em função da alta conservação destas sequências transcritas (VARSHNEY et al. 2005).

A maioria dos marcadores SSR e outras sequências de natureza repetitiva vem sendo consideradas marcas neutras dentro dos genomas. No entanto, a neutralidade de uma marca só pode ser confirmada com a presença de um mapa físico, de ligação ou da própria sequência completa. Estas sequências somadas a outros elementos genéticos como *introns*, transposons e pseudogenes foram denominadas de DNA lixo (*Junk DNA*), por se acreditar que não possuíam nenhuma função (PALAZZO;

GREGORY, 2014). Entretanto, diversas funções vêm sendo atribuídas às sequências de natureza repetitiva, como a organização da cromatina em termos de manutenção da estrutura do DNA, centrômeros, telômeros e cromossomos. Outras funções estão relacionadas a regulação gênica e também a regulação de processos do próprio DNA, como replicação, ciclo celular e recombinação (LI et al. 2002).

Os SSRs também estão presentes no genoma plastidial (cpSSR) e mitocondrial (mtSSR), sendo de grande importância em estudos evolutivos, filogeográficos e de fluxo gênico com base na dispersão de pólen e sementes, em função da herança uni-parental, ausência de recombinação (permuta - *crossing-over*) e baixa taxa de mutação destas organelas (PROVAN et al. 2001). A caracterização e validação de marcas SSR é de extrema importância para espécies nativas pouco estudadas. Ainda não existem marcadores microssatélites genômicos específicos para *P. lambertii*. A espécie conta apenas com 156 cpSSRs identificados por Vieira et al. (2014), mas que ainda precisam ser validados em populações naturais da espécie.

Por sua vez, *A. angustifolia* possui apenas 35 marcadores SSRs específicos caracterizados e validados (SALGUEIRO et al. 2005; SCHMIDT et al. 2007). Outros 10 marcadores foram identificados, suas sequências foram depositadas na base de nucleotídeos do NCBI, mas as marcas não chegaram a ser validadas em populações da espécie e publicadas como *Primer Note* (PATREZE et al. 2008, dados não publicados). Apesar deste bom número de marcas disponíveis, os estudos de diversidade genética da espécie utilizam apenas um painel contendo os mesmos 8-10 marcadores, em função do baixo polimorfismo revelado pelas marcas restantes (BITTENCOURT; SEBBENN, 2007; PATREZE; TSAI, 2010; CRISTOFOLINI 2013; SANT'ANNA et al. 2013). O genoma plastidial da espécie já foi sequenciado e anotado pelo mesmo grupo responsável pelo sequenciamento do genoma plastidial de *P. lambertii*, assim, novos cpSSR estarão disponíveis para *A. angustifolia* (VIEIRA, L. DO N., 2016), dados não publicados.

### 3 QUESTÕES ORIENTADORAS

O sequenciamento dos transcriptomas de amostras obtidas em uma das áreas de ocorrência natural das espécies auxiliará na identificação do maior número possível de genes e isoformas para as duas espécies?

As três amostras sequenciadas para *A. angustifolia* serão suficientes para a detecção da totalidade de genes e isoformas, baseando-se nos genomas de coníferas já sequenciadas?

Apesar da curta distância evolutiva entre os gêneros *Podocarpus* e *Araucaria* (Figura 1), os transcriptomas de folhas das duas espécies apresentarão grandes diferenças entre si?

Novos genes e isoformas, ainda não caracterizados para espécies de coníferas, serão identificados nos transcriptomas de *Araucaria angustifolia* e *Podocarpus lambertii*?

Será possível a identificação e caracterização de marcas moleculares SSRs e SNPs consistentes e polimórficas a partir de dados de sequências transcritas?

## 4 OBJETIVOS

### 4.1 Objetivos gerais

Caracterizar os transcriptomas de amostras de *Araucaria angustifolia* e *Podocarpus lambertii*, visando a compreensão dos perfis de expressão gênica e identificação de marcadores moleculares SSRs e SNPs.

### 4.2 Objetivos específicos

4.2.1 Caracterizar os genes transcritos, tamanho de transcriptomas e transcritos por amostras de *A. angustifolia* e *P. lambertii*;

4.2.2 Caracterizar genes ainda não identificados em coníferas e inexistentes em angiospermas;

4.2.3 Identificar e caracterizar *in silico* marcadores moleculares SSRs e SNPs a partir dos dados de RNA-Seq em *A. angustifolia* e *P. lambertii*;

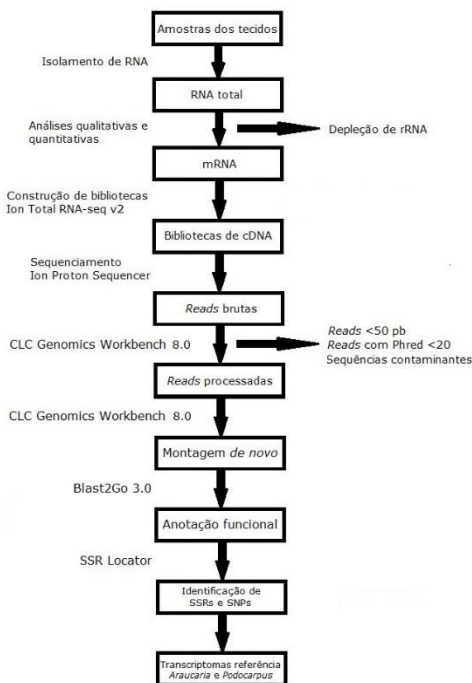
4.2.4 Avaliar os transcriptomas obtidos em relação aos transcriptomas referência de angiospermas e ao disponibilizado para *A. angustifolia* (ELBL, 2014) e para outras coníferas tais como *Picea abies* e *Pinus pinaster*.



## 5 MATERIAL E MÉTODOS

O esquema abaixo resume todas as etapas realizadas para a obtenção de transcriptomas vegetais e mineração de marcadores moleculares com a utilização da plataforma de sequenciamento *Ion Proton™ Semiconductor Sequencer* (Thermo Fisher Scientific) (Figura 8). Tanto nas variadas aplicações de *Next Generation Sequencing*, como em RNA-Seq, diversas metodologias, kits e equipamentos podem ser empregados para o sequenciamento completo dos transcritos de determinada amostra. A descrição a seguir é detalhada visando a reprodutibilidade futura.

Figura 8. Representação esquemática do fluxo de protocolos para a obtenção dos transcriptomas, anotação funcional e mineração de marcadores moleculares em *A. angustifolia* e *P. lambertii*



Fonte: Autor

## 5.1 Área de estudo e coleta das amostras

De acordo com o mapa fitogeográfico de Klein (1978), a área de estudo é classificada como: Campos com capões, florestas ciliares e bosques de pinheiros com predominância de ervas (Gramíneas, Ciperáceas, Leguminosas e Compostas). A região apresenta temperatura média anual inferior a 18 °C, ocorrência de geadas, altitude por volta dos 1.000 m e precipitação média anual de 1.600 mm. O clima é o Cfb de Köppen, mesotérmico subtropical úmido sem estação seca e apresentando verões frescos (BERNARDI, 2015).

As amostras foliares e de hastes foram coletadas em um fragmento de floresta ciliar de uma propriedade particular na localidade Coxilha Rica, Lages/Santa Catarina, durante os dias 09 a 11 de junho de 2014, com exceção dos estróbilos masculinos que foram coletados posteriormente, (Figura 9, Tabela 1). As coletas foram realizadas sob prévia autorização do proprietário, Dr. José Antônio Ribas Ribeiro. A propriedade foi adquirida em 1954, sendo o último corte de araucárias realizado no ano de 1971. A área é caracterizada pela paisagem de campo com criação de gado, presença de capões e mata ciliar nas margens do rio Pelotinhas, pertencente à bacia hidrográfica do rio Uruguai. A presença do gado é constante nas áreas de mata, interferindo principalmente na regeneração e desenvolvimento das espécies.

Figura 9. Imagem de satélite da área de coleta. Mata ciliar nas margens do rio Pelotinhas. Propriedade do Dr. José Antônio Ribas Ribeiro, localidade da Coxilha Rica / Lages / SC



Fonte: Google Earth 2015

As coletas, das duas plantas masculinas de *A. angustifolia* e da planta masculina de *P. lambertii*, foram realizadas em triplicatas por se tratarem de amostras ambientais, mesmo assim podem ocorrer variações significativas nos níveis de transcrição entre as replicatas biológicas da mesma amostra (McINTYRE et al. 2011). O horário de coleta das amostras, nos três dias de coletas foi entre 10 h e 11 h. Esse critério foi utilizado em função de qualquer perfil transcricional ser altamente dinâmico e dependente de fatores ambientais como temperatura, umidade relativa do ar, regime hídrico do solo, luminosidade, interação com micro-organismos, entre outros fatores.

Tabela 1. Coordenadas geográficas das coletas de amostras para RNA-Seq

<b>Amostra</b>	<b>Latitude / Longitude</b>
<b>Folhas - <i>Podocarpus lambertii</i> ♂</b>	28° 02' 49'' S / 50° 17' 47'' O
<b>Folhas e hastes - <i>Araucaria angustifolia</i> ♂</b>	28° 02' 50'' S / 50° 17' 50'' O
<b>Pólen - <i>Araucaria angustifolia</i> ♂</b>	28° 02' 52'' S / 50° 17' 47'' O

Fonte: Autor

Foram utilizadas luvas nitrílicas descartáveis, pinças e tesouras (lavadas em água DEPC 0,1%, queimadas em estufa a 220 °C por 4 h e autoclavadas) e tubos do tipo *falcon* livres de DNases e RNases. As amostras coletadas foram acondicionadas nos tubos, sendo adicionado três volumes de reagente de estabilização de RNA, RNAlater® (Thermo Fisher Scientific). Após fechados, os tubos foram vedados com a utilização da película de parafina Parafilm® (Pechiney, Chicago, IL, EUA) e mantidos a 4 °C até o isolamento de RNA. O armazenamento das amostras a 4 °C foi possível pois o reagente de estabilização inativa as RNases intra e extra-celulares, permitindo o armazenamento de tecidos, com baixíssima degradação de RNA, em temperatura de refrigeração por até 30 dias.

A coleta dos estróbilos masculinos maduros foi realizada entre os dias 10 e 12 de novembro de 2014 na mesma localidade descrita anteriormente, mas de uma planta masculina diferente da coleta de folhas e hastes. Foi coletado um estróbilo por dia, de aproximadamente 15 cm, entre 10h e 11h AM. As estruturas reprodutivas foram mantidas em sacos plásticos (Tipo *ziplock*) e armazenadas a 4 °C do momento da coleta até a chegada ao laboratório. Os grãos de pólen foram retirados dos

estróbilos, acondicionados em tubos tipo “falcon” de 15 ml e armazenados a -80 °C até o isolamento de RNA. O pólen maduro é retirado com facilidade dos estróbilos após leves batidas. Nesta etapa, foi tomado o cuidado para evitar que impurezas provenientes dos estróbilos fossem adicionadas aos grãos retirados (Figura 10, D). A estrutura dos grãos de pólen proporciona estabilidade ao RNA total, em função disso as amostras foram transportadas a 4 °C (TWELL et al. 2006). As distintas datas de coletas se deram em função da maturação do pólen ocorrer apenas em novembro, sendo que a data de coleta das amostras foliares e de hastes foi o período de deiscência das sementes (pinhões) de *A. angustifolia*.

## 5.2 Amostras

As amostras selecionadas de *A. angustifolia* para a rodada de RNA-Seq foram folhas (acículas), hastes e pólen, enquanto que de *P. lambertii* foram analisadas amostras de folhas. Foram selecionadas folhas medido aproximadamente 2 cm para *A. angustifolia* e 5 cm para *P. lambertii*, e sem a presença visível de sintomas de patógenos. As folhas de *A. angustifolia* (Figura 10, C) são aciculares, coriáceas, glabras e possuem de 3 cm a 6 cm de comprimento. As acículas possuem baixa área superficial específica e espessa cutícula, dificultando a perda de água. A vascularização é formada por feixes de xilema na face superior (adaxial) e floema na face inferior (abaxial) das acículas, além da presença de canais resiníferos (LORENZI, 2008).

A amostra foliar de *Podocarpus* foi incluída visando a comparação dos perfis de transcrição de duas coníferas nativas e com maior distribuição no Brasil. As folhas de *P. lambertii* (Figura 10, A) são caducas, alternas, lineares a estreito-elípticas, coriáceas, de margens revolutas, base atenuada e possuindo de 2 cm a 6 cm de comprimento (LORENZI, 2008).

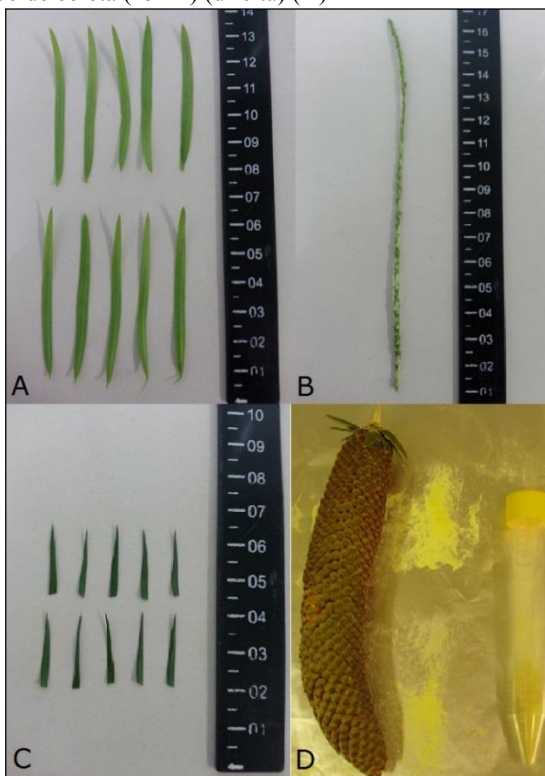
O genoma de cpDNA expressa aproximadamente 50 genes relacionados a fotossíntese. Em função disso os plastomas completos de *Araucaria* e *Podocarpus* foram utilizados como referência de alinhamento e montagem (GREEN, 2011) (Item 5.10). Os outros transcritos relacionados a fotossíntese são expressos pelo genoma nuclear, que recebeu diversas transferências de material genético plastidial durante a evolução das plantas superiores (MARTIN et al. 1998).

A amostra de hastes de *A. angustifolia* (Figura 10, B), tinha comprimento médio de 15 cm e sem a presença das acículas. As hastes

possuem tecidos condutores como xilema e floema, são fotossintetizantes e possuem tecidos de revestimento como córtex e periderme. É uma estrutura flexível e fina, dependendo do tamanho e idade.

A última amostra da rodada de sequenciamento foi composta por grãos de pólen de *A. angustifolia*, isolados de estróbilos maduros (Figura 10, D). Esta amostra foi incluída pois espera-se encontrar importantes informações relacionadas ao metabolismo basal (*housekeeping genes*) dos grãos de pólen e a sinalização celular envolvida antes da germinação do tubo polínico.

Figura 10. Folhas de *P. lambertii* (A); Haste de *A. angustifolia* (B); Acículas de *A. angustifolia* (C); Estróbilo masculino de *A. angustifolia* (esquerda), pólen (centro) e tubo de coleta (15 ml) (direita) (D)



Fonte: Autor

Ao total, foram obtidas 12 bibliotecas de cDNA (4 amostras x 3

replicatas biológicas cada). Cada biblioteca de cDNA foi sequenciada 12 vezes (Item 5.8). Após todas as corridas de sequenciamento, as leituras foram agrupadas em um único arquivo por biblioteca.

### **5.3 Isolamento de RNA total**

Diversas variações dos protocolos de isolamento de RNA total descritos a seguir foram testadas em função das distintas características das amostras coletadas, como conteúdo de lignina, água, proteínas, compostos fenólicos, entre outros.

Todas as vidrarias utilizadas no preparo das soluções de isolamento, cadinhos, pistilos e espátulas foram lavadas com água, detergente neutro, água destilada, água DEPC 0,1%, incubadas em estufa a 220 °C e posteriormente autoclavadas. Todas as superfícies de trabalho (capela de exaustão de gases, bancadas, pipetas rotor de centrífuga, agitador e vórtex) foram limpas com peróxido de hidrogênio 3% e com a solução de descontaminação de RNases RNaseZAP® (Thermo Fisher Scientific) 20 min antes do uso.

Todo o trabalho de isolamento de RNA foi realizado em sala exclusiva para análise de RNA (com presença de ante-sala), localizada no Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal (LFDGV). A limpeza do chão e remoção de resíduos, com instrumentos exclusivos da sala de RNA, foi realizada sempre antes e depois de cada atividade. O aparelho de ar condicionado permaneceu desligado durante todas as etapas descritas a seguir, visando minimizar os riscos de contaminação advindos do ambiente (aerossóis). Durante todas as atividades foram utilizados jalecos (lavados a cada uso) e toucas, luvas e máscaras descartáveis.

#### **5.3.1 CTAB e Cloreto de lítio**

O RNA total das hastes de Araucaria e folhas de Araucaria e Podocarpus foi isolado, baseando-se no protocolo proposto por Lorenz, Yu e Dean (2010), com modificações. No início de cada isolamento, as amostras foram lavadas em água DEPC (Apêndice A) visando retirar o excesso do reagente de estabilização RNAlater®. Imediatamente após isso, 4 g de amostra foram macerados em nitrogênio líquido até a formação de um pó fino. Com o auxílio de espátula, o material macerado foi transferido para tubos tipo falcon contendo 20 ml de tampão de isolamento (Apêndice B) pré-aquecido a 65 °C. Logo após, os tubos

foram incubados em banho-maria a 65 °C por 15 min, com leve agitação a cada 3 min. A extração orgânica foi realizada com a adição de 1 volume de clorofórmio. Os tubos foram agitados a 90 RPM por 10 min e centrifugados em seguida a 7.800 xg por 10 min.

A fase aquosa (sobrenadante) foi transferida para novo tubo de 50 ml e em seguida adicionado ½ volume de clorofórmio. Os tubos foram agitados a 90 RPM por 5 min e centrifugados em seguida a 7.800 xg por 10 min. A fase aquosa foi novamente transferida para um novo tubo e novamente foi adicionado ½ volume de clorofórmio. Os tubos foram brevemente agitados em vórtex e centrifugados em seguida a 7.800 xg por 15 min. Por fim, a fase aquosa foi transferida para novo tubo de 50 ml. Para a precipitação do RNA, foi adicionado 1 volume de cloreto de lítio 4 M. O RNA total precipitou durante 12 h a 4 °C, sendo coletado por centrifugação a 7.800 xg por 50 min a 4 °C. Neste protocolo, o cloreto de lítio possui por função precipitar o RNA total sem precipitar o DNA total, que é considerado um contaminante. Contudo, por segurança, o DNA total é digerido ao final do isolamento de RNA.

O sobrenadante foi descartado e o *pellet* de RNA foi ressuscitado em 800 µl de tampão SSTE a 60 °C (Apêndice B). A solução foi transferida para microtubo de 2 ml e foram adicionados 800 µl de clorofórmio. Os tubos foram brevemente agitados em vórtex e centrifugados a 13.000 xg por 5 min. A fase aquosa foi transferida para um microtubo de 1,5 ml onde foram adicionados 1 ml de etanol absoluto e 50 µl de acetato de sódio 3 M. Os microtubos foram mantidos em freezer -80 °C durante 30 min para a precipitação do RNA total. Após isso, os tubos foram centrifugados a 13.000 xg por 5 min. O *pellet* de RNA foi brevemente lavado com etanol 75%, seco em temperatura ambiente e em seguida ressuscitado em 20 µl de água livre de DNase e RNase (Ambion). O DNA total contaminante foi digerido com a utilização da enzima Turbo™ DNase (Ambion). A reação de digestão foi realizada em um volume de 50 µl, contendo 1 µl de enzima (2 U), 5 µl de tampão enzimático 10 X, 40 µl da amostra de RNA total e água ultra-pura (q.s.p.). As reações foram mantidas a 37 °C durante duas horas e em seguida foram adicionados 5 µl de tampão de inativação. Os tubos foram agitados brevemente em vórtex e mantidos em temperatura ambiente por 10 min, com agitação leve a cada 2 min. Em seguida, as reações inativadas foram centrifugadas a 11.000 xg por 10 min. O sobrenadante contendo o RNA total isolado foi transferido para outro microtubo e mantido a -80 °C até o momento das análises posteriores.

### **5.3.2 Kit Nucleospin RNA Plant®**

O RNA total das amostras de grãos de pólen foi isolado com a utilização do kit Nucleospin Plant RNA (Macherey-Nagel, Düren, Alemanha). Cerca de 100 mg de pólen foram rapidamente submersos em nitrogênio líquido visando a formação de cristais de gelo e ruptura das estruturas celulares dos grãos de pólen. Após a total evaporação do nitrogênio líquido o material foi transferido para um microtubo contendo  $\beta$ -mercaptoetanol (1%) e tampão de lise celular. As demais etapas foram baseadas nas recomendações do fabricante, com modificações, e estão descritas no Anexo A.

### **5.4 Quantificação e verificação da qualidade do RNA**

A partir desta etapa, todos os procedimentos laboratoriais foram realizados no laboratório de sequenciamento de ácidos nucleicos do Núcleo de Fixação de Nitrogênio (NFN), localizado no departamento de Bioquímica, do Centro Politécnico da Universidade Federal do Paraná. As amostras foram transportadas em solução, de Florianópolis até Curitiba, congeladas em gelo seco (-78,5 °C).

A qualidade do RNA isolado é de extrema importância para o sucesso das bibliotecas de cDNA que serão sequenciadas em qualquer plataforma de sequenciamento de DNA de segunda geração. Em função disso, todas as amostras foram submetidas a diversas análises, visando a garantia de um bom RNA total inicial, assim minimizando possíveis problemas futuros relacionados ao preparo das bibliotecas de cDNA, sequenciamento e análise de dados.

Inicialmente, amostras de RNA isoladas foram submetidas à eletroforese em gel desnaturante de agarose 1,0% (Apêndice C), visando detectar a qualidade das bandas de rRNA 18S e 25S. Estes rRNA são os mais intensamente expressos em amostras vegetais, logo a qualidade das bandas reflete a qualidade geral da amostra. O gel também pode revelar a presença não adventícia de DNA nuclear, extra-nuclear (mitocondrial e plastidial), contaminação por polissacarídeos e degradação de mRNA.

A presença de contaminantes, principalmente proteínas e compostos fenólicos, nas amostras de RNA total, foi verificada com a utilização do espectrofotômetro NanoDrop 2000 (Thermo Fischer Scientific, Waltham, EUA). Foi aplicado 1  $\mu$ l de RNA total por amostra e



a leitura foi realizada com base no parâmetro RNA-40. Foram consideradas amostras de RNA total sem a presença de contaminantes amostras com os valores das relações 260/280 e 260/230 entre 1,8 e 2,2, respectivamente.

O RNA total foi quantificado com a utilização do fluorímetro Qubit™ 1.0 (Thermo Fisher Scientific), com o emprego do kit Qubit™ RNA assay (Thermo Fisher Scientific). Foi formado um mix contendo 1 µl do Componente A (fluoróforo) e 199 µl do Componente B (tampão), por amostra. Para a quantificação no aparelho, foram pipetados 199 µl do mix e adicionado 1 µl da amostra a ser quantificada. Os tubos foram brevemente agitados em vórtex, incubados em temperatura ambiente por dois minutos e em seguida foi realizada a calibração do equipamento (quantificação dos padrões) e quantificação das amostras.

A integridade das amostras foi medida com a utilização do equipamento 2100 Bioanalyzer (Agilent Technologies, Wilmington, EUA), com o uso do kit Agilent RNA 6000 Pico (Agilent Technologies) e Agilent RNA 6000 Pico Ladder (Agilent Technologies). A classe de ensaio utilizada foi a RNA total de eucariotos. Cada amostra recebeu um valor entre 0 e 10, onde 10 representa a amostra 100% íntegra. Este valor é denominado de RIN (*RNA Integrity Number*) e foi proposto por Schroeder et al. (2006). Foram consideradas íntegras as amostras com RIN acima de 6,5. Outro indicador de integridade analisado foi a relação 28S/18S. Esta relação é baseada no tamanho em kb das moléculas de rRNA, que é aproximadamente de 5 kb e 2 kb. A divisão teórica entre as quantidades desses rRNAs é de 2,7:1, mas a relação 2:1 sempre foi considerada como padrão ouro de integridade de mRNA. Embora essa relação tenha sido criada para a análise de mRNA de mamíferos (possuidor de rRNA 28S), o equipamento Agilent 2100 Bioanalyzer estima a quantidade do rRNA 25S quando a análise é realizada para RNA total de eucariotos.

## **5.5 Precipitação de RNA total**

As amostras apresentando baixas relações de espectrofotometria (relações 260/280 e 260/230 abaixo de 1,8) foram precipitadas visando a remoção de contaminantes. A precipitação do RNA total foi realizada com a adição de etanol absoluto (3X o volume de RNA total), 5 µl de acetato de sódio (3 M) e 1 µl de glicogênio (20 µg/µl). As soluções foram brevemente agitadas em vórtex e deixadas a -20 °C durante 12 h para a precipitação dos ácidos nucleicos.

Após esse período, as amostras foram centrifugadas a 11.000 xg por 30 min visando a concentração do RNA total em um único *pellet*. Em seguida, o sobrenadante foi descartado e foram adicionados 200 µl de etanol 70% por 1 min. As amostras foram novamente centrifugadas a 11.000 xg por 5 min, seguido pela remoção do sobrenadante. Os tubos foram secos em bomba de vácuo por 30 min para a remoção do etanol residual. A ressuspensão foi realizada em 20 µl de água ultra-pura (livre de RNase e DNase). As amostras precipitadas foram novamente analisadas em espectrofotômetro (NanoDrop 2000) para a detecção de contaminantes. O protocolo de precipitação do RNA total é novamente efetuado caso as relações de absorvância se mantenham insatisfatórias. Porém, o RIN das amostras pode diminuir a medida em que aumenta a manipulação do RNA total.

## **5.6 Depleção de RNA**

A depleção do rRNA foi realizada com a utilização do kit Ribominus Plant Kit for RNA-Seq (Invitrogen, Carlsbad, EUA), com utilização do rack magnético MagnaRack™ (Thermo Fisher Scientific). Neste procedimento, foram efetuadas duas rodadas de depleção de rRNA (Anexo B).

A integridade e quantidade do mRNA enriquecido foram avaliadas com a utilização do equipamento 2100 Bioanalyzer (Agilent Technologies, Wilmington, EUA), com o uso do kit Agilent RNA 6000 Pico (Agilent Technologies) e Agilent RNA 6000 Pico Ladder (Agilent Technologies). A classe de ensaio utilizada foi a RNA total de eucariotos.

## **5.7 Construção das bibliotecas de cDNA**

As 12 bibliotecas de cDNA foram construídas com a utilização do kit Ion Total RNA-Seq v.2 (Thermo Fisher Scientific). Este kit permite a construção de bibliotecas de cDNA, e posterior, sequenciamento de todo o RNA presente nas amostras (Total RNA-Seq - Sequenciamento dos RNAs totais, incluindo RNAs indesejáveis, como os ribossomais) (Anexo C). Por isso, a depleção realizada na etapa anterior é de vital importância, pois a cobertura de sequenciamento dos mRNA pouco expressos aumenta significativamente a medida em que a depleção ribossomal é bem-sucedida (O'Neil et al. 2013). Contudo, ainda existe o risco do sequenciamento de rRNAs residuais. No presente trabalho foram efetuadas duas rodadas de depleção dentro do mesmo kit.

### 5.7.1 Fragmentação do RNA

O mRNA foi fragmentado com a utilização da enzima RNase III. Para cada replicata biológica, foi formado um *mix* contendo 1 µl da enzima RNase III, 1 µl de tampão de reação 10X e 8 µl de mRNA. O tubo foi levemente agitado em vórtex e em seguida centrifugado (*spin*). A reação de digestão foi realizada em termociclador (Veriti Thermal Cycler, Thermo Fisher Scientific) a 37 °C durante 10 min. Em seguida, os tubos foram colocados em gelo e foram adicionados 20 µl de água nuclease *free* a cada replicata.

### 5.7.2 Purificação do RNA fragmentado

O tubo contendo as *beads* magnéticas foi agitado manualmente visando a ressuspensão das *beads*. Em seguida, foram pipetados 5 µl de *beads* e 90 µl da solução concentrada de ligação (*binding solution*) em 12 poços da placa de processamento. A solução foi misturada com o auxílio de uma pipeta P100 ajustada para 50 µl. Após isso, 30 µl de mRNA fragmentado foram transferidos para a placa de processamento (uma replicata por poço da placa).

Foram adicionados 150 µl de etanol absoluto a cada um dos 12 poços da placa de processamento. A solução foi misturada com o auxílio de uma pipeta P200 ajustada para 150 µl e ficou em temperatura ambiente por 5 min visando a ligação magnética com as *beads*. Após isso, a placa de processamento foi transferida para o *rack* magnético, onde permaneceu por 6 min para separação das *beads*. O sobrenadante foi aspirado e descartado com o auxílio de uma pipeta P100 ajustada para 50 µl. Em seguida, foram adicionados 150 µl da solução concentrada de lavagem (*Wash solution*). As amostras foram incubadas em temperatura ambiente por 30 s e após isso o sobrenadante foi novamente aspirado e descartado. A placa de processamento permaneceu em temperatura ambiente por mais 5 min para a evaporação do etanol residual.

Para a eluição do RNA fragmentado, foram adicionados 12 µl de água nuclease *free* pré-aquecida a 37 °C. As *beads* foram misturadas com a água com auxílio de pipeta P10 µl. A placa de processamento permaneceu em temperatura ambiente por 1 min e depois mais 1 min no *rack* magnético. O sobrenadante contendo o RNA fragmentado e purificado foi recolhido e acondicionado em um novo microtubo.

### **5.7.3 Avaliação da qualidade e distribuição de tamanho do RNA fragmentado**

O RNA fragmentado foi quantificado com a utilização do fluorímetro Qubit™ 1.0 (Thermo Fisher Scientific), com o emprego do kit Qubit™ RNA assay (Thermo Fisher Scientific) como descrito anteriormente. A distribuição de tamanho foi realizada com a utilização do equipamento 2100 Bioanalyzer (Agilent Technologies, Wilmington, EUA), com o uso do kit Agilent RNA 6000 Pico (Agilent Technologies) e Agilent RNA 6000 Pico Ladder (Agilent Technologies). A classe de ensaio utilizada foi a RNA total de eucariotos. O processo de fragmentação produz fragmentos de 35 pb até milhares de pb, dependendo da amostra. A faixa ideal tamanho dos fragmentos de mRNA é de 100 pb a 200 pb.

### **5.7.4 Hibridização e ligação do RNA**

O *mix* de hibridização foi preparado, em gelo, da seguinte forma: 13 (nº de replicatas + margem de segurança de 8,33%) x 2 µl de Ion Adaptor Mix v2 + 3 µl de Hybridization Solution, totalizando 5 µl. Em seguida, os 5 µl do *mix* de hibridização foram pipetados em um novo microtubo contendo 3 µl da amostra de RNA fragmentado. A nova solução foi brevemente misturada com o auxílio de pipeta P10 e brevemente centrifugada (*spin*). Os tubos foram colocados em termociclador (Veriti Thermal Cycler) onde foi realizada a hibridização do RNA fragmentado com os adaptadores, com as seguintes condições de tempo e temperatura: 65 °C por 10 min, seguido de 30 °C por 5 min.

Após a hibridização, os reagentes de ligação foram adicionados aos tubos já contendo 8 µl da reação de hibridização. Foram adicionados 10 µl do tampão de ligação 2X e 2 µl da enzima de ligação, totalizando 20 µl da nova reação. A nova reação foi brevemente misturada com o auxílio de pipeta P10 e brevemente centrifugada (*spin*). As reações de ligação foram realizadas em termociclador (Veriti Thermal Cycler) a 30 °C durante 30 min.

### **5.7.5 Transcrição reversa (RT)**

O *mix* para a reação de transcrição reversa foi preparado, em gelo, da seguinte forma: 13 (nº de replicatas + margem de segurança de 8,33%) x 2 µl de água nuclease *free* + 4 µl de tampão 10X RT + 2 µl de dNTP *mix*

(2,5 mM) + 8 µl de primer RT v2, totalizando 16 µl de *mastermix*.

Em seguida, 20 µl da reação de ligação foram adicionados aos 16 µl de *mastermix*. Os tubos foram agitados em vórtex e brevemente centrifugados (*spin*). Os tubos foram incubados a 70 °C por 10 min e em seguida resfriados em gelo. Foram adicionados 4 µl do *mix* da enzima 10X SuperScript®III. Os tubos foram brevemente agitados e foram incubados em termociclador (Veriti Thermal Cycler) a 42 °C durante 30 min.

### 5.7.6 Purificação do cDNA

O cDNA foi purificado visando a remoção de mRNA residual, dNTP e *primers*. O tubo contendo as *beads* magnéticas foi agitado manualmente visando a ressuspensão das *beads*. Em seguida, foram pipetados 5 µl de *beads* e 120 µl da solução concentrada de ligação (*binding solution*) em 12 poços da placa de processamento. A solução foi misturada com o auxílio de uma pipeta P100 ajustada para 50 µl. Em seguida, 60 µl de água nuclease *free* foram adicionados e misturados aos 40 µl da reação RT. Estes 100 µl foram então transferidos para a placa de processamento.

Foram adicionados 125 µl de etanol absoluto a cada um dos 12 poços da placa de processamento. A solução foi misturada com o auxílio de uma pipeta P200 ajustada para 150 µl e ficou em temperatura ambiente por 5 min visando a ligação magnética com as *beads*. Após isso, a placa de processamento foi transferida para o *rack* magnético, onde permaneceu por 6 min para separação das *beads*. O sobrenadante foi aspirado e descartado com o auxílio de uma pipeta P100 ajustada para 50 µl. Em seguida, foram adicionados 150 µl da solução concentrada de lavagem (*Wash solution*). As amostras foram incubadas em temperatura ambiente por 30 segundos e após isso o sobrenadante foi novamente aspirado e descartado. A placa de processamento permaneceu em temperatura ambiente por mais 5 min para a evaporação do etanol residual.

Para a eluição do cDNA, foram adicionados 12 µl de água nuclease *free* pré-aquecida a 37 °C. As *beads* foram misturadas com a água com auxílio de pipeta P10. A placa de processamento permaneceu em temperatura ambiente por 1 min e depois mais 1 min no *rack* magnético. O sobrenadante contendo o cDNA purificado foi recolhido e acondicionado em um novo microtubo.

### 5.7.7 Amplificação e purificação do cDNA

As amostras purificadas foram amplificadas em PCR, sem a adição de *barcodes*, com a utilização do seguinte *mix*: 13 (n° de replicatas + margem de segurança de 8,33%) x 45 µl do pré *mix* Platinum PCR Super Mix High Fidelity + 1 µl do primer Ion 5' PCR v2 + 1 µl do *primer* Ion 3' PCR v2. Em um novo microtubo de PCR, foram adicionados os 47 µl do *mix* de amplificação e 6 µl do cDNA amplificado no passo anterior. A solução foi suavemente agitada e centrifugada (spin).

Em termociclador (Veriti Thermal Cycler), o cDNA foi amplificado com a utilização da seguinte ciclagem: desnaturação inicial a 94 °C por 2 min, 2 ciclos contendo: 94 °C por 30 s, 50 °C por 30 s e 68 °C por 30 s. Em seguida foram realizados 16 ciclos contendo: 94 °C por 30 s, 62 °C por 30 s e 68 °C por 30 s. A reação foi finalizada com extensão final a 68 °C por 5 min. O cDNA amplificado foi purificado visando a remoção de dNTPs e *primers* residuais. O tubo contendo as *beads* magnéticas foi agitado manualmente visando a ressuspensão das *beads*. Em seguida, foram pipetados 5 µl de *beads* e 180 µl da solução concentrada de ligação (*binding solution*) em 12 poços da placa de processamento. Com o auxílio de uma pipeta P10 a solução foi misturada.

Em seguida, os 53 µl da reação de amplificação de cDNA foram transferidos para a placa de processamento. Foram adicionados 130 µl de etanol absoluto a cada um dos 12 poços da placa de processamento. A solução foi misturada com o auxílio de uma pipeta P200 ajustada para 150 µl e ficou em temperatura ambiente por 5 min visando a ligação magnética com as *beads*. Após isso, a placa de processamento foi transferida para o *rack* magnético, onde permaneceu por 6 min para separação das *beads*. O sobrenadante foi aspirado e descartado com o auxílio de uma pipeta P100.

Em seguida, foram adicionados 150 µl da solução concentrada de lavagem (*Wash solution*). As amostras foram incubadas em temperatura ambiente por 30 segundos e após isso o sobrenadante foi novamente aspirado e descartado. A placa de processamento permaneceu em temperatura ambiente por mais 5 min para a evaporação do etanol residual. Para a eluição do cDNA amplificado, foram adicionados 15 µl de água nuclease *free* pré-aquecida a 37 °C. As *beads* foram misturadas com a água com auxílio de pipeta P10. A placa de processamento permaneceu em temperatura ambiente por 1 min e depois mais 1 min no *rack* magnético. O sobrenadante contendo o cDNA amplificado purificado foi recolhido e acondicionado em um novo microtubo.

### **5.7.8 Avaliação da quantidade e distribuição de tamanho do cDNA amplificado**

O cDNA das bibliotecas foi quantificado com a utilização do equipamento 2100 Bioanalyzer, com o uso do kit High Sensitivity DNA (Agilent Technologies) e High Sensitivity DNA Ladder (Agilent Technologies). A classe de ensaio utilizada foi a High Sensitivity DNA. Foi quantificada a porcentagem de DNA de tamanho menor que 160 pb com a utilização da faixa de tamanho (*size range*) de 50 pb a 160 pb. A concentração molar (nM) foi calculada com a utilização da faixa de tamanho (*size range*) de 50 pb a 1.000 pb. A cada uma das 12 bibliotecas foi adicionado um *barcode* específico, o que possibilitou a corrida das 12 bibliotecas em um único *chip* semicontudor. Com base na concentração molar obtida, as bibliotecas de cDNA foram diluídas na concentração de 11 pM em tampão TE, fornecido pelo kit, para a reação de sequenciamento.

### **5.8 Sequenciamento de cDNA por *chips* semi-condutores**

As 12 bibliotecas de cDNA foram sequenciadas doze vezes cada (Ion Chip PI v.2, Thermo Fisher Scientific), na plataforma de sequenciamento de segunda geração Ion Proton™ Semiconductor Sequencer (Thermo Fisher Scientific). Assim, cada replicata biológica constituiu uma biblioteca de transcritos que foi sequenciada 12 vezes cada. Para cada biblioteca, as leituras oriundas das 12 corridas foram agrupadas em um único arquivo com formato .fastq.

As bibliotecas foram sequenciadas com a utilização dos *chips* de sequenciamento PI v.2 (Thermo Fisher Scientific). Cada *chip* de sequenciamento pode sequenciar até 80 milhões de leituras brutas por corrida. Mas esse número depende da qualidade do RNA e das bibliotecas de cDNA, da eficiência no carregamento das amostras no chip, entre outras variáveis.

As corridas de sequenciamento foram preparadas e realizadas pela Dra. Michelle Zibetti Tadra Sfeir, responsável pelo equipamento do Departamento de Bioquímica do Centro Politécnico da UFPR. O tempo médio de duração de cada uma das 12 corridas foi de aproximadamente 2 h. Os dados das corridas foram transferidos pelo servidor do sequenciador Torrent Suite™ (Thermo Fisher Scientific) para o servidor do Núcleo de Fixação de Nitrogênio (NFN) da UFPR ([www.genopar.org](http://www.genopar.org)).

## 5.9 *Trimming e montagem de novo*

Inicialmente foi realizado o descarte das leituras oriundas de *beads* policlonais e *barcodes*. Estas *beads* contêm mais de um *template* (fragmento de cDNA). O sequenciamento nas plataformas Ion Torrent PGM™ e Ion Proton™ já prevêem certa porcentagem de conteúdo policlonal de até 40% do total de leituras obtidas. A remoção destas leituras foi realizada com a utilização do *software* Torrent Suite (Thermo Fisher Scientific).

A remoção das leituras (*Trimming*) com baixa qualidade (Limite: 0.01), leituras com presença de mais de 2 bases ambíguas e leituras menores que 50 pb foi realizada com a utilização do *software* CLC Genomics Workbench versão 8.0. (<http://www.clcbio.com>).

A montagem *de novo* dos transcriptomas também foi realizada com a utilização do *software* CLC Genomics Workbench versão 8.0. Este *software* não é baseado em linha de comando (UNIX) e foi desenvolvido, entre outras aplicações, para a montagem de transcriptomas sem referência (método *de novo*) e com a utilização de leituras curtas não pareadas “*Unpaired sequences*”.

Os parâmetros para a montagem de novo foram os seguintes: Tamanho mínimo de sequências contíguas de 120 pb; *Word size* de 20 nucleotídeos e *Bubble size* de 50 nucleotídeos. As opções de mapeamento das leituras foram as seguintes: Custo de pareamento incorreto (*Mismatch*): 2; Custo de *indel*: 3; Fração de comprimento: 0.5; Fração de similaridade: 0.8. Após a montagem de novo, as sequências contíguas com cobertura acima de 5 foram extraídas e utilizados para a anotação funcional com a utilização do *software* Bast2Go (Item 4.11).

## 5.10 Mapeamento e análise referenciada (*Reference based*)

As análises referenciadas foram realizadas com a utilização do *software* CLC Genomics Workbench, versão 8.0. O genoma cloroplastidial de *P. lambertii*, GenBank: KJ010812, foi usado como referência para a montagem e anotação dos transcritos plastidiais da espécie na amostra de folha. O genoma referência de cpDNA apresenta 133.734 pb com 37,1% de conteúdo GC. São encontrados 118 genes, entre eles 31 genes codificando para tRNA, 58 genes codificando para proteínas e 20 sequências codificando para sub-unidades ribossomais (Vieira et al. 2014).



A sequência completa do genoma cloroplastidial de *A. angustifolia*, embora não anotada funcionalmente, foi gentilmente cedida por Vieira, L. do N.. Este genoma de 145.558 pb foi utilizado para a montagem referenciada das amostras de *Araucaria* (dados ainda não publicados).

Na montagem e análise dos transcritos mitocondriais das duas espécies, o genoma mitocondrial de *Cycas taitungensis*, GenBank: AP009381, foi tomado como referência (Chaw et al. 2008). O genoma mitocondrial de *C. taitungensis* é um dos poucos genomas de mtDNA sequenciados para gimnospermas, possuindo o tamanho total de 414.903 pb, 39 genes codificando para proteínas, 22 genes codificando tRNA e 3 genes de rRNA. Somadas, as sequências codificantes representam 10,1% do genoma mitocondrial. Além disso, o genoma apresenta 15,1% de sequências repetitivas e 25 introns.

Alternativamente, todos os transcriptomas também foram analisados de maneira referenciada contra o genoma completo de *Arabidopsis thaliana*. As leituras “*trimadas*” no item 5.09 foram utilizadas para a montagem referenciada. Estes, foram mapeados tanto nas regiões gênicas quanto nas regiões inter-gênicas. Os parâmetros de mapeamento foram: custo de *mismatch*: 2; custo de *indel*: 3; fração de tamanho: 0,8; fração de similaridade: 0,8. Além disso, as leituras foram mapeadas em ambas orientações do genoma de referência. Os valores de expressão de cada transcrito foram calculados com base no parâmetro RPKM (*Reads Per Kilobase per Million Mapped Reads*) (MORTAZAVI et al. 2008).

## 5.11 Análise funcional e caracterização dos transcriptomas

A anotação funcional proveniente da montagem de novo foi realizada com a utilização do *software* Blast2GO v. 3.0.9. (BioBam Bioinformatics, <http://www.biobam.com>, CONESA et al. 2005, GÖTZ et al. 2008). Inicialmente, foi realizado um *Blast* local, independente dos servidores do NCBI, com a utilização do algoritmo BLASTX (e-Value cut-off: 1.0 e-5 e devolução dos 20 melhores blast hits) das sequências contíguas montadas contra uma base de dados de sequências de proteínas não redundantes, com sequências oriundas de bases como GenPept, Swissprot, PIR, PDF, PDB e NCBI RefSeq, totalizando 22.690 MB de tamanho. Após a realização do BLASTX, as sequências contíguas foram mapeadas e anotadas funcionalmente de acordo com os termos do *Gene Ontology* (GOSlims), nas categorias de funções moleculares, processos

biológicos e componentes celulares (ASHBURNER et al. 2000). Figuras contendo as distribuições das anotações dos termos GO (level 2) foram construídos com a utilização da ferramenta *Web Gene Ontology Annotation Plot* (WEGO) (YE et al. 2006).

A análise e anotação funcional dos transcriptomas analisados de maneira referenciada foi toda efetuada com o *software* CLC Genomics Workbench, versão 8.0. Cada uma das quatro amostras formou um grupo, composto pelas suas respectivas replicatas biológicas. Os resultados do mapeamento, para cada referência, foram comparados grupo a grupo, sendo que o valor de expressão utilizado foi o RPKM. Inicialmente, os dados foram transformados na escala de log 10 e normalizados com base na totalidade dos valores.

### 5.12 Identificação de SNPs e SSRs putativos

A identificação dos SNPs putativos foi realizada com a utilização do *software* CLC Genomics Workbench, versão 8.0. A detecção das variantes foi realizada com a utilização das leituras mapeados contra o genoma completo de *Arabidopsis thaliana*. Foram utilizados critérios rigorosos para a identificação das marcas a fim de reduzir o risco de assumir falsos positivos, como frequência mínima de alelo (MAF) de 50%, cobertura mínima de 100 X.

Os SSRs putativos foram identificados, a partir de sequências não redundantes, com tamanho mínimo de 600 pb e cobertura mínima de 50 X, com a utilização do *software* SSR Locator (MAIA et al. 2008). O tamanho de 600 pb foi estabelecido como mínimo para o desenho de iniciadores direto e reverso, após a identificação do motivo SSR. O valor de cobertura de 50 X tem por função garantir a fidelidade das sequências visando minimizar eventuais erros no desenho dos iniciadores. O número mínimo de repetições (rep.) por motivo SSR foi estabelecido da seguinte forma: mono-: 12 rep.; 2-mer: 6 rep.; 3-mer: 4 rep.; 4-mer a 10-mer: 3 rep.. Dentro de cada sequência contígua, o espaçamento mínimo entre SSRs foi de 100 pb.

Após identificados, os iniciadores direto (*forward*) e reverso (*reverse*) foram desenhados com a utilização do *software* Primer3 (ROZEN; SKALETSKY, 2000). Os parâmetros para o desenho dos SSR foram: temperatura de anelamento (hibridização) a 60 °C; tamanho de oligonucleotídeos entre 18-mer e 22-mer; conteúdo GC entre 40% e 50% (ideal: 50%), presença de apenas uma base C ou G nas últimas 3 bases da extremidade 3' de ambos os iniciadores e tamanho de produtos

amplificados (*amplicons*) entre 100 pb e 550 pb, sem a presença de estruturas secundárias como *loops*, *hairpins*, *dimers* e *cross-dimers*. O intervalo de 450 pb foi escolhido visando a futura criação de amplos painéis *multiplex* de genotipagem em sequenciador automático de DNA, baseado somente em SSRs expressos (gênicos). Cada painel poderá conter de 8 a 12 marcadores SSR marcados na extremidade 5' do iniciador *forward* com os seguintes fluoróforos: 6-FAM (azul); NED (amarelo); VIC (verde); PET (vermelho).

## 6 RESULTADOS

### 6.1 Qualidade e quantidade de RNA total

A qualidade e a quantidade das amostras de RNA são fatores essenciais para o sucesso da construção das bibliotecas de cDNA e posterior sequenciamento dos transcriptomas. Amostras degradadas ou com a presença de contaminantes podem inibir o funcionamento das enzimas envolvidas no preparo das bibliotecas e no sequenciamento, acarretando em um número elevado de leituras curtas, baixo número total de leituras obtidas e bases com baixos valores de qualidade. Os indicadores de qualidade e quantidade de RNA total das amostras foram considerados satisfatórios para a análise de RNA-Seq. O protocolo de precipitação de RNA total (Item 5.5) foi utilizado apenas nas replicatas #1 e #3 de pólen de *A. angustifolia* por apresentarem relação 260/230 inferior a 1,8 (Tabela 2). Posteriormente ao uso deste protocolo, as amostras apresentaram valor acima de 1,8 para a relação 260/230.

Tabela 2. Indicadores de integridade, pureza e quantidade das amostras de RNA total pré-depleção de rRNA

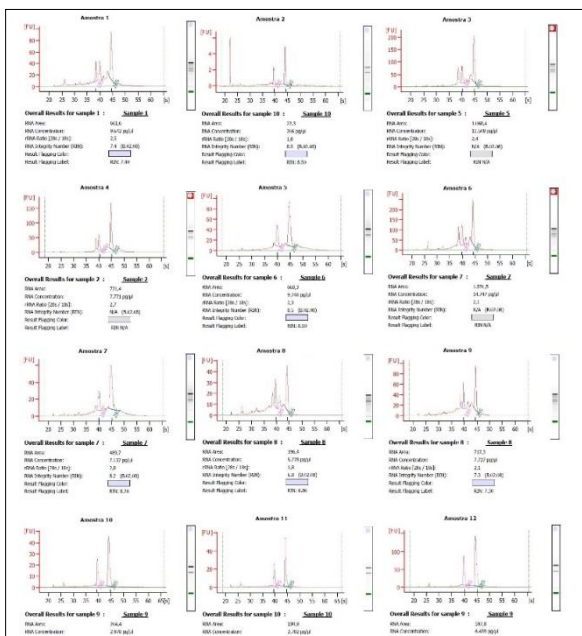
Amostra	Código	RIN*	28S/18S*	260/280**	260/230**	ng/ $\mu$ l total***
Folhas - <i>P. lambertii</i> #1	1	7,40	2,5	2,05	2,16	1.720,0
Folhas - <i>P. lambertii</i> #2	2	8,50	1,8	2,16	2,29	952,0
Folhas - <i>P. lambertii</i> #3	3	N/A****	2,4	2,14	2,29	2.000,0
Folhas - <i>Araucaria</i> #1	4	N/A****	2,7	2,00	1,87	147,0
Folhas - <i>Araucaria</i> #2	5	8,50	2,3	2,05	2,05	478,0
Folhas - <i>Araucaria</i> #3	6	N/A****	2,1	1,98	2,11	2.000,0
Hastes - <i>Araucaria</i> #1	7	8,20	2,8	2,04	2,14	224,0
Hastes - <i>Araucaria</i> #2	8	6,80	1,8	1,97	1,85	602,0
Hastes - <i>Araucaria</i> #3	9	7,30	2,1	2,02	2,21	326,0
Pólen - <i>Araucaria</i> #1	10	9,20	1,9	2,08	1,62	38,8
Pólen - <i>Araucaria</i> #2	11	9,10	1,9	2,15	2,10	42,6
Pólen - <i>Araucaria</i> #3	12	9,40	2,0	2,12	1,72	60,6

Dados oriundos do \*Agilent Bioanalyzer 2100, \*\*NanoDrop 2000 e \*\*\*Qubit 1. \*\*\*\*RIN não fornecido pelo instrumento. A principal causa para o não fornecimento do valor de RIN é a detecção de um pico na região rápida (*fast*

region), entre os picos de RNA 5S e 18S. Contudo, as amostras foram aceitas por apresentarem eletroferogramas com padrões de RIN entre 7,0 e 9,0 (Figura 11).  
 Fonte: Autor

Uma compilação dos eletroferogramas gerados pelo equipamento Agilent 2100 Bioanalyzer para cada amostra estudada é apresentada na Figura 11. Os principais indicadores de qualidade de RNA total fornecidos nesta análise são os valores de integridade de RNA (RIN) e a relação 28S/18S, calculada a partir da intensidade dos principais picos dos rRNAs.

Figura 11. Eletroferogramas das amostras pré-depleção ribossomal utilizando o kit RNA 6000 Nano. Picos não marcados na base correspondem a rRNA adicionais. Amostras 1 a 3: RNA total isolado de folhas de *P. lamberti*; Amostras 4 a 6: RNA total isolado de folhas de *A. angustifolia*; Amostras 7 a 9: RNA total isolado de hastes de *A. angustifolia*; Amostras 10 a 12: RNA total isolado de pólen de *A. angustifolia*



Fonte: Autor

## 6.2 Depleção de RNA

Os valores da quantificação do mRNA após a depleção dos RNAs ribossomais são apresentados na Tabela 3. A leitura da biblioteca número 10 não foi realizada pois o chip do equipamento Agilent 2100 Bioanalyzer realiza apenas 11 leituras por chip de corrida, sendo os demais espaços destinados aos padrões e demais controles. A biblioteca de cDNA da replicata faltante foi iniciada com base na média das outras duas replicatas de pólen.

Tabela 3. Quantificação de mRNA pós-depleção ribossomal

Biblioteca	Amostra	pg/ $\mu$ l
#1	Folhas	192.723
	<i>P. lambertii</i>	
#2	Folhas	63.841
	<i>P. lambertii</i>	
#3	Folhas	13.467
	<i>P. lambertii</i>	
#4	Folhas	101.522
	<i>A. angustifolia</i>	
#5	Folhas	54.460
	<i>A. angustifolia</i>	
#6	Folhas	51.096
	<i>A. angustifolia</i>	
#7	Hastes	93.220
	<i>A. angustifolia</i>	
#8	Hastes	69.718
	<i>A. angustifolia</i>	
#9	Hastes	4.070
	<i>A. angustifolia</i>	
#10	Pólen	*
	<i>A. angustifolia</i>	
#11	Pólen	3.542
	<i>A. angustifolia</i>	
#12	Pólen	882
	<i>A. angustifolia</i>	

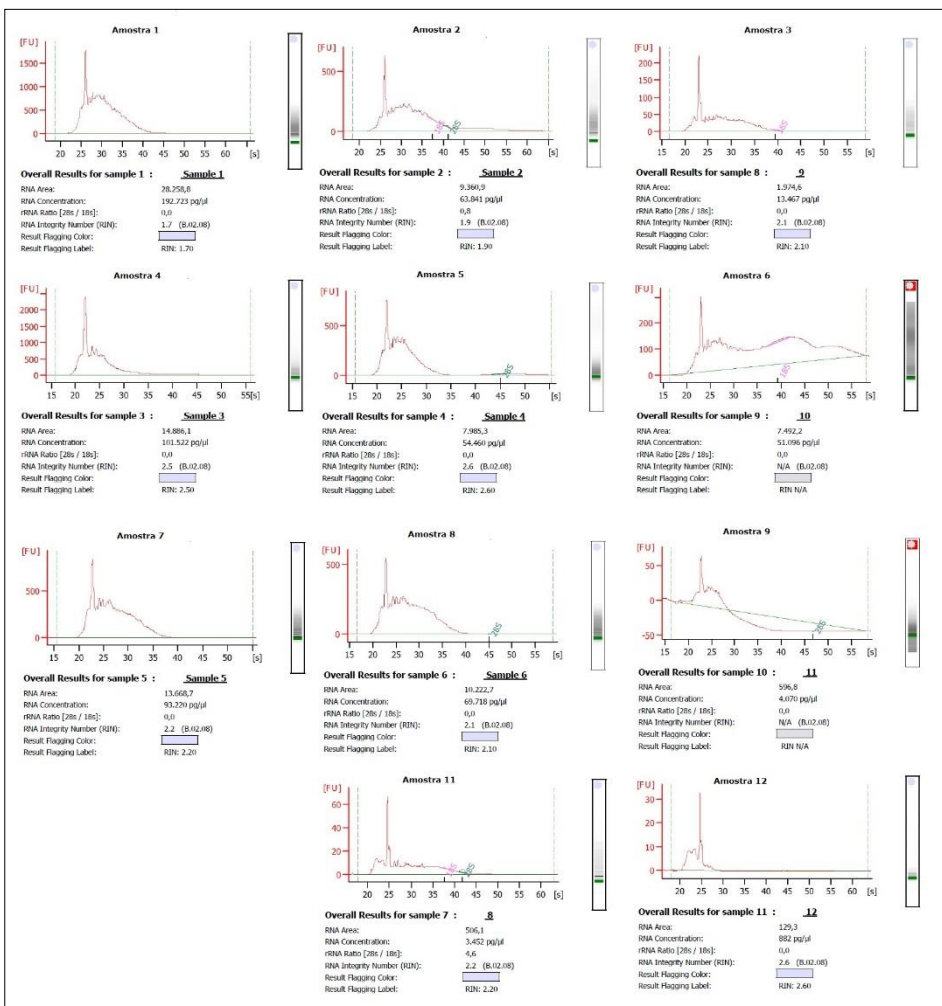
Dados oriundos do Agilent Bioanalyzer 2100

\*Leitura não realizada, biblioteca iniciada com base nos valores das outras replicatas

Fonte: Autor

Os eletroferogramas desta análise (Figura 12), também são utilizados para verificar a presença de rRNA residual caso sejam identificados picos nas posições dos rRNAs 18S e 28S, com a eventual necessidade de uma segunda ou terceira rodada de depleção ribossomal.

Figura 12. Eletroferogramas das amostras pós-depleção ribossomal utilizando o kit RNA 6000 Nano. Amostras 1 a 3: RNA total isolado de folhas de *P. lamberti*; Amostras 4 a 6: RNA total isolado de folhas de *A. angustifolia*; Amostras 7 a 9: RNA total isolado de hastes de *A. angustifolia*; Amostras 11 e 12: RNA total isolado de pólen de *A. angustifolia*



Fonte: Autor

Nestes eletroferogramas, as amostras de mRNA possuem RIN baixo em função da baixíssima presença de rRNA (ausência de picos), pois o indicador de qualidade é calculado com base nos rRNA 18S e 25S. Estes valores não representam a integridade do mRNA, que é analisada em função do formato dos picos após o marcador (*ladder*). Todas as amostras de mRNA foram consideradas com integridade satisfatória e prosseguiram para a construção das bibliotecas de cDNA (Figura 12).

### 6.3 Sequenciamento, *trimming* e montagem dos transcriptomas

Após a realização de 12 corridas de sequenciamento, as leituras foram agrupadas em um único arquivo por biblioteca, totalizando 12 arquivos multifasta. Foram obtidos um total de 341.278.253 milhões (45,6 Gb) de leituras brutas com comprimento médio de leitura de 56,71 pb. A porcentagem média de leituras utilizáveis foi de 40,85% após a retirada dos *barcodes*, leituras de baixa qualidade, sequências adaptadoras, policlonais e contaminantes (Tabela 4), reduzindo a quantidade de dados biológicos para 18 Gb. Após o *trimming*, o tamanho médio das leituras aumentou significativamente para 81,71 pb.

Os maiores transcriptomas sequenciados foram os de folhas e hastes de araucária com mais de 106 milhões e 98 milhões de leituras brutas, respectivamente. Após o controle de qualidade, estes transcriptomas continuaram a ser os maiores com mais de 43 e 38 milhões de leituras processadas, respectivamente. Entretanto, o transcriptoma de pólen de araucária foi o transcriptoma que mais apresentou sequências contíguas com cobertura acima de 5X, com mais de 20 mil *contigs* na soma das replicatas biológicas. O menor transcriptoma sequenciado foi o de folhas de *P. lambertii* com aproximadamente 56 milhões de leituras brutas, 26 milhões de leituras processadas e mais de 12 mil sequências contíguas com cobertura acima de 5X (Tabela 4).

Somente as leituras com qualidade (*Phred score*) acima de 20 foram utilizados para a montagem *de novo*. Desta montagem, somente as sequências contíguas com cobertura acima de 5X foram anotadas funcionalmente de acordo com os termos (GO), com a utilização do *software* Blast2Go (Tabela 4).



Tabela 4. Resumo das leituras, bases e sequências contíguas obtidas por meio de RNA-Seq das amostras analisadas

Biblioteca*	Total de leituras brutas	Total de bases brutas	Leituras com alta qualidade (≥ QV 20)	Tamanho médio das leituras pré-trimming	Tamanho médio das leituras pós-trimming	Contigs (Cobertura > 5)
FPL-1	11.088.379	677.744.884	5.147.921 (46,42%)	77,06	97,4	1.835
FPL-2	23.306.824	1.444.916.047	12.055.425 (51,72%)	60,23	83,05	5.307
FPL-3	21.499.990	1.258.458.076	9.514.592 (44,25%)	58,21	83,98	5.076
FAA-1	28.329.792	1.635.654.279	12.290.435 (43,38%)	57,22	82,82	4.413
FAA-2	43.521.307	2.102.054.049	13.634.866 (31,33%)	48,18	75	5.085
FAA-3	35.101.552	2.132.243.356	17.078.643 (48,65%)	59,55	82,78	5.342
HAA-1	31.229.808	1.670.474.899	12.359.417 (39,57%)	52,07	77,5	3.011
HAA-2	30.996.988	1.528.696.508	10.587.475 (34,15%)	48,53	72,55	2.991
HAA-3	36.203.778	2.154.692.603	15.512.055 (42,84%)	57,92	86,53	6.846
PAA-1	34.181.932	1.907.323.472	14.863.610 (43,48%)	54,74	78,98	8.946
PAA-2	10.672.054	640.610.186	4.856.725 (45,51%)	57,56	84,2	2.309
PAA-3	35.145.849	1.743.189.630	11.508.336 (32,74%)	49,28	75,77	8.882
<b>Total</b>	<b>341.278.253</b>	<b>18.896.057.989</b>	<b>139.409.500 (40,85%)</b>			<b>60.043</b>

\*FPL – folhas *P. lambertii*; FAA – folhas *A. angustifolia*; HAA – hastes *A. angustifolia*; PAA – pólen *A. angustifolia*.

Fonte: Autor

O conteúdo GC das amostras analisadas ficou abaixo dos 50%, com exceção de uma replicata biológica de hastes de *A. angustifolia* com valor de 50,8%. O valor médio das replicatas de folhas de *P. lambertii* foi de 46,1%, seguido de 45,4% para pólen de *A. angustifolia*, 46,96% e 48,66% para acículas e hastes de *A. angustifolia*, respectivamente (Tabela 5).

Tabela 5. Proporção de nucleotídeos (%) e conteúdo GC% após o controle de qualidade das bibliotecas sequenciadas

<b>Biblioteca*</b>	<b>A</b>	<b>T</b>	<b>C</b>	<b>G</b>	<b>C+G%</b>
<b>FPL-1</b>	28,0	26,4	19,7	25,9	<b>45,6</b>
<b>FPL-2</b>	27,9	26,0	20,3	25,8	<b>46,1</b>
<b>FPL-3</b>	27,0	26,4	21,0	25,6	<b>46,6</b>
<b>FAA-1</b>	26,8	25,9	21,2	26,1	<b>47,3</b>
<b>FAA-2</b>	28,5	25,8	19,6	26,1	<b>45,7</b>
<b>FAA-3</b>	27,0	25,1	21,1	26,8	<b>47,9</b>
<b>HAA-1</b>	24,7	24,5	23,1	27,7	<b>50,8</b>
<b>HAA-2</b>	26,7	24,8	21,7	26,8	<b>48,5</b>
<b>HAA-3</b>	27,5	25,8	20,9	25,8	<b>46,7</b>
<b>PAA-1</b>	29,0	26,6	19,2	25,2	<b>44,4</b>
<b>PAA-2</b>	27,6	25,6	20,4	26,4	<b>46,8</b>
<b>PAA-3</b>	28,4	26,6	19,5	25,5	<b>45,0</b>

\*FPL – folhas *P. lambertii*; FAA – folhas *A. angustifolia*; HAA – hastes *A. angustifolia*; PAA – pólen *A. angustifolia*.

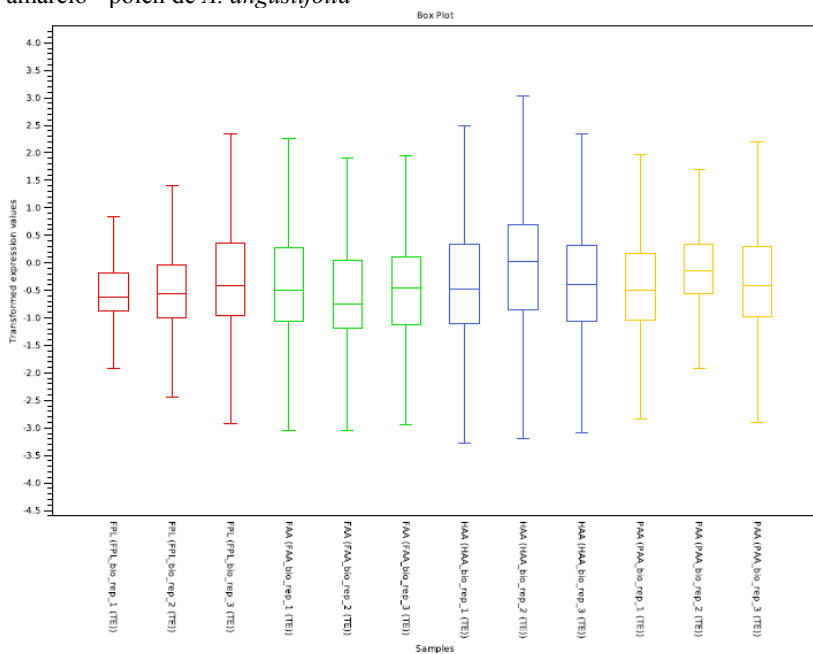
Fonte: Autor

#### 6.4 Análise funcional e caracterização dos transcriptomas

A distribuição dos valores das replicatas biológicas de cada amostra (*Box plot*, Figura 13), revelam algumas discrepâncias na distribuição dos dados de RNA-Seq. Cada “caixa” é formada pelo primeiro quartil, pelo terceiro quartil e pela mediana. As hastes superiores e inferiores se estendem até seus respectivos limites. Caso existam valores fora dos limites estes dados são considerados como *outliers* e representados por asteriscos. Nas corridas analisadas não foram identificados *outliers*, no entanto as corridas apresentaram grandes variações no que se refere a distribuição dos limites superiores e

inferiores. Possíveis variações entre as replicatas biológicas já eram esperadas por se tratarem de amostras de RNA e coletadas a campo (ambientais). Assim, a manipulação do RNA nas etapas de isolamento, depleção de rRNA, purificação e construção das bibliotecas de cDNA podem vir a degradar o mRNA. Outra fonte de variação pode estar nos distintos níveis de expressão gênica das próprias replicatas no momento da coleta das amostras.

Figura 13. *Box plot*. Mapeamento contra o genoma completo de *Arabidopsis thaliana*. As cores representam amostras distintas; vermelho - folhas de *P. lambertii*; verde - folhas de *A. angustifolia*; azul - hastes de *A. angustifolia*; amarelo - pólen de *A. angustifolia*

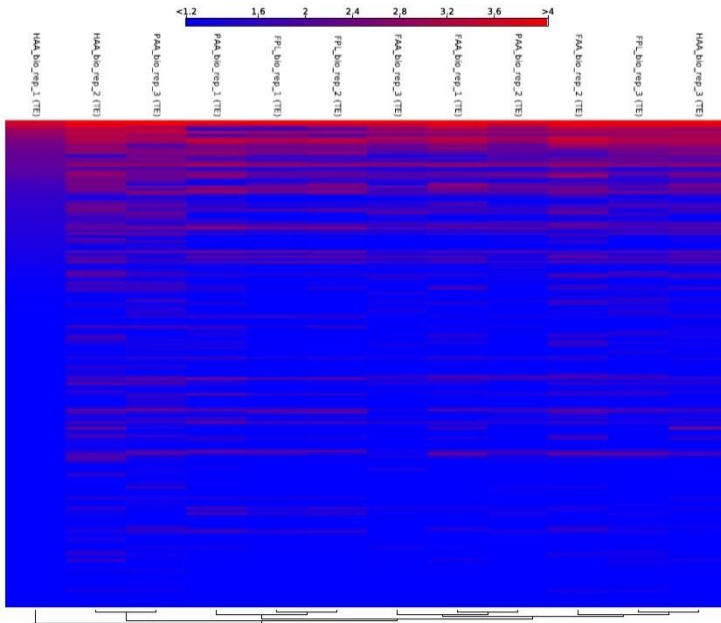


Fonte: Autor

Outro recurso relacionado a visualização de um grande volume de dados transcriptômicos e ao controle de qualidade das amostras sequenciadas são os *Heat maps* (Figura 14). Estas figuras são ferramentas utilizadas para a visualização de diferenças quantitativas nos níveis de expressão gênica das amostras. Além disso, um dendrograma é gerado a partir destes gráficos bi-dimensionais, evidenciando a similaridade entre as amostras. Cada coluna no *Heat map* representa uma replicata biológica

de uma amostra sequenciada, sendo que cada linha representa a expressão de determinado gene. A cor vermelha indica altos níveis de expressão gênica, a cor azul indica o contrário. Esta análise foi realizada a partir do mapeamento das leituras contra o genoma completo de *Arabidopsis thaliana*. Apesar das grandes diferenças genômicas entre as espécies deste trabalho e *A. thaliana*, a análise dos *Heat maps* oferece informações relevantes no que se refere a similaridade e ao controle de qualidade das amostras sequenciadas.

Figura 14. *Heat map* mostrando o padrão de expressão das amostras e suas replicatas biológicas quando do mapeamento referenciado contra *Arabidopsis thaliana*



FPL: folhas *P. lambertii*; FAA: folhas *A. angustifolia*; HAA: hastes *A. angustifolia*; PAA: pólen *A. angustifolia*.

Fonte: Autor

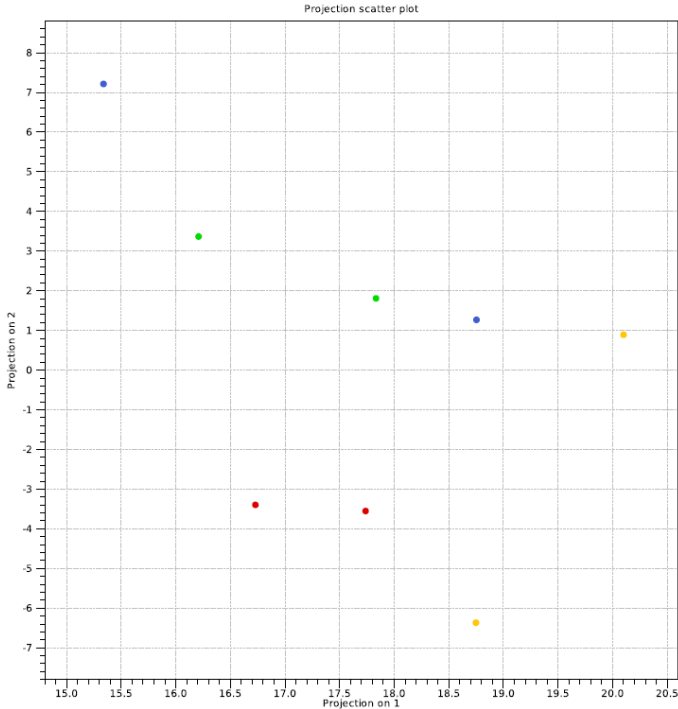
Com base nas medianas presentes no *Box plot* (Figura 13) e no padrão de expressão dos *Heat maps* (Figuras 14), foi retirada da análise transcriptômica a replicata biológica mais discrepante de cada amostra sequenciada, como por exemplo a replicata n°1 de hastes de *A. angustifolia* (Figura 14, primeira coluna). Deste modo, a análise dos

transcriptomas de cada amostra foi realizada com duas replicatas biológicas cada.

A análise dos componentes principais (PCA) das replicatas selecionadas para cada amostra é apresentada na Figura 15. As replicatas de folhas de *A. angustifolia* (pontos verdes) mostram certa proximidade entre si. O mesmo ocorre com as amostras de folhas de *P. lambertii* (pontos vermelhos), onde ocorre a maior proximidade entre replicatas das amostras sequenciadas. Já as amostras de hastes (pontos azuis) e pólen (pontos amarelos) não apresentam agrupamento entre si. As replicatas de hastes mostram certa proximidade das replicatas de folhas de *A. angustifolia*, onde a menor distância entre dois pontos de toda a PCA é entre uma replicada de hastes e uma replicada de folhas. Os dois pontos das amostras de hastes apresentam a maior distância entre si em relação as quatro amostras analisadas.

Em análise transcriptômica via RNA-Seq, como em outros ensaios, as variações entre as replicatas podem ser de origem técnica como também de origem biológica. Partes das variações encontradas entre as amostras e replicatas nos gráficos de caixa (Figura 13), *Heat map* (Figura 14) e na P.C.A (Figura 15) podem ser atribuídas aos variados níveis de expressão gênica entre as amostras coletadas nos distintos dias. Além disso, o armazenamento das amostras e o processamento de RNA desde o momento do isolamento até a construção das bibliotecas de cDNA podem ter causado as discrepâncias entre as replicatas.

Figura 15. Análise dos componentes principais (PCA) quando da utilização do genoma de *Arabidopsis thaliana* como referência. A cor em vermelho representa as replicatas das amostras de folhas de *P. lamberti*; verde representa folhas de *A. angustifolia*; azul representa hastes de *A. angustifolia*; amarelo representa pólen de *A. angustifolia*

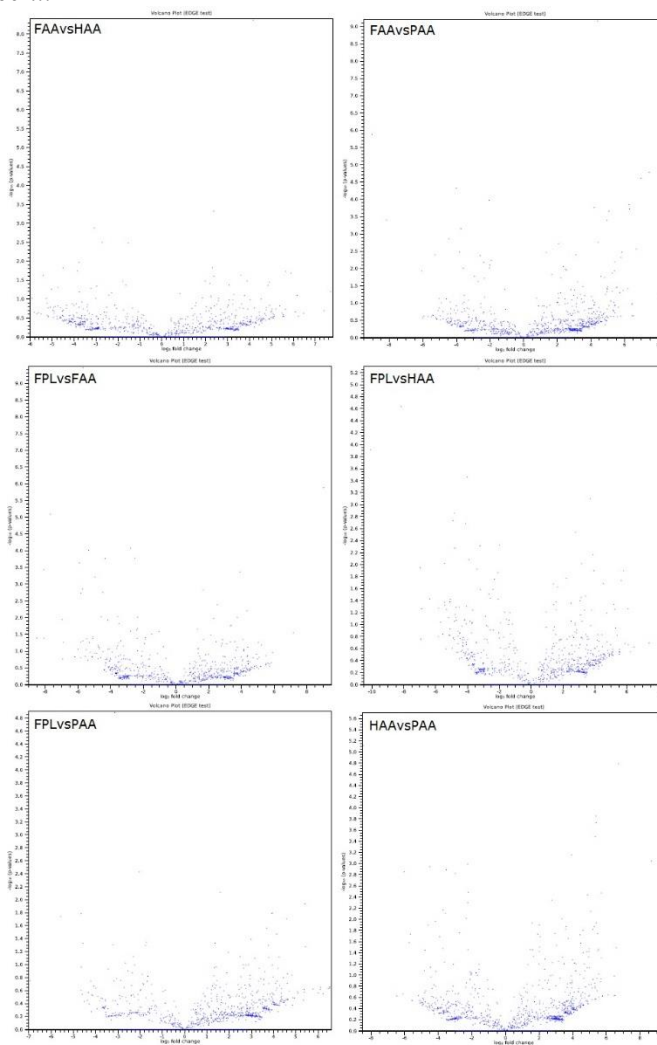


Fonte: Autor

Para uma melhor visualização das diferenças entre as amostras, o padrão de expressão dos transcritos foi comparado par-a-par com a utilização dos gráficos vulcão (*Volcano plots*) (Figura 16). As figuras mostram a relação entre os *p-values* e a magnitude da diferença dos valores de expressão gênica das amostras. Os pontos azuis no gráfico representam os transcritos identificados em ambas amostras da comparação. No eixo y são plotados os valores dos *p-values* em  $-\log_{10}$ . No eixo x são plotados os valores de “*fold change*” dos transcritos. Quanto maior a diferença na expressão de um gene, mais extremo será o ponto no eixo x. E quanto mais significativo, menor o *p-value*, sendo maior o valor em  $-\log_{10}$ . Os pontos de interesse, possuidores de diferenças significativas na expressão irão aparecer na parte elevada dos

gráficos.

Figura 16. *Volcano plot* das amostras comparadas par-a-par e mapeadas contra o genoma completo de *Arabidopsis thaliana*. FAA: Folhas de *A. angustifolia*; HAA: Hastes de *A. angustifolia*; PAA: Pólen de *A. angustifolia*; FPL: Folhas de *P. lambertii*



Fonte: Autor

A análise das figuras mostra que a grande maioria dos genes analisados não possui diferenças significativas nos níveis de expressão, pois estão abaixo do valor 2 na escala de  $-\log_{10}$ , o que representa um p-value de 0,01. A comparação mais homogênea entre amostras foi a realizada entre folhas e hastes de *A. angustifolia*. O maior número de diferenças significativas nos níveis de expressão estão nas comparações da amostra foliar de *P. lambertii* com as amostras de foliares, de hastes e de pólen de *A. angustifolia* (Figura 16).

Após a realização destas etapas relacionadas ao controle de qualidade dos resultados de sequenciamento, similaridade e agrupamento entre as amostras, foi realizado o BLASTX das sequências contíguas provenientes da montagem de novo.

As duas espécies que mais contribuíram com *Top-hits* na análise de BLASTX foram *Picea sitchensis*, com mais de 5700 *Top-hits* e *Amborella trichopoda* com mais de 1500 *Top-hits* (Figura 17). Os *Top-hits* são os melhores alinhamentos obtidos na análise de BLASTX, possuindo o menor *e-value* (maior confiança) entre todos os *hits* obtidos. O gênero *Picea* é um dos gêneros de coníferas mais bem caracterizados nas áreas genômicas, transcriptômicas e proteômicas, possuindo elevado volume de dados biológicos nas mais importantes bases de dados (RALPH et al. 2006; RALPH et al. 2008; CHEN et al. 2012; RAHERISON et al. 2012; BIROL et al. 2013).

*Amborella trichopoda* é uma espécie de angiosperma considerada basal, sendo o mais recente ancestral comum de todas as angiospermas existentes. Além disso, a espécie possui distribuição restrita as ilhas da Nova Caledônia, local possuidor de espécies do gênero *Podocarpus* (KELCH, 1997; KELCH, 1998) e local de endemismo de 13 espécies do gênero *Araucaria* (SETOGUCHI et al. 1998). O genoma de *A. trichopoda* já foi completamente sequenciado (AMBORELLA GENOME PROJECT, 2013) e revela características semelhantes aos genomas de gimnospermas. Além disso, Rice et al. (2013) reportaram a transferência horizontal de genomas mitocondriais completos de outras espécies de angiospermas para dentro do genoma de *A. trichopoda*. A espécie possui cinco genomas mitocondriais distintos e independentes, fruto das diversas transferências de genes ao longo da evolução da espécie. A elevada similaridade das sequências de *A. angustifolia* e *P. lambertii* com sequências de *A. trichopoda* sugere que as coníferas nativas da Nova Caledônia também transferiram DNA para esta angiosperma basal. No entanto, esta hipótese deverá ser testada com a realização de análises filogenéticas específicas.





O número de transcritos anotados com sucesso com a utilização do *software* Blast2Go a partir dos resultados do BLASTX foi variável entre as corridas de sequenciamento. Em *A. angustifolia*, foram anotados 4.354 unigenes distintos a partir do transcriptoma de folhas, 3.643 unigenes do transcriptoma de hastes e 9.354 unigenes do transcriptoma de pólen. Em *P. lambertii* foram identificados 4.649 unigenes no transcriptoma de folhas.

Os resultados das categorizações funcionais dos transcriptomas de *A. angustifolia* estão explicitados nas Figuras 18, 19, 20 e 21. Como foi sequenciado apenas uma amostra de *P. lambertii*, este transcriptoma foi denominado de transcriptoma referência da espécie (Figura 21). A comparação dos termos anotados entre os transcriptomas de folhas de *A. angustifolia* e *P. lambertii* é apresentada na Figura 23.

Os diferentes genes anotados foram classificados nos respectivos termos (#GO) e separados em três distintas categorias. Estas categorias e sub-categorias reúnem os genes ligados a processos biológicos (BP), genes responsáveis pelas funções moleculares (MF) e genes responsáveis aos componentes celulares (CC) (ASHBURNER et al. 2000).

Em todos os transcriptomas de *A. angustifolia*, o maior número de sub-categorias foi alocado na categoria de processos biológicos, seguido das categorias de componentes celulares e funções moleculares. No transcriptoma de grãos de pólen foram identificadas 307 sub-categorias, sendo 40 sub-categorias de componentes celulares, 196 sub-categorias de processos biológicos e 71 sub-categorias de funções moleculares. Foram identificadas as mesmas 71 sub-categorias de termos GO nos transcriptomas de folhas e hastes de *A. angustifolia* e folhas de *P. lambertii*, sendo 17 sub-categorias de componentes celulares, 42 sub-categorias de processos biológicos e 12 sub-categorias de funções moleculares. No entanto, o número de genes anotados funcionalmente foi maior no transcriptoma de folhas de *A. angustifolia*.

Na categoria funcional de Processos Biológicos (BP), os termos mais frequentes em todos os transcriptomas analisados foram de transcritos relacionados com processos de biossíntese, processos celulares e processos que envolvem nucleobases (Figuras 18, 19 e 20). Os genes que atuam nos processos de biossíntese são genes que mediam reações para a formação de substâncias, consumindo energia para a transformação de substâncias simples em complexas, como por exemplo, o processo fotossintético. Os processos celulares englobam todos os processos que ocorrem no interior das células, visando o crescimento e manutenção das mesmas. Os processos que envolvem nucleobases são

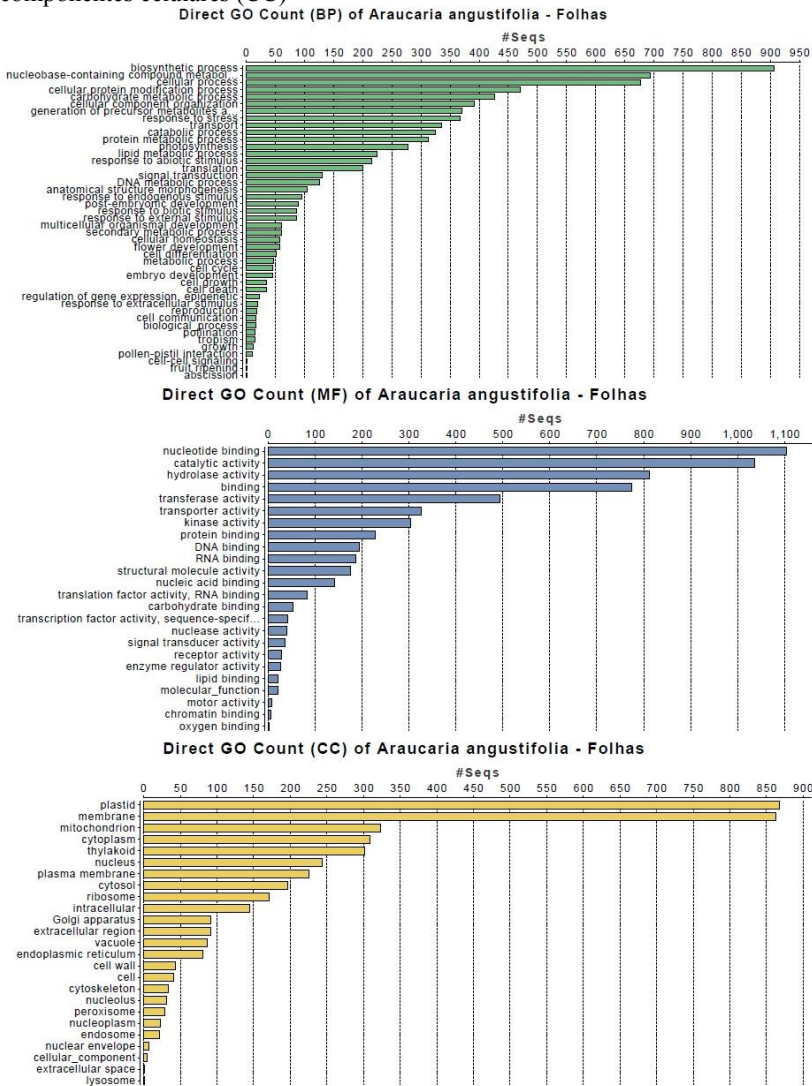
todos os processos metabólicos onde nucleotídeos, nucleosídeos e ácidos nucleicos atuam, como por exemplo, a transcrição e a tradução.

Na categoria de funções moleculares, os termos mais frequentes foram os relacionados a atividade catalítica e a ligação ou acoplamento em geral (binding). A atividade catalítica remete a atividade enzimática realizada por proteínas (enzimas) e também por RNA (ribozimas). Estas enzimas identificadas em cada transcriptoma foram distribuídas em seis principais classes: Oxidoredutares; Transferases; Hidrolases; Liasas; Isomerases e Ligases. (Apêndice D).

Em relação aos componentes celulares das amostras sequenciadas, foram identificados mais transcritos relacionados ao plastídeos nos transcriptomas de folhas de *A. angustifolia* e *P. lambertii* do que nos transcriptomas de hastes e pólen de *A. angustifolia*. Nestes dois últimos transcriptomas, os transcritos relacionados a formação das membranas celulares, com exceção da membrana plasmática, foram predominantes. A predominância de transcritos relacionados aos plastídeos nos transcriptomas de folhas ocorre em função das folhas serem fotossinteticamente ativas. Já a presença destes mesmos transcritos nos grãos de pólen se dão em função dos plastídeos de reserva energética, como os leucoplastos.

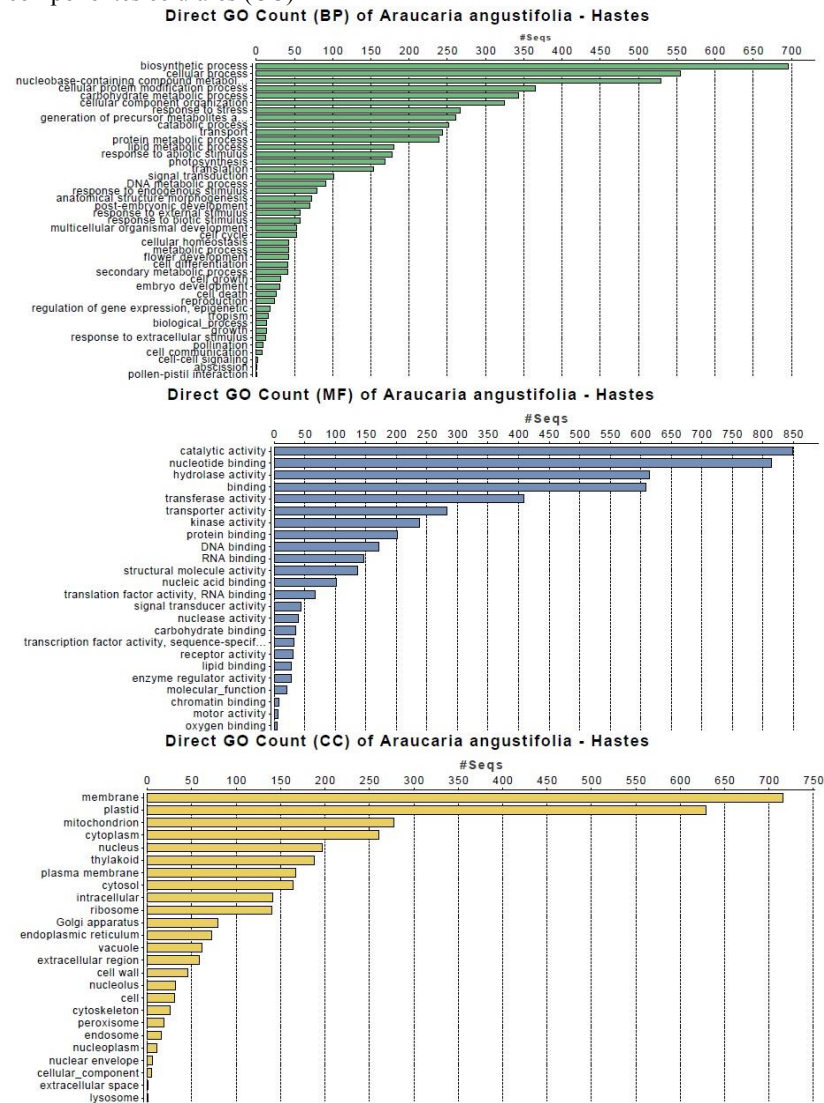
Ainda em relação ao pólen de *A. angustifolia*, foram encontrados transcritos responsáveis pela formação da matriz proteica extracelular (Componentes Celulares, Figura 20). Estes transcritos não foram identificados nos demais transcriptomas e são responsáveis pela formação da exina que recobre e protege os grãos de pólen. Além disso, pode-se constatar a presença de transcritos mitocondriais na categoria de componentes celulares, apesar de em *A. angustifolia* os grãos de pólen não transmitirem mitocôndrias, apenas os cloroplastos (Figura 20). Estes transcritos mitocondriais fazem parte do transcriptoma tardio dos grãos de pólen, necessários durante o início da germinação dos tubos polínicos. Outros termos exclusivos no transcriptoma de grãos de pólen são: simplasto, atividade antioxidante, reserva nutricional, biogênese de componentes celulares, sistema imune, locomoção, processos rítmicos e formação anatômica estrutural. Estes termos exclusivos representam menos de 1 % dos genes anotados no transcriptoma de grãos de pólen. Os transcriptomas de folhas e hastes de *A. angustifolia* apresentaram comportamento muito similar, sendo que no transcriptoma de folhas o número de unigenes anotados foi maior, com exceção dos seguintes termos: regulação enzimática, transdução molecular, transporte e crescimento.

Figura 18. Distribuições das anotações funcionais com utilização dos termos GO no transcriptoma de folhas de *A. angustifolia*. As barras do histograma representam o número de unigenes identificados nas mais frequentes categorias funcionais de processos biológicos (BP), funções moleculares (MF) e componentes celulares (CC)



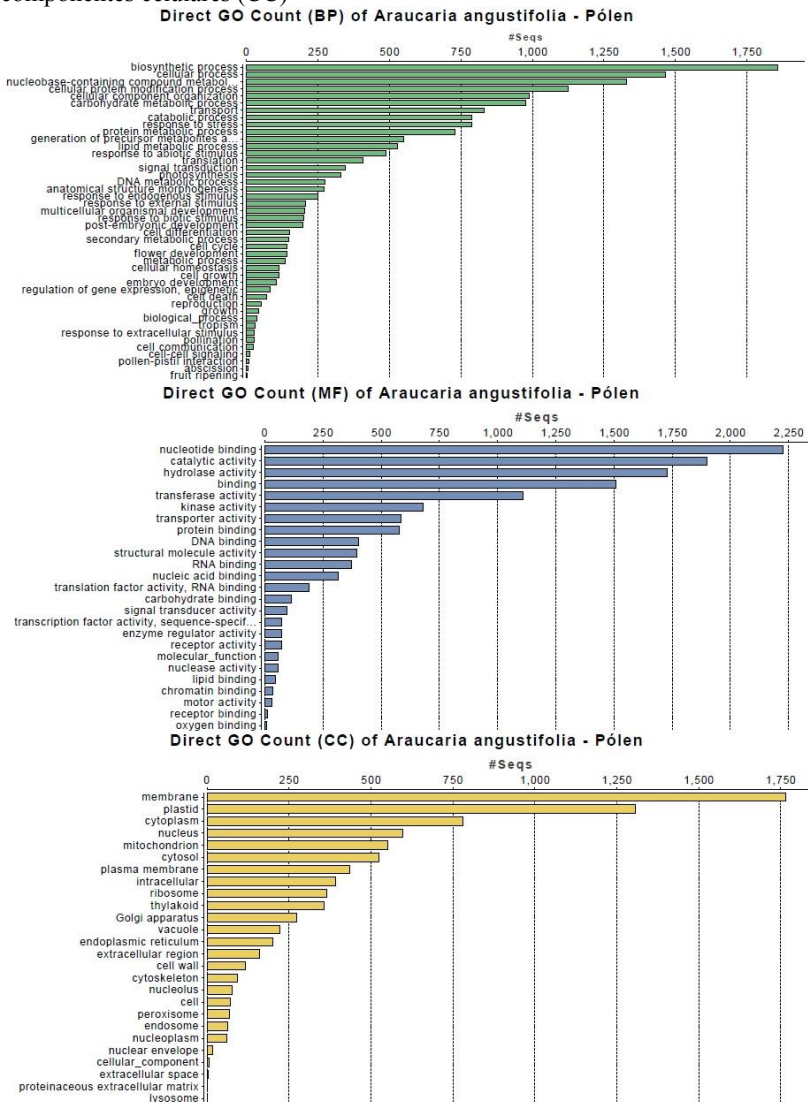
Fonte: Autor

Figura 19. Distribuições das anotações funcionais com utilização dos termos GO no transcriptoma de hastes de *A. angustifolia*. As barras do histograma representam o número de unigenes identificados nas mais frequentes categorias funcionais de processos biológicos (BP), funções moleculares (MF) e componentes celulares (CC)



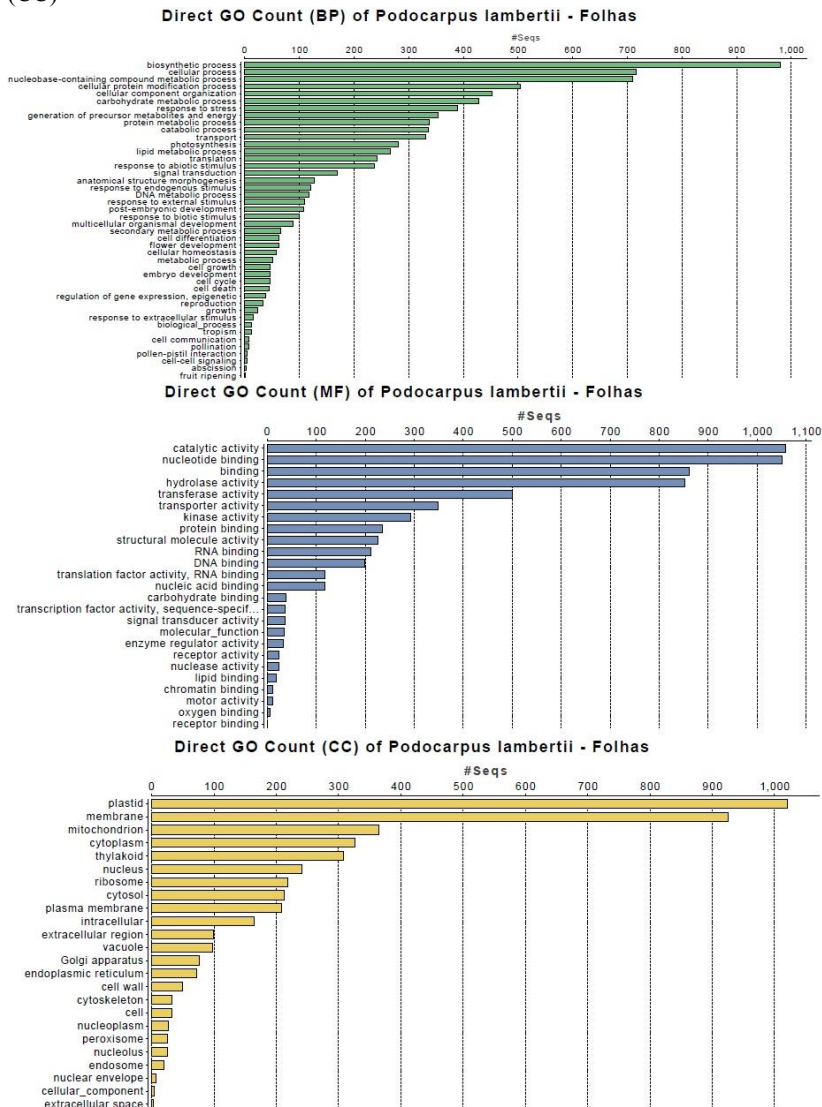
Fonte: Autor

Figura 20. Distribuições das anotações funcionais com utilização dos termos GO no transcriptoma de pólen de *A. angustifolia*. As barras do histograma representam o número de unigenes identificados nas mais frequentes categorias funcionais de processos biológicos (BP), funções moleculares (MF) e componentes celulares (CC)



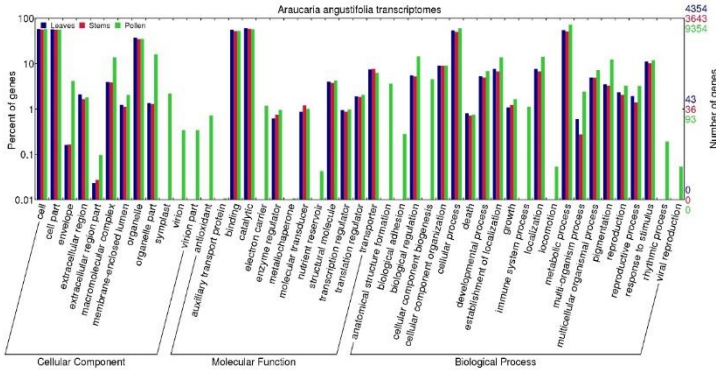
Fonte: Autor

Figura 21. Distribuições das anotações funcionais com utilização dos termos GO no transcriptoma de folhas de *P. lambertii*. As barras do histograma representam o número de unigenes identificados nas mais frequentes categorias funcionais de processos biológicos (BP), funções moleculares (MF) e componentes celulares (CC)



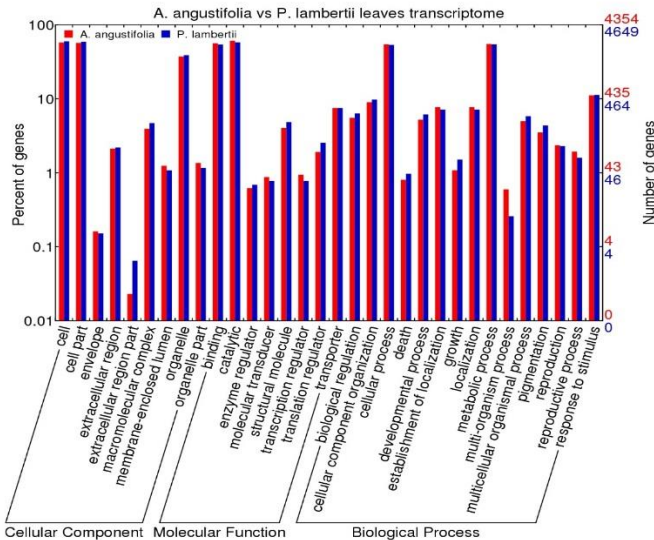
Fonte: Autor

Figura 22. Distribuição dos termos GO (level 2) mais abundantes nos transcriptomas de folhas (azul), hastes (vermelho) e pólen (verde) de *A. angustifolia*



Fonte: Autor

Figura 23. Distribuição dos termos GO (level 2) mais abundantes nos transcriptomas de folhas de *A. angustifolia* (vermelho) e *P. lambertii* (azul)



Fonte: Autor.



## 6.5 Identificação de SNPs e SSRs putativos

Ao total, foram identificados 224 e 35 SNPs putativos para *A. angustifolia* e *P. lambertii*, respectivamente, com base em 82.114 seqüências contíguas (Apêndices E e F). Este resultado corresponde a 1 SNP putativo a cada 56,78 kb. Inserções ou deleções de apenas 1 pb não foram considerados como marcas SNPs. Para a identificação de marcadores putativos em *A. angustifolia* foram utilizados conjuntamente os dados de sequenciamento das três amostras analisadas, representando o genoma de dois indivíduos distintos. Isto significa que para *A. angustifolia*, a busca por SNPs putativos foi realizada a partir de quatro genomas haploides, contra dois de *P. lambertii*. Embora o volume de dados utilizados para *A. angustifolia* tenha sido três vezes maior do que em *P. lambertii*, o número de marcas putativas foi seis vezes maior. Nos transcritos analisados, a frequência de SNPs por kb em *A. angustifolia* foi praticamente o dobro do que em *P. lambertii* (Tabela 6).

Tabela 6. Resumo dos SNPs putativos identificados em *A. angustifolia* e *P. lambertii*

	<i>A. angustifolia</i>	<i>P. lambertii</i>
Número de <i>contigs</i> analisados	47.825	12.218
Total de bases analisadas	11.183.469	3.521.710
Número de seqüências contendo mais de um SNP	37	05
Número total de SNPs	224	35
Frequência de SNPs	1 a cada 49,93 kb, <b>ou, 0,02 SNP/kb</b>	1 a cada 100,62 kb, <b>ou, 0,01 SNP/kb</b>

Fonte: Autor

Em 37 seqüências de *A. angustifolia* foram encontrados mais de um SNP putativo, para *P. lambertii* este fato ocorreu apenas 5 vezes, sendo que grande parte dos transcritos analisados possui menos de um kilobase de tamanho. Assim, a frequência ou densidade de SNP/kb seja apenas um valor teórico, servindo como referência do número de mutações em relação ao total de bases analisadas. De maneira análoga, poderia ter sido utilizada a porcentagem de bases mutantes em relação ao total de bases analisadas, sem prejuízo algum na informação acessada.

Para *A. angustifolia*, a quantidade de mutações encontradas nos SNPs putativos do tipo transições foi em maior número do que do tipo transversões. Para *P. lambertii*, foi identificada somente uma transição a

mais que transversão. A relação Ti/Tv foi 2,15 para *A. angustifolia* e 1,05 para *P. lambertii*, sendo caracterizada pela divisão do número de transições (Ti) pelo número de transversões (Tv) de bases nitrogenadas (Tabela 7).

Tabela 7. Distribuição dos SNPs putativos em *A. angustifolia* e *P. lambertii* por tipo de mutação

SNP	<i>A. angustifolia</i>	<i>P. lambertii</i>
<b>Transições (Ti)</b>		
A <-> G (purina <-> purina)	72 (32,14%)	07 (20,00%)
C <-> T (pirimidina <-> pirimidina)	81 (36,16%)	11 (31,42%)
<b>Total de transições</b>	<b>153 (68,30%)</b>	<b>18 (51,42%)</b>
<b>Transversões (Tv)</b>		
A <-> T	16 (07,14%)	05 (14,29%)
G <-> T	24 (10,72%)	02 (5,72%)
C <-> G	10 (04,46%)	07 (20,00%)
A <-> C	21 (09,38%)	03 (08,57%)
<b>Total de transversões</b>	<b>71 (31,70%)</b>	<b>17 (48,58%)</b>
<b>Total geral</b>	<b>224</b>	<b>35</b>
<b>Relação Ti/Tv</b>	<b>2,15</b>	<b>1,05</b>

Transições (base púrica vs. base púrica, base pirimídica vs. base pirimídica)

Transversões (base púrica vs. Base pirimídica)

Fonte: Autor

As porcentagens dos tipos de transições e transversões não apresentaram semelhanças entre *A. angustifolia* e *P. lambertii*. Uma provável razão para este fato pode ser o baixo número de sequências contíguas utilizadas para a busca de SNPs putativos para *P. lambertii* em comparação com *A. angustifolia*.

Além dos SNPs, foram identificados 113 *indels* para *P. lambertii* e 431 *indels* para *A. angustifolia*, variando de 1 pb a 10 pb e totalizando 544 *indels* para as duas espécies. A relação entre o número de substituições (SNP) e entre o número de *indels* (S/I) foi de 0,31 para *P. lambertii* e de 0,52 para *A. angustifolia*.

Em relação aos marcadores SSR, somando os dados das duas espécies, foram identificados 100 SSR putativos em 661 sequências contíguas, correspondendo a 1 SSR putativo a cada 5,82 kb. Embora o número de sequências putativas identificadas nas análises de RNA-Seq

tenha sido maior para *A. angustifolia* do que para *P. lambertii*, a frequência de marcas por sequência de DNA foi muito próxima entre as espécies. Já o número de sequências contendo motivos SSR foi praticamente o mesmo (Tabela 8).

Tabela 8. Resumo dos SSRs putativos identificados em *A. angustifolia* e *P. lambertii*

	<i>A. angustifolia</i>	<i>P. lambertii</i>
Nº de <i>contigs</i> não reduntantes analisados	488	173
Total de bases analisadas	413.332	169.638
Número de sequências contendo SSRs	62 ( <b>12,70%</b> )	22 ( <b>12,71%</b> )
Número de sequências contendo mais de um SSR	8	4
Número total de SSRs	72	28
Pares de iniciadores desenhados	35	21
Frequência de SSRs	1 a cada 5,74 kb, <b>ou, 0,17 SSR/kb</b>	1 a cada 6,05 kb, <b>ou, 0,16 SSR/kb</b>

Fonte: Autor

Os maiores valores para número de SSRs ocorrem em *A. angustifolia*, pois os valores apresentados para *P. lambertii*, são provenientes apenas do sequenciamento de uma amostra (folhas) contra três amostras de *A. angustifolia* (folhas, hastes e pólen) (Tabela 4). Dos 100 motivos SSR identificados, em 44 não foi possível o desenho de um dos iniciadores. Isto se deu em função dos motivos repetitivos estarem muito próximos das extremidades 5' e 3' das sequências contíguas, impossibilitando o desenho de um dos dois iniciadores. Os resultados do desenho dos iniciadores, como a sequência dos iniciadores direto (F) e reverso (R), temperatura de anelamento dos oligos ( $T_m$  °C), tamanho esperado de *amplicons* variando entre 113 pb e 565 pb, tipo de motivo SSR e início de fim destes motivos dentro das sequências contíguas analisadas estão inseridos nos Apêndices G e H. Em apenas três casos foram identificadas diferenças de mais de 1 °C entre as temperaturas de hibridização dos iniciadores do mesmo par. Os resultados do BLASTX destas sequências contíguas contendo motivos SSRs estão nos apêndices I e J. Além da identificação da proteína expressa pelo gene possuidor do motivo SSR, é apresentado o local de expressão desta proteína, seja ele nuclear, plastidial ou mitocondrial. Adicionalmente, os valores de *E-value* destes alinhamentos e a espécie do *Top hit* também são apresentados.

Apenas três SSR compostos foram identificados, sendo eles

(GCTTTC)4-(GAG)4, (T)12-(AAT)4 e (CAA)5-(GAG)4-(CAG)4. Os marcadores restantes foram considerados SSR perfeitos. Apesar da busca por SSRs putativos ter sido realizada para a identificação de motivos de até 10-mer, somente foram localizados motivos de até 7-mer. O predomínio dos motivos microssatélites para as duas espécies, de acordo com o tamanho de motivos repetitivos foram para as categorias mono- (39,7 %), 4-mer (22,22 %) e 3-mer (20,63%) (Tabela 9). O tamanho total das regiões SSRs, considerando a soma das repetições, variou de 12 pb até 24 pb. Os tamanhos dominantes nos motivos mononucleotídeos foram de 12 pb e 13 pb. O tamanho de 12 pb também foi predominante nos motivos trinucleotídeos e tetranucleotídeos (Apêndices G e H).

Tabela 9. Distribuição dos motivos dos SSRs putativos identificados em *A. angustifolia* e *P. lambertii* de acordo com o número de repetições

Espécie	Mono-	2-mer	3-mer	4-mer	5-mer	6-mer	7-mer
<i>A. angustifolia</i>	14	2	9	8	2	3	1
<i>P. lambertii</i>	11	2	4	6	0	1	0

Fonte: Autor

Em relação aos motivos mononucleotídeos encontrou-se, igual número de motivos A/T e C/G para *A. angustifolia*. Em *P. lambertii*, 72,8% dos motivos SSR mononucleotídeos foram A/T e 27,2% foram C/G. Neste tipo de motivo, adeninas e timinas, assim como citosinas e guaninas, são tratados da mesma maneira, pois dependendo da orientação em que o fragmento é sequenciado pode-se detectar um ou outro nucleotídeo complementar. Nos motivos dinucleotídeos os motivos AT/TA foram os mais predominantes (Apêndices G e H).

Como a identificação dos SSRs putativos foi realizada a partir de sequências transcritas, estes motivos podem codificar para um ou dois diferentes aminoácidos se repetindo em tandem dentro de um determinado peptídeo. Nos motivos mononucleotídeos, as sequências homopolímeras transcritas codificam para os aminoácidos fenil alanina (A), lisina (T), prolina (G) e glicina (C). Nos motivos trinucleotídeos e nos hexanucleotídeos a fase de leitura dos aminoácidos geralmente não é deslocada, apenas quando uma das bases do motivo está fora de determinado *codon*. Nos demais motivos, a adição ou subtração de uma repetição deslocará a fase de leitura dos aminoácidos a cada diferente alelo.

## 7 DISCUSSÃO

### 7.1 Isolamento, qualidade e quantidade de RNA total

A integridade do RNA total (RIN) está relacionada a diversos fatores, tais como a composição do tecido analisado, nível de maturidade fisiológica, tempo e temperatura de manipulação e armazenamento e presença de RNAses intra e extra-celulares.

As melhores integridades de RNA total foram obtidas nos isolamentos de RNA de pólen de *A. angustifolia*. Este resultado já era esperado, pois o grão de pólen e seu RNA total são muito resistentes a desidratação e a temperaturas ambientais. O mRNA deve estar íntegro para logo iniciar a síntese proteica que visa a germinação do tubo polínico, portanto é baixa a presença de RNAses endógenas nestas estruturas reprodutivas (TWEEL et al. 2006).

O isolamento de RNA total dos grãos de pólen foi realizado com a utilização do kit baseado em coluna de sílica. O tempo de manipulação durante o isolamento é menor em comparação aos outros protocolos utilizados. Outro fator é a não maceração do pólen. As amostras foram apenas congeladas em nitrogênio líquido visando o rompimento das membranas dos grãos de pólen. A maceração mecânica pode romper cadeias longas de mRNA, diminuindo o RIN da amostra.

A desvantagem está no fato deste método ter apresentado o menor rendimento de isolamento (ng/μl) em comparação aos outros métodos. Se for considerada a dinâmica da expressão gênica do grão de pólen maduro, se encontram poucos transcritos responsáveis pelo metabolismo basal e poucos transcritos responsáveis pela sinalização da germinação do tubo polínico. Portanto, a quantidade de mRNA obtido após a depleção ribossomal é representativa em relação ao genoma paterno e ao baixo nível de expressão gênica dos grãos de pólen (transcriptoma tardio).

O rendimento de isolamento do método baseado em CTAB e LiCl foi superior principalmente pela maior quantidade inicial de material vegetal. Foram utilizados 4 g de amostra neste método, contra 100 mg dos isolamentos em coluna de membrana de sílica, sendo essa massa de 100 mg a quantidade máxima de amostra suportada pela coluna. Proporcionalmente, os isolamentos baseados em coluna fornecem um rendimento semelhante aos protocolos abertos, como o baseado em CTAB e LiCl que isolam RNA total com a utilização de 40 vezes mais material vegetal. Por outro lado, não foi possível obter 4 g de pólen por

estrólibo de *A. angustifolia* para a utilização do protocolo baseado em CTAB e LiCl.

A quantificação de mRNA enriquecido, por meio do equipamento 2100 Agilent Bioanalyzer, foi considerada satisfatória, com exceção das replicatas de pólen que já continham baixa quantidade isolada de RNA total. No entanto, todas as bibliotecas de cDNA puderam ser construídas e a baixa quantidade de mRNA para as amostras de pólen não representou problema para o sequenciamento das amostras. Além disso, nas amostras de pólen de *A. angustifolia* foram obtidos os maiores números de seqüências contíguas provenientes da montagem de novo.

Apesar das depleções ribossomais terem sido consideradas satisfatórias levando em consideração os resultados apresentados pelo Bioanalyzer, muitos transcritos anotados funcionalmente pertencem às classes de rRNAs nucleares. Uma nova depleção poderia ter sido realizada para a quase total eliminação dos rRNA. Porém, mais mRNA poderia ter sido perdido e degradado durante este processo.

## 7.2 Sequenciamento das bibliotecas de cDNA

Os quatro transcriptomas deste trabalho são os primeiros transcriptomas de coníferas sequenciados com a utilização da plataforma de sequenciamento de nova geração Ion Proton. Os genomas destas espécies não-modelo são extremamente grandes e complexos, necessitando assim elevado volume de sequenciamento para uma caracterização mínima dos transcritos.

O volume de dados biológicos utilizados (mais de 341 milhões de leituras brutas e 139 milhões de leituras processadas com tamanho médio 81,71 pb) foi considerado satisfatório em comparação a outros estudos transcriptômicos já realizados em *A. angustifolia* e outras coníferas. O número médio de leituras de alta qualidade sequenciadas por biblioteca de cDNA foi de 11.617.458. O transcriptoma de culturas embriogênicas de *A. angustifolia* (ELBL et al. 2014) foi montado com a utilização de mais de 326 milhões de leituras brutas e 188 milhões de leituras processadas, com comprimento médio de 200 pb. O transcriptoma de acículas de *Picea abies*, outra conífera próxima a Araucaria e Podocarpus, foi montado e analisado com a utilização de mais de 136 milhões de leituras brutas, sendo 30 milhões de leituras pareadas e 18 milhões de leituras não pareadas processadas (CHEN et al. 2012). Niu et al. (2013) analisaram o transcriptoma de *Pinus tabuliformes* com a utilização de apenas 822.891 leituras de alta qualidade com comprimento

médio de 382 pb, sendo esse um número muito baixo visto que os genomas das coníferas são extremamente grandes, podendo passar dos 35 Gb tamanho, portanto, mais de 200 X maior que o genoma de *Arabidopsis thaliana* (MACKAY et al. 2012).

O conteúdo GC (porporção de citosina e guanina) das amostras sequenciadas no presente trabalho foi considerado alto em relação a outros transcriptomas vegetais já sequenciados. Garg et al. (2011) caracterizaram os transcriptomas de grão-de-bico, soja e arábido, obtendo valores médios de conteúdo CG de 40,3%, 40,9% e 42,5%, respectivamente. Em *Catharanthus roseus*, outra angiosperma, o valor encontrado foi de 40,65% (VERMA et al. 2014). Tal valor não foi informado nos transcriptomas de embriões somáticos de *A. angustifolia* (ELBL et al. 2014). Em outras coníferas como *Pinus tabuliformes*, o conteúdo CG médio do transcriptoma foi de 43,2% (NIU et al. 2013) e *Pinus dabeshanensis* o valor foi de 44,58% (XIANG et al. 2015). O conteúdo CG dos transcritos em coníferas é mais elevado do que nas espécies de angiospermas. Este valor fornece informações relacionadas ao genoma das espécies, como estrutura (número e tamanho de *introns*), evolução, regulação gênica, estabilidade e conservação das sequências em função das três ligações de hidrogênio entre bases de diferentes cadeias de DNA (VINOGRADOV, 2003).

### 7.3 Análise funcional e caracterização dos transcriptomas

Foram observadas grandes diferenças entre as replicatas biológicas. A grande quantidade de variação pode ser decorrente do fato de que as coletas das replicatas aconteceram em dias distintos. As condições ambientais podem ter alterado os padrões de expressão gênica, resultando nas diferenças observadas nos gráficos de caixa, *heat map*, P.C.A e gráfico vulcão. Portanto, essas diferenças não são surpresas e geneticamente, esperadas.

As replicatas de folhas de *A. angustifolia* ficaram próximas entre si na PCA, sugerindo alto nível de homogeneidade entre as amostras. O mesmo fato ocorreu com as replicatas de folhas de *P. lambertii*. Entretanto, o não agrupamento entre as amostras de hastes revelado pela análise dos componentes principais, pode ser em função da translocação via floema de diferentes mRNAs expressos em outras partes da planta translocados para as hastes. Já a distância entre as replicatas de pólen pode ser em função da amostragem de grande quantidade de grãos de pólen, em estádios de desenvolvimento distintos e com transcriptomas

incompletos ou parcialmente degradados.

Um dos principais objetivos do sequenciamento dos transcriptomas do presente estudo é a obtenção dos perfis de transcrição das amostras em um dos ambientes de ocorrência natural das espécies, visando a utilização destes dados para a caracterização genética, uso e conservação. Além disso, um catálogo completo de genes expressos em *A. angustifolia* e *P. lambertii* é necessário para melhorar a compreensão das funções biológicas e relações evolutivas destas espécies com outras coníferas e angiospermas. Em *A. angustifolia*, foram anotados 4.354 unigenes distintos a partir do transcriptoma de folhas, 3.643 unigenes do transcriptoma de hastes e 9.354 unigenes do transcriptoma de pólen. Em *P. lambertii* foram identificados 4.649 unigenes no transcriptoma de folhas. A maioria destes unigenes anotados representam funções basais comuns presentes em angiospermas e gimnospermas em geral, sendo muitos destes genes expressos constitutivamente.

Do total de unigenes que foram para a etapa de anotação funcional, somente 29,25% dos transcritos foram anotados funcionalmente no transcriptoma de folhas de *A. angustifolia*, 33,45% para pólen de *A. angustifolia*, 25,28% para hastes de *A. angustifolia* e 28,43% para folhas de *P. lambertii*. Este valor é abaixo do esperado, podendo ter ocorrido em função de muitas sequências contíguas formadas na montagem de novo possuírem baixo tamanho (< 200 pb). Além disso, há outras sequências contíguas que podem ser específicas de *A. angustifolia* ou *P. lambertii*, assim não se obtendo correspondência (homólogos) após o BLASTX e anotação funcional. Os valores obtidos no presente trabalho são mais baixos comparativamente aos valores anotados funcionalmente nos transcriptomas de *Pinus dabeshanensis* (79,91%), *Pinus halepensis* (72%), *Picea glauca* (70%), *Taxus mairei* (64,4%) e *Pinus pinaster* (62,2%) (XIANG et al. 2015, PINOSIO et al. 2014; PAVY et al. 2005, HAO et al. 2011, CANALES et al. 2014).

A distribuição das anotações nas diferentes categorias funcionais mostrou uma grande diversidade de transcritos anotados, apesar da baixa porcentagem de anotação de sequências contíguas, evidenciando a alta representatividade dos transcriptomas sequenciados. Apesar das diferenças detectadas em relação as replicatas biológicas dentro de cada amostra, os perfis gerais de anotação funcional das amostras foliares foram semelhantes a transcriptomas foliares sequenciados em outras coníferas como em *Pinus tabuliformes* (NIU et al. 2013) e *Pinus dabeshanensis* (XIANG et al. 2015). Já em relação ao transcriptoma de embriões somáticos de *A. angustifolia* (ELBL et al. 2014), as únicas



semelhanças estão na predominância de transcritos que efetuam processos metabólicos e celulares, transcritos relacionados a atividade catalítica e acoplamento e transcritos relacionados a componentes da célula em si. As demais categorias não apresentam semelhanças com os perfis obtidos no presente trabalho.

#### 7.4 Identificação de SNPs e SSRs putativos

O número de marcadores SNPs putativos encontrados é altamente satisfatório, visto que este é o primeiro trabalho que visa identificar este tipo de marcador em *A. angustifolia* e *P. lambertii*. Porém, a frequência encontrada de marcas SNP para *P. lambertii* (0,01 SNP/kb) e *A. angustifolia* (0,02 SNP/kb) é extremamente baixa em comparação com outras espécies de angiospermas e coníferas. Estudos prévios em coníferas revelam que a frequência de SNP por kb varia de 3 SNP/Kb em *Pinus radiata* até 26 SNP/Kb em *Pseudotsuga menziesii*. Em angiospermas, esta frequência pode chegar até 60 SNP/Kb em *Populus tremula* (SEBASTIANI et al. 2007). A baixa frequência de SNPs no genoma das plantas e principalmente nas regiões expressas pode indicar uma elevada conservação destes genomas e também estabilidade ambiental no sentido da baixa eficiência frequência de fatores causadores de estresse, tais como salinidade, seca, patógenos, entre outros. As amostras analisadas foram coletadas de um ambiente natural das espécies aparentemente conservado, com exceção da presença do gado. De outro lado, pode-se levantar a hipótese de que o genoma das duas espécies aqui avaliadas resistem a possíveis modificações potencialmente causadas pelos fatores de estresse. Portanto, a baixa frequência de mutações pode ser reflexo desta estabilidade ambiental e a relação SNP/kb pode atuar como um indicador para tal situação.

A relação entre o número de transições e transversões obtidas (Ti/Tv) foi de 1,05 para *P. lambertii* e de 2,15 para *A. angustifolia*. O número de transições é geralmente maior que o de transversões, pois a substituição de bases nitrogenadas de mesma estrutura molecular é um evento mutacional muito mais provável de ocorrer do que uma substituição de uma base púrica por uma pirimídica e vice-versa. Além disso, as transições são mutações que na grande maioria as vezes não alteram a fase de leitura dos aminoácidos na síntese proteica. Plantas de *Arabidopsis thaliana* cultivadas em condições controladas apresentaram Ti/Tv de aproximadamente 2,5. Já plantas da mesma espécie cultivadas sob condições de estresse ambiental apresentaram Ti/Tv de 1,6. Neste

caso, o número de transversões de bases aumenta significativamente a ponto de reduzir o valor da relação (JIANG et al. 2014). Em *P. lambertii* esta relação quase equalitária entre as transições e transversões pode ser reflexo número de bases analisadas até o momento, que é três vezes menor do que o número analisado para *A. angustifolia*.

As sequências SSR estão distribuídas em todas as regiões do genoma dos eucariotos, incluindo as regiões codificantes e não codificantes, como *introns*, espaçadores inter-gênicos e regiões não traduzidas (UTRs). Devido a abundância destas marcas, co-dominância, padrão mendeliano de segregação, facilidade operacional e altíssima reprodutibilidade, este marcador molecular é amplamente utilizado em análise de diversidade genética, parentesco, *fingerprinting* e mapas genéticos de ligação (VARSHENEY et al. 2005). No entanto, é necessária a caracterização molecular das sequências flanqueadoras destes motivos repetitivos antes da sua real utilização. O BLASTX das sequências contíguas apresentando motivos SSR revelou a origem dos transcritos possuidores de marcadores putativos, contudo a localização dos motivos SSR dentro destes transcritos ainda não foi realizada, sendo de grande importância para o entendimento da funcionalidade dos motivos dentro das regiões codantes. Estes motivos podem estar situados nas regiões codantes dos genes, podendo ser empregados para a detecção de variantes funcionais.

Em plantas, o número de transcritos que apresentam sequências SSR estimado 10 anos atrás, estava está entre 1% a 5% (VARSHENEY et al. 2005). Neste estudo, 12,70% das sequências contíguas de *A. angustifolia* e 12,71% das sequências contíguas de *P. lambertii* apresentaram sequências microssatélites. Em *Araucaria* foi encontrado 0,17 SSR/kb, já em *P. lambertii* a frequência foi de 0,16 SSR/kb, evidenciando um padrão muito semelhante deste tipo de marcador nas duas espécies. O valor de 0,15 SSR/kb foi encontrado para *Pinus halepensis* por Pinosio et al. 2014. Estas frequências são baixas comparadas aos valores encontrados por Ueno et al. (2012), que identificaram de 0,6 a 1,1 motivos SSR/kb em outras sete coníferas. Já em *Pinus dabeshanensis* a densidade de marcas SSR em sequências transcritas foi de 0,04 SSR/kb (XIANG et al. 2015), portanto, mais baixas que as duas espécies estudadas no presente trabalho. Com o avanço na obtenção de dados, os parâmetros generalizados para plantas podem ser alterados, como foi a estimativa de percentagem de transcritos que apresentam sequências SSR. Uma possível explicação para as distintas frequências entre as coníferas analisadas pode estar nos diferentes

critérios utilizados pelos autores para a busca dos motivos SSRs nos dados transcriptômicos.

Comparando as frequências encontradas dos marcadores, observa-se que as marcas SSRs são muito mais frequentes que as marcas SNPs. O número de marcadores encontrados para *A. angustifolia* é superior em relação ao encontrado para *P. lambertii* em função do maior número de amostras sequenciadas e conseqüentemente maior número de transcritos obtidos. Dos SSRs putativos identificados para *A. angustifolia*, 19 são de origem nuclear, 11 provenientes do genoma plastidial e 3 transcritos no genoma mitocondrial. Somente em 2 seqüências não foi possível a identificação via BLASTX. Em *P. lambertii* não foi possível a identificação da origem de apenas um transcrito. Dos restantes, 8 SSRs são de origem nuclear, 3 de origem mitocondrial e 9 de origem plastidial. Os 9 SSRs plastidiais caracterizados neste trabalho para *P. lambertii* já foram previamente identificados no sequenciamento do genoma plastidial da espécie (VIEIRA et al. 2014), sendo o presente trabalho o primeiro a identificar a transcrição destas marcas SSRs.

O número de marcas SSR identificadas nos transcritos poderia ter sido maior, no entanto a busca pelos motivos repetitivos foi efetuada com a utilização de seqüências com tamanho mínimo de 600 pb. Esta metodologia foi empregada pois seqüências maiores possuem maior probabilidade de terem iniciadores desenhados em relação a seqüências curtas. Nos transcriptomas do presente estudo, realizando a restrição de tamanho, foi possível o desenho de iniciadores em 56% dos SSRs. Xiang et al. (2015) identificaram 1.966 SSRs no transcriptoma de *P. dabeshanensis*, no entanto a procura dos motivos SSRs foi realizada em todos os transcritos sequenciados, não havendo restrição de tamanho. Do total de 1.966 SSRs somente em 431 (21,93%) os autores conseguiram realizar o desenho dos iniciadores. No transcriptoma da angiosperma *Raphanus sativus* foram 1.4641 motivos SSR, mas o desenho de iniciadores somente foi possível em 2.511 (17,5%) dos casos (WANG et al. 2012). Resultado semelhante ocorreu no transcriptoma de *Hevea brasiliensis*, onde de 17.297 motivos foram desenhados apenas 1.709 (9,88%) primers SSR (MANTELLO et al. 2014).

Marcadores SSR são de extrema importância para análises de variabilidade genética, como análises de fluxo gênico, parentesco e filogeografia. Ainda não foram desenvolvidos marcadores SSR específicos para *P. lambertii*. A espécie conta apenas com 7 marcadores SSRs transferidos de *Podocarpus sellowii* (DANTAS et al. 2015). Para *A. angustifolia* existem somente 35 marcadores SSR caracterizados para a

espécie (SALGUEIRO et al. 2005; SCHMIDT et al. 2007), sendo que diversos trabalhos ainda utilizam marcadores desenvolvidos para outras espécies do gênero *Araucaria* em função do baixo polimorfismo dos marcadores espécie-específicos (BITTENCOURT e SEBBENN, 2007; CRISTOFOLINI 2013, SANT'ANNA et al. 2013).

O predomínio de marcas SSR mononucleotídicas e tetranucleotídicas nos transcriptomas deste estudo não vai ao encontro aos padrões encontrados para outras espécies de coníferas. Nos transcriptomas de *P. pinaster*, *P. tabuliformes* e *P. dabeshanensis*, a grande maioria de marcadores identificados foram dos tipos trinucleotídeos seguidos pelos motivos dinucleotídeos (FERNÁNDEZ-POZO et al. 2011; NIU et al. 2013; CANALES et al. 2014; XIANG et al. 2015).

Microssatélites provenientes de sequências transcritas geralmente apresentam menor polimorfismo em relação a marcadores localizados em regiões não codificantes. Uma explicação para este fato é a menor pressão de seleção sofrida pelas regiões não codificantes e a alta conservação das regiões codificantes (CHABANE et al. 2005). Contudo, a caracterização genética destes marcadores moleculares em populações de *A. angustifolia* e *P. lamberti* fornecerão informações que permitirão a escolha dos SSRs mais polimórficos e que apresentem padrão mendeliano de segregação (HEDRICK, 2000). O uso de marcadores de sequências transcritas, pode apresentar utilidade em estudos evolutivos ou de adaptação, bem como de identificação de genótipos ou populações para certas finalidades (ex: concentração de certas substâncias de importância econômica). Além disso, a elevada conservação das sequências transcritas pode permitir a transferibilidade destas marcas para outras espécies dos gêneros *Araucaria* e *Podocarpus*, permitindo a realização de estudos comparativos entre espécies do mesmo gênero. Além de *P. lambertii*, existem outras oito Podocarpaceas nativas do Brasil que compartilham algumas áreas de ocorrência natural, com exceção de *R. piresii* que possui distribuição restrita no estado de Rondônia. Já *Araucaria araucana*, outra espécie da família Araucariaceae na América do Sul, não compartilha áreas de ocorrência com *A. angustifolia*. Dentre as espécies de Podocarpaceas que co-habitam as mesmas regiões e as Araucariaceas que não o fazem, os marcadores transferidos, caso apresentem variação funcional, poderiam ser utilizados para a realização de comparações destas variações funcionais em ambientes distintos ou não.

A utilização dos dados de RNA-Seq para o descobrimento de marcadores moleculares como SSRs e SNPs se mostrou altamente viável,

evidenciando a versatilidade da técnica como em outros estudos realizados em coníferas (PARCHMAN et al. 2010, XIANG et al. 2015, PINOSIO et al. 2014). O número de marcas moleculares SNPs identificadas é semelhante ao de experimentos que visam exclusivamente a identificação de marcadores a partir de NGS de DNA genômico (ZALAPA et al. 2012). O número de potenciais SNPs seria maior caso tivessem sido sequenciados e comparados mais indivíduos das duas espécies. No entanto, o número de marcas SSRs identificadas foi semelhante a trabalhos tradicionais com bibliotecas genômicas de DNA e sequenciamento Sanger (KLABUNDE et al. 2014, SANTOS et al. 2008, SCHMIDT et al. 2007). Comparativamente ao sequenciamento, a preparação destas bibliotecas demanda tempo e intensivo trabalho de bancada. Além do mais, o custo envolvido na preparação destas bibliotecas se assemelha ao custo de uma corrida de sequenciamento em uma plataforma de NGS. Embora já identificados no presente trabalho, os novos marcadores SSRs e os primeiros SNPs ainda precisam ser validados em 80 indivíduos pertencentes a quatro populações naturais destas espécies para poderem ser utilizados em estudos de diversidade genética. As regiões contendo as marcas SSRs são amplificadas via PCR e genotipadas em função das diferenças de tamanho dos fragmentos. Já as marcas SNPs são amplificadas via PCR e sequenciadas bidirecionalmente com utilização da química Sanger.

A relação entre o número de substituições (SNP) e entre o número de inserções ou deleções (S/I) foi de 0,31 para *P. lambertii* e de 0,52 para *A. angustifolia*. Entre os diferentes tipos as sequências transcritas, o valor de S/I identificado para a maioria das espécies analisadas foi pelo menos maior que 1,0 (CHEN et al. 2009). Quanto menor esta relação, maior o número de inserções ou deleções identificadas. Eventos mutacionais que podem causar drásticas consequências na fase de leitura durante o processo de transcrição, tem menor chance de ser acumulados nas gerações seguintes. Já alguns dos SNPs em sequências transcritas possuem a vantagem de não alterar a fase de leitura dos códons (*frameshift mutations*), somente nos casos em que a mutação codifique para o mesmo aminoácido, de acordo com a redundância do código genético.

## 8 CONCLUSÕES

Este é o primeiro estudo de transcriptômica em larga escala realizado no Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal, sendo as conclusões são muito significativas. Inicialmente, a realização deste trabalho permitiu a internalização de muitos procedimentos moleculares e de bioinformática, isto significa que estudos futuros que necessitem tais ferramentas demandarão menos esforço.

A metodologia utilizada permitiu que a montagem dos transcriptomas de *A. angustifolia* e *P. lambertii* com a utilização de leituras curtas (81,71 pb) e método de alinhamento de novo foi possível mesmo sem a presença dos genomas de referência das próprias espécies. Neste sentido, mesmo sem o genoma disponível, foi possível conhecer parte do mesmo que estava expresso nas amostras coletadas. Por se tratar de um estudo realizado com amostras coletadas em ambiente de ocorrência natural das espécies foram constatadas divergências entre as replicatas biológicas, que devem refletir diversos efeitos, como efeitos epigenéticos, em razão da origem ambiental e a cobertura dos sequenciamentos em função do elevado tamanho e complexidade dos genomas.

Certamente a descoberta de milhares de unigenes sequenciados representam uma relevante coleção de recursos genômicos de *A. angustifolia* e *P. lambertii*, que servirão como base para a realização de diversos ensaios moleculares. Adicionalmente, este estudo proporcionou ainda identificar novos marcadores moleculares, como os EST-SSR putativos, não reduntantes e com iniciadores desenhados, bem como os primeiros SNPs para *Araucaria angustifolia* e *Podocarpus lambertii*. Novamente, além do avanço no conhecimento da expressão gênica, estes novos marcadores são extremamente importantes e, de certa forma, esperados para a realização de futuros estudos populacionais associados a evolução, diversidade genética neutra, quase neutra e mesmo funcional, para fins de uso (ex: melhoramento genético) e conservação das espécies.

Dentre as descobertas, é relevante destacar a baixa frequência de marcas SNP comparativamente a outras espécies. Embora, o presente estudo não atinja por completo a espécie, o fato pode ser um forte indicio de elevada conservação dos genomas e principalmente dos genes destas duas espécies. Mais do que isso, poderia desencadear a realização de outros estudos relacionados aos efeitos diferenciais de fatores adversos.

Por fim, os resultados obtidos também contribuíram para o avanço científico relacionado a composição do genoma codificante e

parâmetros a ele relacionados. Com os dados obtidos, certos intervalos de variação de índices (ex: porcentagem de transcritos que apresentam sequências SSR) deverão ser revistos em futuros artigos de revisão.

## 9 PERSPECTIVAS

Os dados gerados neste estudo permitem a realização de novas análises, tais como a caracterização inicial de mecanismos epigenéticos como os miRNA, caracterização dos fatores de transcrição e análises filogenéticas de sequências de genes geradas nos transcriptomas em comparação com transcriptomas de outras coníferas e angiospermas.

Uma etapa importante e ainda não efetuada do fluxo de atividades das análises transcriptômicas são as validações dos quatro transcriptomas gerados neste trabalho via PCR quantitativa em tempo real (qPCR). Serão utilizados os genes de referência AaEIF4B-L, AaPP2A e outros quatro genes expressos constitutivamente e já avaliados para este fim por Elbl et al. (2015b), a partir dos dados dos primeiros transcriptomas da espécie (ELBL et al. 2015a).

Assim como foi realizado para os marcadores SSRs putativos, será realizada a anotação funcional das primeiras marcas SNPs geradas para *A. angustifolia* e *P. lambertii*. Esta etapa permitirá realizar a escolha de marcas que possam ser utilizadas em estudos funcionais, de diversidade genética e fluxo gênico.

Os dados brutos de RNA-Seq deste trabalho serão depositados na base de dados do NCBI, via *Sequence Read Archive* (SRA) para utilização futura por outros estudantes e pesquisadores. O cadastro do projeto deste estudo já foi realizado e atende pelo nome de “BioProject ID: PRJNA308171” para as amostras de *A. angustifolia* e “PRJNA311154” para a amostra de *P. lambertii*. As sequências contendo os motivos SSRs e SNPs putativos também serão submetidas nas bases de dados de nucleotídeos e SNPs, respectivamente.

Adicionalmente aos transcriptomas deste estudo, novos perfis de expressão gênica podem ser gerados, visando aumentar o número de genes sequenciados e compreender as peculiaridades de outros tecidos reprodutivos e vegetativos de *A. angustifolia* e *P. lambertii*, tais como tecidos embrionários (imatuross, maduros e em processo germinativo) e caulinares (floema, felogênio e feloderme).

## 10 BIBLIOGRAFIA

AHUJA, M.R. 2001. Recent advances in molecular genetics of forest trees. *Euphytica* 121:173-195.

AHUJA, M.R. e NEALE D.B. 2005. Evolution of genome size in conifers. *Silvae Genetica* 54:126-137.

AHN, Y.K.; TRIPATHI, S.; CHO, Y.I.; KIM, J.H.; LEE, H.E.; KIM, D.S.; GYU, J.; WOO, J.G. e CHO, M.C. 2013. De novo transcriptome assembly and novel microsatellite information in *Capiscum annum* varietis Saengryeg 211 and Saengryeg 213. *Botanical Studies* 54:58.

AMBORELLA GENOME PROJECT. 2013. The Amborella Genome and the Evolution of Flowering Plants. *Science* 342.

AULER, N.M.F.; REIS, M.S.; GUERRA, M.P. e NODARI, R.O. 2002. The genetics and conservation of *Araucaria angustifolia*: I. Genetic structure and diversity of natural populations by of means of non-adaptive variation in the state of Santa Catarina. *Genetics and Molecular Biology* 25:329–338.

ASHBURNER, M.; BALL, C.A.; BLAKE, J.A.; BOTSTEIN, D.; BUTLER, H.; CHERRY, J.M.; DAVIS, A.P.; DOLINSKI, K.; DWIGHT, S.S.; EPPIG, J.T.; HARRIS, M.A.; HILL, D.P.; ISSEL-TARVER, L.; KASARSKIS, A.; LEWIS, S.; MATESE, J.C.; RICHARDSON, J.E.; RINGWALD, M.; RUBIN, G.M. e SHERLOCK, G. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* 25:25-29.

ARCHER, S.; SCIFRES, C. e BASSHAM, C.R. 1988. Autogenic succession in a subtropical savanna: conversion of grassland to thorn woodland. *Ecological Monographs* 58:111–127.

BABU, S.C.V. e GRASSMANN, M. 2011. Assessing integrity of plant RNA with the Agilent 2100 Bioanalyzer. Disponível em [www.agilent.com/genomics/RNA](http://www.agilent.com/genomics/RNA) , acesso em 28 de agosto de 2015.

BEHLING, H.; DE PATTA-PILLAR, V.; ORLÓCI, L. e BAUERMANN, S.G. 2004. Late Quaternary *Araucaria* forest, grassland (Campos), fire and climate dynamics, studied by high-resolution pollen, charcoal and



multivariate analysis of the Camará do Sul core in southern Brazil. *Palaeogeography, Palaeoclimatology, Palaeoecology* 203:277-297.

BEHLING, H.; DE PATTA-PILLAR, V. e BAUERMANN, S.G. 2005. Late Quaternary grassland (Campos), gallery forest, fire and climate dynamics, studied by pollen, charcoal and multivariate analysis of the São Francisco de Assis core in western Rio Grande do Sul (southern Brazil). *Review of Paleobotany & Palinology* 133:235-248.

BERNARDI, A. 2015. Estrutura genética e demográfica de *Podocarpus lambertii* KLOTZCH EX. ENDL. em uma paisagem de campo. Dissertação (Mestrado em Ciências), UFSC, Florianópolis, SC, 128 p.

BIROL, I.; RAYMOND, A.; JACKMAN, S.D.; PLEASANCE, S.; COOPE, R.; TAYLOR, G.A.; KEELING, C.I.; BRAND, D.; VANDERVALK, B.P.; KIRK, H.; PANDOH, P.; MOORE, R.A.; ZHAO, Y.; MUNGALL, A.J. JAQUISH, B.; YANCHUK, A.; RITLAND, C.; BOYLE, B.; BOUSQUET, J.; RITLAND, K.; MACKAY, J.; BOHLMANN, J. e JONES, S.J. 2013. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics* 29:1492-1497.

BITTENCOURT, J.V.M. e SEBBENN, A.M. 2007. Patterns of pollen and seed dispersal in a small, fragmented population of the wind-pollinated tree *Araucaria angustifolia* in southern Brazil. *Heredity* 99:580–591.

BLANCA, J.; CAÑIZARES, J.; ROIG, C.; ZIASOLO, P.; NUEZ, F. e PICÓ, B. 2011. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genomics* 12:104.

BOAVIDA, L.C.; BECKER, J.D. e FEIJÓ, J.A. 2005. The making of gametes in higher plants. *The International Journal of Developmental Biology* 49:595-614.

CANALES, J.; BAUTISTA, R.; LABEL, P.; GÓMEZ-MALDONADO, J.; LESUR, I.; FERNÁNDEZ-POZO, N.; RUEDA-LÓPEZ, M.; GUERRERO-FERNÁNDEZ, D.; CASTRO-RODRÍGUEZ, V.; BENZEKRI, H.; CAÑAS, R.A.; GUEVARA, M.A.; RODRIGUES, A.; SEOANE, A.; TEYSSIER, C.; MOREL, A.; EHRENMANN, F.; LE

PROVOST, G.; LALANNE, C.; NOIROT, C.; KLOPP, C.; REYMOND, I.; GARCÍA-GUTIÉRREZ, A.; TRONTIN, J.F.; LELU-WALTER, M.A.; MIGUEL, C.; CERVERA, M.T.; CANTÓN, F.R.; PLOMION, C.; HARVENGT, L.; AVILA, C.; CLAROS, M.G. e CÁNOVAS, F.M. 2014. De novo assembly of maritime pine transcriptome: implications for forest breeding and biotechnology. *Plant Biotechnology Journal* 12:286–299.

CAÑAS, R.A.; CANALES, J.; GÓMEZ-MALDONADO, J.; ÁVILA, C. e CÁNOVAS, F.M. 2014. Transcriptome analysis in maritime pine using laser capture microdissection and 454 pyrosequencing. *Tree Physiology* 34(11):1278-1288.

CARVALHO, A.; PAIVA, J.; LOUZADA, J. e LIMA-BRITO, J. 2013. The transcriptomics of secondary growth and wood formation in conifers. *Molecular Biology International* 2013:12.

CENTRO NACIONAL DE CONSERVAÇÃO DA FLORA – CNCFLORA. 2015. Mapa de distribuição de *Podocarpus lambertii* Klotzsch ex Endl. Disponível em <[www.http://cncflora.jbrj.gov.br/portal/pt-br/profile/Podocarpus%20lambertii](http://cncflora.jbrj.gov.br/portal/pt-br/profile/Podocarpus%20lambertii)>. Acesso em 07/09/2015.

CHABANE, K.; ABLETT, G.A.; CORDEIRO, G.M.; VALKOUNN, J. e HENRY, R.J. 2005. EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. *Genetic Resources and Crop Evolution* 52:903-909.

CHAW, S.-M.; SHIH, A.C.-C.; WANG, D.; WU, Y.-W.; LIU, S.-M. E CHOU, T.-C. 2008. The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. *Molecular Biology and Evolution* 25:603-615.

CHEN, J.Q.; WU, Y.; YANG, H.; BERGELSON, J.; KREITMAN, M. e TIAN, D. 2009. Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. *Molecular Biology and Evolution* 26(7):1523-1531.

CHEN, J.; UEBBING, S.; GYLLENSTRAND, N. e LAGERCRANTZ, U. 2012. Sequencing of the needle transcriptome from Norway spruce

(*Picea abies* Karst L.) reveals lower substitution rates, but similar selective constraints in gymnosperms and angiosperms. *BMC Genomics* 13:589.

CHIAMOLERA, L.B.; ÂNGELO, A.C. e BOEGER, M.R. 2010. Resposta morfoanatômica da folha de *Podocarpus lambertii* Klotzsch ex Endl (Podocarpaceae) implantado em duas áreas com diferentes graus de sucessão às margens do Reservatório Iraí – PR. *Biotemas* 23:1-11.

CHOUDHARY, O.M. e TRIVEDI, S. 2010. Microsatellite or Simple Sequence Repeat (SSR) instability depends on repeat characteristics during replication and repair. *Journal of Cell and Molecular Biology* 8(2):21-34.

CONESA, A.; GÖTZ, S.; GARCÍA-GÓMEZ, J.M.; TEROL, J.; TALÓN, M. e ROBLES, M. 2005. Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674-3676.

CONSELHO ESTADUAL DO MEIO AMBIENTE – CONSEMA. 2014. Lista Oficial das Espécies da Flora Ameaçada de Extinção no Estado de Santa Catarina. RESOLUÇÃO CONSEMA Nº 51, de 05 de dezembro de 2014. p.12.

COX, C.B. e MOORE, P.D. 2009. Biogeografia: uma abordagem ecológica e evolucionária. 7 ed. Rio de Janeiro. LTC. 398p.

CRISTOFOLINI, C. 2013. Dinâmica da diversidade genética em *Ararucaria angustifolia* (Bertol.) Kuntze em paisagem de campo no estado de Santa Catarina. Dissertação (Mestrado em Ciências), UFSC, Florianópolis, SC, 93p.

DANTAS, L.G.; ESPOSITO, T.; SOUZA, A.C.B.; FÉLIX, L.; AMORIM, L.L.B.; BENKO-ISEPPON, A.M.; BATALHA-FILHO, H. e PEDROSA-HARAND, A. 2015. Low genetic diversity and high differentiation among relict populations of the neotropical gymnosperm *Podocarpus sellowii* (Klotz.) in the Atlantic forest. *Genetica* 143:21-30.

DAVIES, K. 2010. The \$1,000 Genome: The Revolution in DNA Sequencing and the New Era of Personalized Medicine. 1 ed. New York.

Free Press. 342p.

DOYLE, J.A. 1998. Phylogeny of vascular plants. *Annual Review of Ecology, Evolution and Systematics* 29:567–99.

DUARTE, L.S. e DILLENBURG, L.R. 2000. Ecophysiological responses of *Araucaria angustifolia* (Araucariaceae) seedlings to different irradiance levels. *Australian Journal of Botany* 48:531-537.

DUARTE, L.S.; DOS SANTOS, M.M.G.; HARTZ, S.M. e PILLAR, V.D. 2006. Role of nurse plants on *Araucaria* Forest expansion over grassland in south Brazil. *Austral Ecology* 31:520-528.

ELBL, P.; LIRA, B.S.; ANDRADE, S.C.S.; JO, L.; WENDT dos SANTOS, A.L.; COUTINHO, L.L.; FLOH, E.I.S. e ROSSI, M. 2015a. Comparative transcriptome analysis of early somatic embryo formation and seed development in Brazilian pine, *Araucaria angustifolia* (Bertol.) Kuntze. *Plant Cell, Tissue Organ Culture* 120:903-915.

ELBL, P.; NAVARRO, B.V.; de OLIVEIRA, L.F.; ALMEIDA, J.; MOSINI, A.C.; dos SANTOS, A.L.W.; ROSSI, M. e FLOH, E.I.S. 2015b. Identification and evaluation of reference genes for quantitative analysis of Brazilian pine (*Araucaria angustifolia* Bertol. Kuntze) gene expression. *Plos One* 10(8).

ESCAPA, I.H. e CATALANOT, S.A. 2013. Phylogenetic Analysis of Araucariaceae: Integrating molecules, morphology, and fossil. *International Journal of Plant Science* 174(8):1153-1170.

FARJON, A. 2008. A natural history of conifers. Timber Press Inc. Londres, Reino Unido. 720p.

FARJON, A. 2010. A handbook of the world's conifers. Volume 1. Koninklijke Brill, Leiden, Holanda. 1.111p.

FARJON, A. 2013. *Podocarpus lambertii*. The IUCN Red List of Threatened Species. Version 2014.3. <[www.iucnredlist.org](http://www.iucnredlist.org)>. Acesso em 05 fevereiro de 2015.

FERNÁNDEZ-POZO, N.; CANALES, J.; GUERRERO-FERNÁNDEZ,

D.; VILLALOBOS, D.P.; DÍAZ-MORENO, S.M.; BAUTISTA, R.; FLORES-MONTERROSO, A.; GUEVARA, M.Á.; PERDIGUERO, P.; COLLADA, C.; CERVERA, M.T.; SOTO, Á.; ORDÁS, R.; CANTÓN, F.R.; AVILA, C.; CÁNOVAS, F.M. e CLAROS, M.G. 2011. EuroPineDB: a high-coverage web database for maritime pine transcriptome. *BMC Genomics* 12:366.

FONTOURA, S.B.; GANADE, G. e LAROCCA, J. 2006. Changes in plant community diversity and composition across an edge between Araucaria forest and pasture in South Brazil. *Revista Brasileira de Botânica* 29:79–91.

FRANCO, A.M.S. e DILLENBURG, L.R. 2007. Ajuste morfológico e fisiológico em plantas jovens de *Araucaria angustifolia* (Bertol.) Kuntze em resposta ao sombreamento. *Hoehnea* 34:135-144.

GAO, Z.; LUO, W.; LIU, H.; ZENG, C.; LIU, X.; YI, S. e WANG, W. 2012. Transcriptome analysis and SSR / SNP markers information of the blunt snout bream (*Megalobrama amblycephala*). *PLOS ONE* 7(8):1-10.

GARDNER, M. 2013. *Retrophyllum piresii*. The IUCN Red List of Threatened Species. Version 2014.3. <[www.iucnredlist.org](http://www.iucnredlist.org)>. Acesso em 06 fevereiro de 2015.

GARG, R.; PATEL, R.K.; TYAGI, A.K. e JAIN, M. 2011. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research* 18:53-63.

GOLD - Genomes Online Database <<https://gold.jgi-psf.org/>> Acesso em 13 de junho de 2015.

GOLTE, W. 1993. *Araucaria*: Verbreitung und Standortansprüche einer Coniferengattung in vergleichender Sicht. Franz Steiner, Stuttgart, 167 p.

GÖTZ, S.; GARCÍA-GOMEZ, J.M.; TEROL, J.; WILLIAMS, T.D.; NAGARAJ, S.H.; NUEDA, M.J.; ROBLES, M.; TALÓN, M.; DOPAZO, J. e CONESA, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 36(10):3420-3435.

GREEN, B. 2011. Chloroplast genomes of photosynthetic eukaryotes. *Plant Journal* 66:34-44.

GUERRA, M.P.; SILVEIRA, V.; SANTOS, A.L.W.; ASTARITA, L. e NODARI, R.O. 2000. Somatic Embryogenesis in *Araucaria angustifolia* (Bert) O. Ktze.. In: JAIN, S.M.; GUPTA, P.K. e NEWTON, R.J. (Org.). *Somatic Embryogenesis in Woody Plants*. Dordrecht: Kluwer Academic Publishers, p. 457-478.

GUERRA, M.P.; SILVEIRA, V. e REIS, M.S. 2002. Exploração, Manejo e Conservação da Araucária (*Araucaria angustifolia*). In: Luciana Lopes Simões; Clayton Ferreira Lino. (Org.). *Manejo Sustentável da Mata Atlântica: A Exploração de seus Recursos Florestais*. São Paulo: Senac, p. 85-101.

GUERRA, M. P.; STEINER, N.; MANTOVANI, A.; NODARI, R.O.; REIS, M.S. e SANTOS, K.L. 2008. Evolução, ontogênese e diversidade genética em *Araucaria angustifolia*. In: Barbieri, R.L.; Stumpf, E.R.T. (Org.). *Origem e Evolução de Plantas Cultivadas*. Brasília, DF: Embrapa Informação Tecnológica, p. 149-184.

HAO, D.C.; GE, G.B.; XIAO, P.G.; ZHANG, Y.Y. e YANG, L. 2011. The first insight into the tissue specific *Taxus* transcriptome via Illumina second generation sequencing. *PLoS ONE* 6(6).

HEDRICK, P.W. 2000. *Genetics of populations*, 2nd ed. Jonas and Bartlett, Boston, Massachusetts, USA.

HENRY, R.J. 2005. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. 1. Ed. Oxfordshire: CABI Publishing, Oxon, UK, 332p.

HOHEISEL, J.D. 2006. Microarray technology: beyond transcript profiling and genotype analysis. *Nature Reviews Genetics* 7:200-210.

INOUE, M.T.; RODERJAN, C.V. e KUNIYOSHI, Y.S. 1984. Projeto Madeira do Paraná. Sudesul, IBDF, FUPEF, Curitiba.

JEHAN, T. e LAKHANPAUL, S. 2006. Single nucleotide polymorphism (SNP) - Methods and applications in plant genetics: A review. *Indian*

Journal of Biotechnology 5:435-459.

JIANG, C.; MITHANI, A.; BELFIELD, E.J.; MOTT, R.; HURST, L.D. e HARBERD, N.P. 2014. Environmentally responsive genome-wide accumulation of de novo *Arabidopsis thaliana* mutations and epimutations. *Genome Research* 24(11):1821-1829.

JIN, J.P.; ZHANG, H.; KONG, L.; GAO, G.; e LUO, L.C. 2014. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Research* 42(D1):1182-1187.

JUDD, W.S.; CAMPBELL, C.S.; KELLOG, E.A.; STEVENS, P.F. e DONOGHUE, M.J. 2009. Sistemática vegetal um enfoque filogenético. 3. Ed. Porto Alegre, RS: Artmed, 612p.

KELCH, D.G. 1997. The phylogeny of the Podocarpaceae based on morphological evidence. *Systematic Botany* 22(1):113-131.

KELCH, D.G. 1998. Phylogeny of Podocarpaceae: comparison of evidence from morphology and 18S rDNA. *American Journal of Botany* 85:986-996.

KLABUNDE, G.H.F. 2012. Análise filogeográfica entre populações de *Araucaria angustifolia* (Bert.) O. Kuntze em sua área de distribuição natural. Dissertação (Mestrado em Ciências), UFSC, Florianópolis, SC, 85 p.

KLABUNDE, G.H.F.; OLKOSKI, D.; VILPERTE, V.; ZUCCHI, M.I. e NODARI, R.O. 2014. Characterization of 10 new nuclear microsatellite markers in *Acca sellowiana* (Myrtaceae). *Applications in Plant Sciences* 2014:2(6).

KLEIN, R.M. 1960. Aspecto dinâmico do pinheiro brasileiro. *Sellowia* 12:17-44.

KLEIN, R.M. 1978. Mapa Fitogeográfico do Estado de Santa Catarina. In: Reitz, R. (ed.) *Flora Ilustrada Catarinense*, V Parte – Mapa Fitogeográfico. Itajaí: Herbário Barbosa Rodrigues, 24 p.

KRANITZ, M.L., BIFFIN, E., CLARK, A., HOLLINGSWORTH, M.L.,

RUHSAM, M., GARDNER, M.F., THOMAS, P., MILL, R.R., ENNONS, R.A., GAUDEUL, M. LOWE, A.J. e HOLLINGSWORTH, P.M. 2014. Evolutionary Diversification of New Caledonian Araucaria. Plos One 9(10).

KRASSILOV, V.A. 1974. Podocarpus from the upper cretaceous of eastern Asia and its bearing on the theory of conifer evolution. Palaeontology 17:365–370.

LANDER, E.S. et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.

LEDRU, M.P.; SALATINO, M.L.F.; CECCANTINI, G.; SALATINO, A.; PINHEIRO, F. e PINTAUD, J.C. 2007. Regional assessment of the impact of climatic change on the distribution of a tropical conifer in the lowlands of South America. Diversity and Distributions 13:761-771.

LI, B.; FILLMORE, N; BAI, Y.; COLLINS, M.; THOMSON, J.A.; STEWART, R. e DEWEY, C.N. 2014. Evaluation of de novo transcriptome assemblies from RNA-Seq data. BioRxiv preprint.

LI, Y-C.; KOROL, A.B.; FAHIMA, T.; BEILES, A. e NEVO, E. 2002. Microsatellites: genomic distribution, putative functions and mutacional mechanisms: a review. Molecular Ecology 11: 2453-2465.

LIU, J.J.; STURROCK, R.N. e BENTON, R. 2013. Transcriptome analysis of Pinus monticola primary needles by RNA-Seq provides novel insight into host resistance to Cronartium ribicola. BMC Genomics 14:884.

LODISH, H.; BERK, A.; ZIPURSKY, S.L.; MATSUDAIRA, P.; BALTIMORE, D. e DARNELL, J. 2000. Molecular cell biology, seção 11.6: Processing of rRNA and tRNA. 4. Ed. New York: W.H. Freeman.

LONGHI, S.J.; BRENA, D.A.; RIBEIRO, S.B.; GRACIOLI, C.R.; LONGHI, R.V. e MASTELLA, T. 2010. 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.



LORENZ, W.W.; YU, Y.-S. e DEAN, J.F.D. 2010. An improved method of RNA isolation from Loblolly pine (*P. taeda* L.) and other conifer species. *Journal of Visualized Experiments* 36.

LORENZI, H. 2008. Árvores brasileiras: Manual de identificação e cultivo de plantas arbóreas nativas do Brasil. 5. Ed. Nova Odessa, SP: Instituto Plantarum, 384p.

MACKAY, J. e DEAN, J.F.D. 2011. Transcriptomics. In: PLOMION, C.; BOUSQUET, J. e KOLE, C. (Eds.). *Genetics, Genomics and Breeding of Conifers Trees*. Edenbridge Science Publishers and CRC Press, New York.

MACKAY, J.; DEAN, J.F.D.; PLOMION, C.; PETERSON, D.G.; CÁNOVAS, F.M.; PAVY, N.; INGVARSON, P.K.; SAVOLAINEN, O.; GUEVARA, M.A.; FLUCH, S.; VINCENTI, B.; ABARCA, D.; DÍAZ-SALA, C.; CERVERA, M.T. 2012. Towards decoding the conifer genome. *Plant Molecular Biology* 80:555-569.

MACRAE, E. 2007. Extraction of plant RNA. In: HILARIO, E. e MACKAY, J.F. (Eds.). *Protocols for Nucleic Acids by Nonradioactive Probes*. Totowa, EUA. Humana Press Inc. 10 p.

MAGALLÓN, S. e SANDERSON, M.J. 2005. Angiosperm divergence times: The effect of genes, codon positions, and time constraints. *Evolution* 59:1653–1670.

MAIA, L.C.D.; PALMIERI, D.A.; SOUZA, V.Q.D.; KOPP, M.M.; CARVALHO, F.I.F.D. e OLIVEIRA, A.C.D. 2008. SSR Locator: Tool for Simple Sequence Repeat discovery integrated with primer design and PCR simulation. *International Journal of Plant Genomics* 2008:1-9.

MAINERI, C. e PIRES, J.M. 1973. O gênero *Podocarpus* no Brasil. *Silvicultura* 8:1-24.

MAIXNER, A.E. e FERREIRA, L.A.B. 1978. Contribuição ao estudo das essências florestais e frutíferas nativas no Estado do Rio Grande do Sul. *Trigo e Soja*, Porto Alegre, n.28, p.2-31.

MANTELLO, C.C.; CARDOSO-SILVA, C.B.; da SILVA, C.C.; de

SOUZA, L.M.; JUNIOR, E.J.S.J.; GONÇALVES, P.S. e VINCENTINI, R. 2014. De novo assembly and transcriptome analysis of the rubber tree (*Hevea brasiliensis*) and SNP markers development for rubber biosynthesis pathways. *PLoS ONE* 9:7.

MANTOVANI, A., MORELLATO, L.P.C. e REIS, M.S. 2004. Fenologia reprodutiva e produção de sementes em *Araucaria angustifolia* (Bert.) O. Kuntze. *Revista Brasileira de Botânica* 27:787-796.

MARIN-RAGAGNIN, L.I.; CORRÊA-COSTA, E. e HOPPE, J.M. 1994. Maturidade fisiológica de sementes de *Podocarpus lambertii* Klotzsch. *Ciência Florestal* 4:23-41.

MARTIN, W.; STOEBE, B.; GOREMYKIN, V.; HANSMANN, S.; HASEGAWA, M. e KOWALLIK, K.V. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393:162-165.

MARTIN, J.A. e WANG, Z. 2011. Next-generation transcriptome assembly. *Nature Reviews Genetics* 12:671-682.

MATTOS, J.R. 2011. O pinheiro brasileiro. Florianópolis, Editora UFSC, 700p.

McINTYRE, L.M.; LOPIANO, K.K.; MORSE, A.M.; AMIN, V.; OBERG, A.L.; YOUNG, L.J. e NUZHIDIN, S.V. 2011. RNA-seq: technical variability and sampling. *BMC Genomics* 12:93.

METZKER, L.M. 2010. Sequencing technologies - the next generation. *Nature Reviews Genetics* 11:31-46.

MILLER, C.N. 1988. The origin of modern conifer families. In: Beck, C.B. (Ed.). *Origin and evolution of gymnosperms*, 448-486. Columbia, University Press, New York, NY.

MINISTÉRIO DO MEIO AMBIENTE – MMA. 2014. PORTARIA MMA Nº 443, DE 17 DE DEZEMBRO DE 2014, p. 25.

MORGANTE, M. e de PAOLI, E. 2011. Toward the conifer genome sequence. In: PLOMION, C.; BOUSQUET, J. e KOLE, C. (Eds.). *Genetics, Genomics and Breeding of Conifers Trees*. Edenbridge Science

Publishers and CRC Press, New York.

MORTAVAZI, A., WILLIAMS, B.A., McCUE, K., SCHAEFFER, L. e WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by ma-seq. *Nature Methods* 5(7):621-628.

MURRAY, B.G.; LEITCH, I.J. e BENNETT, M.D. 2010. Gymnosperm DNA C-values database (release 4.0, Dec. 2010) <http://www.kew.org/cvalues/>

NAGALAKSHMI, U.; WANG, Z.; WAERN, K.; SHOU, C.; RAHA, D.; GERSTEIN, M. e SNYDER, M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320 (5881):1344-1349.

NEKRUTENKO, A. e TAYLOR, J. 2012. Next generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nature Reviews Genetics* 13:667-672.

NIU, S.-H.; LI, Z.-X.; YUAN, H.-W.; CHEN, X.-Y.; LI, Y. e LI, W. 2013. Transcriptome characterisation of *Pinus tabulaeformis* and evolution of genes in the *Pinus* phylogeny. *BMC Genomics* 14:263.

NYSTEDT, B.; STREET, N.R.; WETTERBOM, A.; ZUCCOLO, A.; LIN, Y.-C.; SCOFIELD, D.G.; VEZZI, F.; DELHOMME, N.; GIACOMELLO, S.; ALEXEYENKO, A.; VICEDOMINI, R.; SAHLIN, K.; SHERWOOD, E.; ELFSTRAND, M.; GRAMZOW, L.; HOLMBERG, K.; HÄLLMAN, J.; KEECH, O.; KLASSON, L.; KORIABINE, M.; KUCUKOGLU, M.; KÄLLER, M.; LUTHMAN, J.; LYSHOLM, F.; NIITTYLÄ, T.; OLSON, A.; RILAKOVIC, N.; RITLAND, C.; ROSSELLÓ, J.A.; SENA, J.; SVENSSON, T.; TALAVERA-LÓPEZ, C.; THEIBEN, G.; TUOMINEN, H.; VANNESTE, K.; WU, Z.-Q.; ZHANG, B.; ZERBE, P.; ARVESTAD, L.; BHALERAO, R.; BOHLMANN, J.; BOUSQUET, J.; GIL, R.G.; HVIDSTEN, T.R.; DE JONG, P.; MACKAY, J.; MORGANTE, M.; RITLAND, K.; SUNDBERG, B.; THOMPSON, S.L.; DE PEER, Y.V.; ANDERSSON, B.; NILSSON, O.; INGVASSON, P.K.; LUNDEBERG, J. e JANSSON, S. 2013. The norway spruce genome sequence and conifer genome evolution. *Nature* 497:579-584.

OLIVEIRA, Y. M. M. de. e ROTTA, E. 1982. Levantamento da estrutura vertical de uma mata de araucária do primeiro planalto paranaense. CONGRESSO FLORESTAL BRASILEIRO, 4. 1982. Curitiba. Anais. Curitiba: EMBRAPA-URPFCS, p.27-41. (EMBRAPA-URPFCS. Documentos, 10).

O'NEIL, D.; GLOWATZ, H. e SCHLUMPBERGER, M. 2013. Ribossomal RNA depletion for efficient use of RNA-Seq capacity. Current Protocols in Molecular Biology 4:19-1.

OXFORD NANOPORE. 2015. Disponível em: <<https://nanoporetech.com/about-us/summary>> Acesso em 03/02/2015.

PACIFIC BIOSCIENCES. 2015. Disponível em: <<http://www.pacificbiosciences.com/>> Acesso em 03/02/2015.

PAGE, C.N. 1989. New and maintained genera in the conifer families Podocarpaceae e Pinaceae. Royal Botanical Garden. Edinburg 45:377-395.

PALAZZO, A.F. e GREGORY, T.R. 2014. The case for junk DNA. PLoS Genetics 10:5.

PARCHMAN, T.L.; GEIST, K.S.; GRAHNEN, J.A.; BENKMAN, C.W. e BUERKLE, C.A. 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. BMC Genomics 11:180.

PATREZE, C.M. e TSAI, S.M. 2010. Intrapopulacional genetic diversity of *Araucaria angustifolia* (Bertol.) Kuntze is different when assessed on the basis of chloroplast or nuclear markers. Plant Systematics and Evolution 284:111-122.

PAVY, N.; PAULE, C.; PARSONS, L.; CROW, J.A.; MORENCY, M.J.; COOKE, J.; JOHNSON, J.E.; NOUMEN, E.; GUILLET-CLAUDE, C.; BUTTERFIELD, Y.; BARBER, S.; YANG, G.; LIU, J.; SCOTT, J.; KIRKPATRICK, R.; SIDDIQUI, A.; HOLT, R.; MARRA, M.; SEGUIN, A.; RETZEL, E.; BOUSQUET, J. e MACKAY, J. 2005. Generation, annotation, analysis and database integration of 16,500 white spruce EST clusters. BMC Genomics 6:144.

PAVY, N.; JOHNSON, J.J.; CROW, J.A.; PAULE, C.; KUNAU, T.; MACKAY, J. e RETZEL, E.F. 2007. ForestTreeDB: a database dedicated to the mining of tree transcriptomes. *Nucleic Acids Research* 35:888-894.

PILLAR, V. D. 2003. Dinâmica da expansão florestal em mosaicos de floresta e Campos no sul do Brasil. In: *Ecosistemas Brasileiros: Manejo e Conservação*. CLAUDINO-SALES, V. (Org.). Fortaleza, Expressão Gráfica e Editora. 209–216pp.

PINHEIRO, C.C. e GANADE G. 2009 Influência do microhabitat no processo de predação de sementes em uma área degradada. *Neotropical Biology and Conservation* 4:20-27.

PINOSIO, S; GONZÁLEZ-MARTINEZ, S.C.; BAGNOLI, F.; CATTONARO, F.; GRIVET, D.; MARRONI, F.; LORENZO, Z.; PAUSAS, J.G.; VERDÚ, M. e VENDRAMIN, G.G. 2014. First insights into the transcriptome and development of new genomic tools of a widespread circum-Mediterranean tree species, *Pinus halepensis* Mill. *Molecular Ecology Resources* 14:846-856.

PLOMION, C; BOUSQUET, J. e KOLE, C. 2011. *Genetics, Genomics and Breeding of Conifers*. Edenbridge Science Publishers and CRC Press, New York.

PROVAN, J.; POWELL, H. e HOLLINGSWORTH, P.M. 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution* 16(3):142-147.

QUINN, C.J.; PRICE, R.A. e GADER, P.A. 2002. Familial concepts and relationships in the conifers based on *rbcL* and *matK* sequence comparisons. *Kew Bulletin* 57:513-531.

RAGAGNIN, L.I.M.; COSTA, E.C. e HOPPE, J.M. 1994. Maturidade fisiológica de sementes de *Podocarpus lambertii* Klotzsch. *Ciência Florestal* 4:23-41.

RAFINSKA, K.; ZIENKIEWICZ, K. e BEDNARSKA, E. 2010. Pollen transcriptome and proteome: Molecular and functional analysis. *Advances in Cell Biology* 2:29-57.

RAHERISON, E.; RIGAUULT, P.; CARON, S.; POULIN, P.-L.; BOYLE, B.; VERTA, J.-P.; GIGUÈRE, I.; BOMAL, C.; BOHLMANN, J. e MACKAY, J. 2012. Transcriptome profiling in conifers and the PiceaGenExpress database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. *BMC Genomics* 13:434.

RALPH, S.G.; YUEH, H.; FRIEDMANN, M.; AESCHLIMAN, D.; ZEZNIK, J.A.; NELSON, C.C.; BUTTERFIELD, Y.S.; KIRKPATRICK, R.; LIU, J.; JONES, S.J.; MARRA, M.A.; DOUGLAS, C.J. e RITLAND, K. 2006. Conifer defense against insects: microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale changes of the host transcriptome. *Plant Cell & Environment* 29:1545-1570.

RALPH, S.; CHUN, H.; KOLOSOVA, N.; COOPER, D.; ODDY, C.; RITLAND, C.E.; KIRKPATRICK, R.; MOORE, R.; BARBER, S.; HOLT, R.A.; JONES, S.J.M.; MARRA, M.A.; DOUGLAS, C.J.; RITLAND, K. e BOHLMANN, J. 2008. A conifer genomics resource of 200,000 spruce (*Picea* spp) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics* 9:484.

REITZ, R. e KLEIN, R.M. 1966. *Araucariáceas*. *Flora Ilustrada Catarinense*. 62 pp.

REITZ, R.; KLEIN, R.M. e REIS, A. 1978. Projeto Madeira de Santa Catarina. *Sellowia* 28:1-320.

RICE, D.W.; ALVERSON, A.A.; RICHARDSON, A.O.; YOUNG, G.J.; SANCHEZ-PUERTA, M.V.; MUNZINGER, J.; BARRY, K.; BOORE, J.L.; ZHANG, Y.; dePAMPHILIS, C.W.; KNOX, E.B. e PALMER, J.D. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* 342.

ROGGE-RENNER, G.D.; STEINER, N.; SCHMIDT, E.C.; BOUZON, Z.L.; FARIAS, F.L. e GUERRA, M.P. 2013. Structural and component characterization of meristema cells in *Araucaria angustifolia* (Bert.) O.

Kuntze zygotic embryo. *Protoplasma* 250:731-739.

ROSADO, R.M.; FERREIRA, A.G.; MARIATH, J.E.A. e COCUCCI, A.R. 1994. Amido no megagametófito de *Araucaria angustifolia* (Bert.) O. Ktze: degradação durante a germinação e desenvolvimento do esporófito. *Acta Botânica Brasileira* 8:35-43.

ROTHBERG, J.M.; HINZ, W.; REARICK, T.M.; SCHULTZ, J.; MILESKI, W.; DAVEY, M.; LEAMON, J.H.; JOHNSON, K.; MILGREW, M.J.; EDWARDS, M.; HOON, J.; SIMONS, J.F.; MARRAN, D.; MYERS, J.W.; DAVIDSON, J.F.; BRANTING, A.; NOBILE, J.R.; PUC, B.P.; LIGHT, D.; CLARK, T.A.; HUBER, M.; BRANCIFORTE, J.T.; STONER, I.B.; CAWLEY, S.E.; LYONS, M.; FU, Y.; HOMER, N.; SEDOVA, M.; MIAO, X.; REED, B.; SABINA, J.; FEIERSTEIN, E.; SCHORN, M.; ALANJARY, M.; DIMALANTA, E.; DRESSMAN, D.; KASINSKAS, R.; SOKOLSKY, T.; FIDANZA, J.A.; NAMSARAEV, E.; McKERNAN, K.J.; WILLIAMS, A.; ROTH, G.T. e BUSTILLO, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475:348-352.

ROZEN S. e SKALETSKY H.J. 2000. Primer3 on the WWW for general users and for biologist programmers. In S. Misener and S. A. Krawetz [eds.], *Methods in molecular biology*, vol. 132: Bioinformatics methods and protocols, 365–386. Humana Press, Totowa, New Jersey, EUA.

SALGUEIRO, F.; CARON, H.; DE SOUZA, M.I.F.; KREMER, A. e MARGIS, R. 2005. Characterization of nuclear microsatellite loci in South American *Araucariaceae* species. *Molecular Ecology Notes* 5:256-258.

SANGER F.; NICKEN, S. e COULSON, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* 74:(12):5463-7.

SANKARAN, M.; RATNAM, J. e HANAN, N.P. 2004. Tree–grass coexistence in savannas revisited – insights from an examination of assumptions and mechanisms invoked in existing models. *Ecology Letters* 7:480–90.

SANT'ANNA, C.; SEBBENN, A.M.; KLABUNDE, G.H.F.;

BITTENCOURT, R.; NODARI, R.O.; MANTOVANI, A. e SEDREZ dos REIS, M. 2013. Realized pollen and seed dispersal within a continuous population of the dioecious coniferous Brazilian pine [*Araucaria angustifolia* (Bertol.) Kuntze]. *Conservation Genetics* 14:601-613.

SANTOS, K.L.; SANTOS, M.O.; LABORDA, P.R.; SOUZA, A.P.; PERONI, N. e NODARI, R.O. 2008. Isolation and characterization of microsatellite markers in *Acacia sellowiana* (Berg) Burret. *Molecular Ecology Resources* 8:998–1000.

SCHLÖGL, P.S.; SANTOS, A.L.W.; VIEIRA, L.N.; FLOH, E.L.S. e GUERRA, M.P. 2012. Gene expression during early somatic embryogenesis in Brazilian pine (*Araucaria angustifolia* (Bert) O. Ktze). *Plant Cell, Tissue and Organ Culture* 108:173-180.

SCHÜSSLER, G. 2014. Subsídios ao uso sustentável da semente (Pinhão) de *Araucaria angustifolia*. Tese (Doutorado em Ciências), UFSC, Florianópolis, SC, XXX p.

SCHMIDT, A.B.; CIAMPI, A.Y.; GUERRA, M.P. e NODARI, R.O. 2007. Isolation and characterization of microsatellite markers for *Araucaria angustifolia* (Araucariaceae). *Molecular Ecology Resources* 7(2):340-342.

SCHROEDER, A.; MUELLER, O.; STOCKER, S.; SALOWSKY, R.; LEIBER, M.; GASSMANN, M.; LIGHTFOOT, S.; MENZEL, W.; GRANZOW, M. e RAGG, T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* 7:3.

SEBASTIANINI, F.; GONZÁLEZ-MARTÍNEZ, S.C. e VENDRAMIN, G.G. 2007. Review on single nucleotide polymorphisms (SNPs) and population genetic studies in conifer species. In: KOSKELA, J.; SAMUEL, C.J.A.; MÁTYÁS, Cs. e FADY, B. (Org.). *Conifer Network*. Bioversity International. Roma, Itália.

SEDREZ dos REIS, M.; MANTOVANI, A.; da SILVA, J.Z.; MARIOT, A.; BITTENCOURT, R.; FERREIRA, D.K.; STEINER, F.; MONTAGNA, T.; da SILVA, A.L.S.; FERNDANDES, C.D.; ALTRAK, G.; FIGUEREDO, L.G..U. 2012. Distribution of genetic diversity and tree



species conservation in Santa Catarina Forests remnants. In: VIBRANS, A.C.; SEVEGNANI, L.; de GASPER, A.L.; LINGNER, D.V. (Eds.). 2012. Inventário florístico florestal de Santa Catarina. Volume I. Diversidade e conservação dos remanescentes florestais. Editora da FURB, Blumenau, Brasil, 352 pp.

SEDREZ dos REIS, M; LADIO, A. e PERONI, N. 2014. Landscapes with Araucaria in South America: evidence for a cultural dimension. *Ecology and Society* 19:2.

SETOGUCHI, H.; OSAWA, T.A.; PINTAUD, J.; JAFFRÉ, T. e VEILLON, J. 1998. Phylogenetic relationships within Araucariaceae based on rbcL gene sequences. *American Journal of Botany* 85(11):1507–1516.

SHENDURE, J. 2008. The beginning of the end for microarrays? 2008. *Nature Methods* 5:585-587.

SILBA, J. 1983. A new species of *Decussocarpus* De Laub. (Podocarpaceae) from Brazil. *Phytologia* 54:460-462.

SILVA, F.C. da e MARCONI, L.P. 1990. Fitossociologia de uma floresta com Araucária em Colombo-PR. *Boletim de Pesquisa Florestal* 20:23-28.

SOUZA, V.C. 2013. Podocarpaceae in Lista de Espécies da Flora do Brasil. Jardim Botânico do Rio de Janeiro.

STANNARD, B.L. 1995. Flora of the Pico das almas: Chapada Diamantina – Bahia, Brazil. Kew: Royal Botanical Gardens, 853 p.

STEFANOVIC, S.; JAGER, M.; DEUTSCH, J.; BROUTIN, J. e MASSELOT, M. 1998. Phylogenetic relationships of conifers inferred from partial 28SrRNA gene sequences. *American Journal of Botany* 85:688-697.

STEFENON, V. M. e CAPRESTANO, C.A. 2009. Monoicy in *A. angustifolia* (Bert.) O. Kuntze (Araucariaceae): I. Morphological aspects of the reproductive structures. *Anais da Academia Brasileira de Ciências* 81(4):701-705.

STEFENON, V. M., GAILING, O. e FINKELDEY, R. 2006. Phylogenetic relationship within genus *Araucaria* (Araucariaceae) assessed by means of AFLP fingerprints. *Silvae Genet* 55:45-52.

STEFENON, V.M., GAILING, O. e FINKELDEY, R. 2007. Genetic Structure of *Araucaria angustifolia* (Araucariaceae) Populations in Brazil: Implications for the in situ Conservation of Genetic Resources. *Plant Biology* 9:516-525.

STEFENON, V. M., GAILING, O. e FINKELDEY, R. 2008. Genetic structure of plantations and the conservation of genetic resources of Brazilian pine (*Araucaria angustifolia*). *Forest Ecology and Management* 255:2718–2725.

STEHMANN, J.R.; FORZZA, R.C.; SALINO, A; SOBRAL, M.; COSTA, D.P. e KAMINO, L.H.Y. 2009. *Plantas da Floresta Atlântica*. Rio de Janeiro: Jardim Botânico do Rio de Janeiro, 516p.

STOCKEY, R.A. 1982. The Araucariaceae: an evolutionary perspective. *Review of Palaeobotany and Palynology* 37:133–154.

SURGET-GROBA, Y. e MONTOYA-BURGOS, J. 2010. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Research* 20:1432-1440.

THOMAS, P. 2013. *Araucaria angustifolia*. The IUCN Red List of Threatened Species. Version 2014.2. <[www.iucnredlist.org](http://www.iucnredlist.org)>. Acesso em 30 julho de 2014.

TWELL, D.; OH, S.A. e HONYS, D. 2006. Pollen development, a genetic and transcriptomic review. [w] Malhó R [red.] The pollen tube. A cellular and molecular perspective. Berlin, Springer-Verlag, *Plant Cell Monographs* 3: 15-45.

UENO, S.; MORIGUCHI, Y.; UCHIYAMA, K.; UJINO-IHARA, T.; FUTAMURA, N.; SAKURAI, T.; SHINOHARA, K. e TSUMURA, Y. 2012. A second generation framework for the analysis of microsatellites in expressed sequence tags and the development of EST-SSR markers for a conifer, *Cryptomeria japonica*.

VARSHNEY, R.K.; GRANER, A. e SORRELLS, M.E. 2005. Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* 23:48–55.

VEBLEN, T.T.; ARMESTO, J.J.; BURNS, B.R.; KITZBERGER, T.; LARA, A.; LEON, B. e YOUNG, K.R. 2005. The coniferous forests of South America. In: ANDERSSON, F.A. (Ed.) *Ecosystems of the world*. New York, Elsevier. Cap.6, p.701-725.

VELOSO, H.P. e GOES-FILHO, L. 1982. *Fitogeografia Brasileira - Classificação Fisionômico-Ecológica da Vegetação Neotropical*. Projeto RADAM Brasil, Brasília, Boletim Técnico, Série Vegetação 1:3-79.

VELOSO, H.P.; RANGEL FILHO, A.L.R. e LIMA, L.C.A. 1991. *Classificação da vegetação brasileira, adaptada a um sistema universal*. Rio de Janeiro, IBGE.

VENTER, J.C. et al. 2001. The sequence of the human genome. *Science* 291:1304-1351.

VERMA, M.; GHANGAL, R.; SHARMA, R.; SINHA, A.K. e JAIN, M. 2014. Transcriptome analysis of *Catharanthus roseus* for gene Discovery and expression. Profiling. *PLoS One* 9(7).

VIEIRA, L. do N.; FAORO, H.; ROGALSKI, M.; FRAGA, H.P. de F.; CARDOSO, R.L. de SOUZA, E.M.; PEDROSA, F. de O.; NODARI, R.O. e GUERRA, M.P. 2014. The complete chloroplast genome sequence of *Podocarpus lambertii*: genome structure, evolutionary aspects, gene content and SSR detection. *PLOS ONE* 9(3).

VIEIRA, L. do N.; ROGALSKI, M.; FAORO, H.; FRAGA, H.P. de F.; dos ANJOS, K.G.; PICCHI, G.F.A.; NODARI, R.O.; PEDROSA, F. de O.; de SOUZA, E.M. e GUERRA, M.P. 2016. The plastome sequence of the endemic Amazonian conifer, *Retrophyllum piresii* (Silba) C.N. Page, reveals different recombination events and plastome isoforms. *Tree Genetics & Genomes* 12:10.

VIGNAL, A.; MILAN, D.; SAN CRISTOBAL, M. e EGGEN, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics, Selection and Evolution* 34:275-305.

VINOGRADOV, A.E. 2003. DNA helix: the importance of being GC-rich. *Nucleic Acids Research* 31:1838-1844.

WANG, M.L.; BARKLEY, N.A. e JENKINS, T.M. 2009. Microsatellite markers in plants and insects. Part I: Applications of biotechnology. *Genes, genomes and genomics* 3:54-67.

WANG, Z.; GERSTEIN, M. e SNYDER, M. 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10:57-63.

WANG, S.; WANG, X.; HE, Q.; LIU, X.; XU, W.; LI, L.; GAO, J. e WANG, F. 2012. Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of EST-SSR markers in radish. *Plant Cell Reports* 31:1437-1447.

WILLIAMS, C. 2009. *Conifer reproductive biology*. Springer. Berlin, 172 p.

XIANG, X.; ZHANG, Z.; WANG, Z.; ZHANG, X. e WU, G. 2015. Transcriptome sequencing and development of EST-SSR markers in *Pinus dabeshanensis*, an endangered conifer endemic to China. *Molecular Breeding* 35:158.

YE, J.; FANG, L.; ZHENG, H.; ZHANG, Y.; CHEN, J.; ZHANG, Z.; WANG, J.; LI, S.; LI, R. e BOLUND, L. 2006. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research* 34:293-297.

ZALAPA, J.E.; CUEVAS, H.; ZHU, H.; STEFFAN, S.; SENALIK, D.; ZELDIN, E.; McCOWN, B.; HARBUT, R. e SIMON, P. 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany* 99:193-208.

ZECHINI, A.A. 2012. Morfometria, produção, fenologia e diversidade genética: subsídios para conservação da *Araucaria angustifolia* (Bert.) O. Ktze. e coleta sustentável do pinhão no planalto catarinense. Dissertação (Mestrado em Ciências), UFSC, Florianópolis, SC, 191 p.

ZIENKIEWICZ, K.; SMOLIŃSKI, D.J. e BEDNARSKA, E. 2006.

Distribution of poly(A)RNA and splicing machinery elements in *Hyacinthus orientalis* L. pollen grains and pollen tubes growing in vitro. *Protoplasma* 227:95-103.

ZIMIN, A., STEVENS, K.A., CREPEAU, M.W., HOLTZ-MORRIS, A., KORIABINE, M., MARÇAIS, G., PUIU, D., ROBERTS, M., WEGRZYN, J.L., de JONG, P.J. NEALE, D.B., SALZBERG, S.L., YORKE, J.A. e LANGLEY, C.H. 2014. Sequencing and Assembly of the 22-Gb Loblolly Pine Genome. *Genetics* 196(3):875-890.

ZIMMER, G.O.; PAZ, C.P. e GANADE, G. 2010. Efeitos de diferentes espécies pioneiras sobre a colonização de *Podocarpus lambertii* em uma área em restauração. *Neotropical Biology and Conservation* 5(3):160-166.

ZONNEVELD, B.J.M. 2012a. Genome sizes of all 19 *Araucaria* species are correlated with their geographical distribution. *Plant Systematics and Evolution* 298:1249-1255.

ZONNEVELD, B.J.M. 2012b. Conifer genome sizes of 172 species, covering 64 to 67 genera, range from 8 to 72 picogram. *Nordic Journal of Botany* 30:001-013.

**APÊNDICE A – Demais soluções utilizadas na rotina laboratorial envolvendo análise de RNA.**

**Água DEPC 0,1% (1 L):**

Água ultrapura – 999 ml

DEPC – 1 ml

Manter sob agitação (90 RPM) durante 12 h, autoclavar

**Etanol 70% (200 ml):**

Água DEPC 0,1% - 50 ml

Etanol absoluto – 140 ml

**Etanol 75% (200 ml):**

Água DEPC 0,1% - 50 ml

Etanol absoluto – 150 ml

**Peróxido de hidrogênio 3% (1 L):**

Água DEPC 0,1% - 900 ml

Peróxido de hidrogênio 30% - 100 ml

**APÊNDICE B – Composição dos tampões e soluções para o isolamento de RNA.**

**Tampão de isolamento (400 ml): Preparar solução a 60 °C sob agitação**

CTAB – 8 g (2%)

PVP – 8 g (2%)

Tris-HCl pH 8,0 – 40 ml Tris-HCl 1M (100 mM)

EDTA – 2,922 g (25 mM)

NaCl – 46,755 g (2 M)

Espermidina – 0,2 g (0,5 g/L)

Água DEPC 0,1% q.s.p.: 400 ml

Autoclavar solução, adicionar 2% de 2-Mercaptoetanol somente antes do isolamento.

**Solução de cloreto de lítio 4 M (100 ml):**

LiCl – 16,9576 g (4 M)

Água DEPC 0,1% q.s.p. 100 ml

Autoclavar solução

**Tampão SSTE (100 ml):**

NaCl – 5,844 g (1 M)

SDS – 0,5 g (0,5%)

Tris-HCl pH 8,0 – 1 µL Tris-HCl 1M, pH 8,0 (10 µM)  
EDTA – 29,2 mg (1 mM)  
Água DEPC 0,1% q.s.p. 100 ml - Autoclavar solução

**HCl 5 N (100 ml):**

Água DEPC – 25 ml  
HCl 37% - 41,059 ml  
Água DEPC 0,1% q.s.p. 100 ml

**Tris-HCl 1 M, pH 8,0 (250 ml):**

Tris – 30,25 g  
Água DEPC 0,1% – 200 ml  
Ajustar o pH para 8,0 com HCl 5N  
Água DEPC 0,1% q.s.p. 250 ml

**Acetato de sódio 3 M, pH 5,2 (100 ml):**

Acetato de sódio – 24,6 g  
Água DEPC 0,1% – 60 ml  
Ajustar o pH para 5,2 com HCl 5N  
Água DEPC 0,1% q.s.p. 100 ml

**APÊNCIDE C – Composição dos tampões e soluções para eletroforese nativa e desnaturante de RNA.**

**Tampão MOPS 10X (1 L):**

MOPS – 83,6 g (0,4 M), pH 7,0  
Acetato de sódio – 8,2 g (0,1 M)  
EDTA – 2,92 g (0,01 M), pH 8,0  
Água DEPC 0,1% q.s.p. 1 L

**Gel de agarose 1,0% (100 ml, eletroforese desnaturante):**

Agarose – 1 g  
Água DEPC 0,1% - 72 ml  
Dissolver a agarose e resfriar a 60 °C  
Formaldeído 37% - 18 ml (manipular dentro de capela de exaustão)  
MOPS 10X – 10 ml

**Gel de agarose 1,0% (100 ml, eletroforese nativa):**

Agarose – 1 g

Tampão TBE 1X – 99 ml

Dissolver a agarose e resfriar a 60 °C

**Tampão TBE 10X (1 L) pH 8,3:**

Tris – 10,78 g (89 mM)

Ácido bórico – 5,50 g (89 mM)

EDTA – 584,48 mg (2 mM)

Água DEPC 0,1% q.s.p. 1 L

**Brometo de Etídeo (10 mg/ml):**

Brometo de etídeo – 1 g

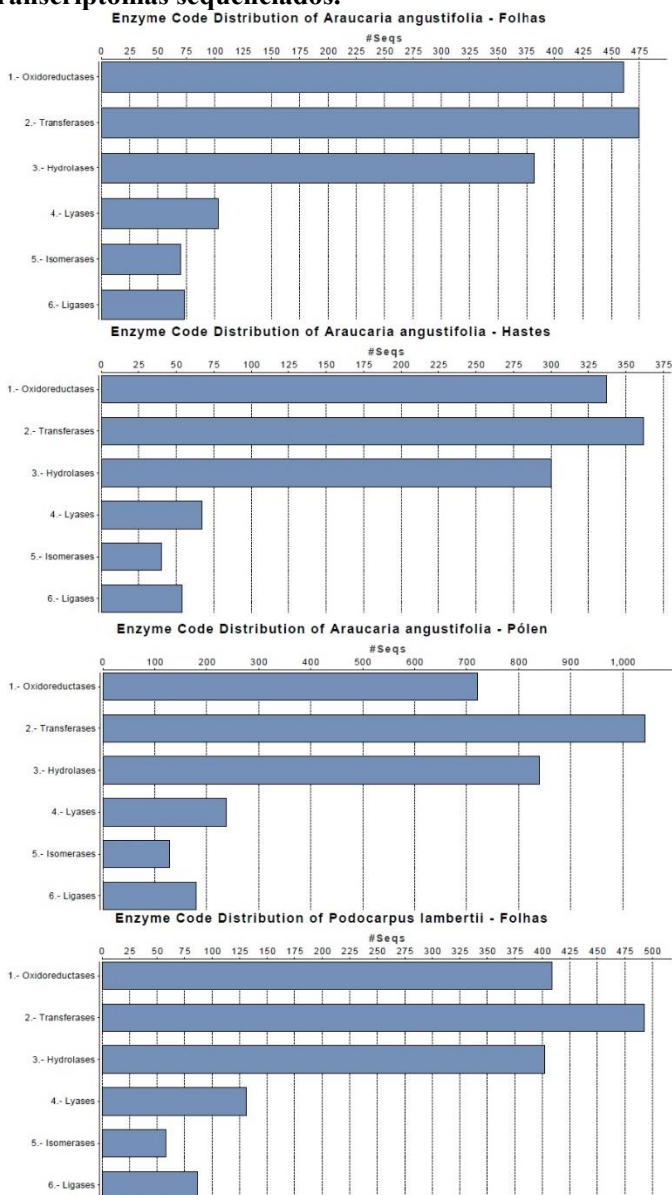
Água DEPC 0,1% – 100 ml

Agitar durante 12 h, acondicionar a 4 °C

Concentração no gel: 5 µg/ml



## APÊNDICE D – Distribuição das classes enzimáticas identificadas nos transcriptomas sequenciados.



Fonte: Autor

**APÊNDICE E – Contigs, posição, base referência e base polimórfica dos SNPs putativos identificados em *P. lambertii*.**

---

FPL_bio_rep_1_contig_71	76	A	T
FPL_bio_rep_1_contig_132	448	A	G
FPL_bio_rep_1_contig_132	459	T	A
FPL_bio_rep_1_contig_241	255	A	G
FPL_bio_rep_1_contig_2303	24	G	A
FPL_bio_rep_1_contig_3561	49	A	C
FPL_bio_rep_2_contig_9	204	T	A
FPL_bio_rep_2_contig_23	79	G	C
FPL_bio_rep_2_contig_91	36	A	G
FPL_bio_rep_2_contig_91	58	T	C
FPL_bio_rep_2_contig_190	194	T	C
FPL_bio_rep_2_contig_489	85	T	C
FPL_bio_rep_2_contig_784	57	A	C
FPL_bio_rep_2_contig_930	33	C	A
FPL_bio_rep_2_contig_1434	85	G	C
FPL_bio_rep_2_contig_2427	44	G	T
FPL_bio_rep_2_contig_3389	118	C	T
FPL_bio_rep_2_contig_5885	42	C	G
FPL_bio_rep_3_contig_19	70	C	G
FPL_bio_rep_3_contig_19	87	A	G
FPL_bio_rep_3_contig_25	188	C	T
FPL_bio_rep_3_contig_45	24	A	T
FPL_bio_rep_3_contig_72	236	T	C
FPL_bio_rep_3_contig_74	115	G	A
FPL_bio_rep_3_contig_78	119	G	C
FPL_bio_rep_3_contig_100	20	A	T
FPL_bio_rep_3_contig_100	37	C	T
FPL_bio_rep_3_contig_100	53	C	G
FPL_bio_rep_3_contig_100	78	C	T
FPL_bio_rep_3_contig_226	208	G	A
FPL_bio_rep_3_contig_456	132	C	T
FPL_bio_rep_3_contig_507	179	C	G
FPL_bio_rep_3_contig_547	360	T	G
FPL_bio_rep_3_contig_841	47	T	C
FPL_bio_rep_3_contig_841	84	C	T

---

Fonte: Autor

**APÊNDICE F - *Contigs*, posição, base referência e base polimórfica dos SNPs putativos identificados em *A. angustifolia*.**

FAA_bio_rep_1_contig_56	73	T	G	FAA_bio_rep_3_contig_64	599	A	T
FAA_bio_rep_1_contig_65	179	G	A	FAA_bio_rep_3_contig_69	99	T	C
FAA_bio_rep_1_contig_65	199	C	T	FAA_bio_rep_3_contig_72	111	A	G
FAA_bio_rep_1_contig_70	107	C	G	FAA_bio_rep_3_contig_96	79	T	C
FAA_bio_rep_1_contig_129	371	A	G	FAA_bio_rep_3_contig_103	87	T	C
FAA_bio_rep_1_contig_131	332	A	G	FAA_bio_rep_3_contig_113	80	T	C
FAA_bio_rep_1_contig_221	181	G	A	FAA_bio_rep_3_contig_145	6	A	G
FAA_bio_rep_1_contig_309	110	G	A	FAA_bio_rep_3_contig_174	169	T	C
FAA_bio_rep_1_contig_476	78	T	G	FAA_bio_rep_3_contig_174	173	T	C
FAA_bio_rep_1_contig_655	33	A	G	FAA_bio_rep_3_contig_256	102	C	T
FAA_bio_rep_1_contig_655	193	T	C	FAA_bio_rep_3_contig_272	244	C	T
FAA_bio_rep_1_contig_655	202	T	C	FAA_bio_rep_3_contig_293	166	A	C
FAA_bio_rep_1_contig_862	48	C	T	FAA_bio_rep_3_contig_336	323	T	G
FAA_bio_rep_1_contig_862	70	C	T	FAA_bio_rep_3_contig_336	327	C	T
FAA_bio_rep_1_contig_862	72	C	A	FAA_bio_rep_3_contig_401	111	C	A
FAA_bio_rep_1_contig_862	85	G	A	FAA_bio_rep_3_contig_429	85	C	A
FAA_bio_rep_1_contig_4398	160	G	A	FAA_bio_rep_3_contig_553	157	G	A
FAA_bio_rep_1_contig_6890	31	C	T	FAA_bio_rep_3_contig_563	201	A	G
FAA_bio_rep_2_contig_23	1511	A	G	FAA_bio_rep_3_contig_579	28	C	T
FAA_bio_rep_2_contig_29	49	C	T	FAA_bio_rep_3_contig_699	99	C	T
FAA_bio_rep_2_contig_29	52	T	A	FAA_bio_rep_3_contig_738	17	G	A
FAA_bio_rep_2_contig_29	57	T	G	FAA_bio_rep_3_contig_1013	18	C	T
FAA_bio_rep_2_contig_29	64	G	C	FAA_bio_rep_3_contig_1141	97	T	C
FAA_bio_rep_2_contig_29	66	T	G	FAA_bio_rep_3_contig_1141	101	C	T
FAA_bio_rep_2_contig_29	68	C	T	FAA_bio_rep_3_contig_1141	106	C	T
FAA_bio_rep_2_contig_32	213	C	T	FAA_bio_rep_3_contig_2348	41	G	T
FAA_bio_rep_2_contig_32	218	A	G	FAA_bio_rep_3_contig_3167	146	G	C
FAA_bio_rep_2_contig_32	220	A	C	FAA_bio_rep_3_contig_4228	64	A	T
FAA_bio_rep_2_contig_32	236	T	G	FAA_bio_rep_3_contig_4228	67	A	T
FAA_bio_rep_2_contig_32	242	T	C	FAA_bio_rep_3_contig_4728	40	A	G
FAA_bio_rep_2_contig_32	246	T	C	FAA_bio_rep_3_contig_7973	145	T	G
FAA_bio_rep_2_contig_46	170	G	A	HAA_bio_rep_1_contig_6	3074	C	T
FAA_bio_rep_2_contig_77	583	G	T	HAA_bio_rep_1_contig_6	3081	C	T
FAA_bio_rep_2_contig_77	586	T	C	HAA_bio_rep_1_contig_6	3105	C	T
FAA_bio_rep_2_contig_80	440	A	G	HAA_bio_rep_1_contig_6	3107	A	C
FAA_bio_rep_2_contig_130	26	A	G	HAA_bio_rep_1_contig_40	55	G	A
FAA_bio_rep_2_contig_168	76	A	T	HAA_bio_rep_1_contig_48	363	G	A
FAA_bio_rep_2_contig_168	98	T	G	HAA_bio_rep_1_contig_48	371	A	C
FAA_bio_rep_2_contig_168	107	A	C	HAA_bio_rep_1_contig_48	386	C	T
FAA_bio_rep_2_contig_168	111	A	G	HAA_bio_rep_1_contig_78	148	T	C
FAA_bio_rep_2_contig_660	114	A	G	HAA_bio_rep_1_contig_78	150	T	G
FAA_bio_rep_2_contig_1322	112	C	T	HAA_bio_rep_1_contig_78	155	C	T
FAA_bio_rep_2_contig_1545	143	A	G	HAA_bio_rep_1_contig_78	166	T	C
FAA_bio_rep_2_contig_1545	151	A	T	HAA_bio_rep_1_contig_300	224	T	C
FAA_bio_rep_3_contig_55	88	G	C	HAA_bio_rep_1_contig_300	243	C	T

Fonte: Autor

## APÊNDICE F – CONTINUAÇÃO

HAA_bio_rep_1_contig_529	38	C	T	HAA_bio_rep_3_contig_1203	107	A	G
HAA_bio_rep_1_contig_760	49	T	G	HAA_bio_rep_3_contig_1203	110	A	G
HAA_bio_rep_1_contig_760	119	A	G	HAA_bio_rep_3_contig_1396	108	A	G
HAA_bio_rep_1_contig_760	122	G	A	HAA_bio_rep_3_contig_1396	114	C	G
HAA_bio_rep_1_contig_786	46	A	G	HAA_bio_rep_3_contig_1750	437	A	G
HAA_bio_rep_1_contig_789	20	G	A	HAA_bio_rep_3_contig_2119	294	T	C
HAA_bio_rep_1_contig_806	157	C	T	HAA_bio_rep_3_contig_2381	233	A	C
HAA_bio_rep_1_contig_1234	156	A	G	HAA_bio_rep_3_contig_2519	66	C	T
HAA_bio_rep_1_contig_1234	158	A	G	HAA_bio_rep_3_contig_2849	18	A	G
HAA_bio_rep_1_contig_1234	160	A	C	HAA_bio_rep_3_contig_3646	36	G	C
HAA_bio_rep_1_contig_1337	20	A	G	HAA_bio_rep_3_contig_3977	113	T	G
HAA_bio_rep_1_contig_1579	80	C	T	HAA_bio_rep_3_contig_3977	115	T	A
HAA_bio_rep_1_contig_1579	87	G	A	HAA_bio_rep_3_contig_3996	107	G	T
HAA_bio_rep_1_contig_2323	50	A	C	HAA_bio_rep_3_contig_4215	184	T	G
HAA_bio_rep_2_contig_3	232	C	T	HAA_bio_rep_3_contig_7317	121	C	A
HAA_bio_rep_2_contig_3	265	C	T	PAA_bio_rep_1_contig_75	232	A	T
HAA_bio_rep_2_contig_112	156	C	T	PAA_bio_rep_1_contig_75	249	C	T
HAA_bio_rep_2_contig_192	148	G	A	PAA_bio_rep_1_contig_174	90	T	G
HAA_bio_rep_2_contig_198	146	A	C	PAA_bio_rep_1_contig_288	69	T	A
HAA_bio_rep_2_contig_343	39	G	T	PAA_bio_rep_1_contig_289	147	T	A
HAA_bio_rep_2_contig_859	86	G	A	PAA_bio_rep_1_contig_291	154	G	A
HAA_bio_rep_2_contig_1049	61	G	A	PAA_bio_rep_1_contig_750	131	T	G
HAA_bio_rep_2_contig_1296	67	T	A	PAA_bio_rep_1_contig_3110	63	G	A
HAA_bio_rep_2_contig_2558	34	G	A	PAA_bio_rep_1_contig_3110	76	T	C
HAA_bio_rep_3_contig_18	68	A	G	PAA_bio_rep_1_contig_4865	161	T	C
HAA_bio_rep_3_contig_19	237	G	A	PAA_bio_rep_1_contig_13473	89	C	G
HAA_bio_rep_3_contig_31	1467	A	G	PAA_bio_rep_1_contig_13473	96	T	C
HAA_bio_rep_3_contig_55	155	T	C	PAA_bio_rep_1_contig_13473	105	A	G
HAA_bio_rep_3_contig_55	174	C	T	PAA_bio_rep_1_contig_13473	109	T	G
HAA_bio_rep_3_contig_102	580	C	A	PAA_bio_rep_2_contig_6	745	T	A
HAA_bio_rep_3_contig_115	494	C	T	PAA_bio_rep_2_contig_27	155	G	A
HAA_bio_rep_3_contig_132	306	G	A	PAA_bio_rep_2_contig_89	57	T	C
HAA_bio_rep_3_contig_132	308	A	T	PAA_bio_rep_2_contig_100	136	A	G
HAA_bio_rep_3_contig_135	256	C	T	PAA_bio_rep_2_contig_220	277	T	C
HAA_bio_rep_3_contig_160	122	C	T	PAA_bio_rep_2_contig_327	196	A	G
HAA_bio_rep_3_contig_187	126	A	C	PAA_bio_rep_2_contig_419	202	T	G
HAA_bio_rep_3_contig_301	16	G	A	PAA_bio_rep_2_contig_419	215	A	C
HAA_bio_rep_3_contig_301	21	A	G	PAA_bio_rep_2_contig_1679	113	C	T
HAA_bio_rep_3_contig_306	155	A	G	PAA_bio_rep_2_contig_1679	120	T	C
HAA_bio_rep_3_contig_473	136	C	G	PAA_bio_rep_2_contig_1679	132	A	T
HAA_bio_rep_3_contig_481	74	C	T	PAA_bio_rep_2_contig_4455	46	C	T
HAA_bio_rep_3_contig_505	88	G	T	PAA_bio_rep_2_contig_4455	83	G	A
HAA_bio_rep_3_contig_505	105	C	T	PAA_bio_rep_2_contig_4455	99	T	G
HAA_bio_rep_3_contig_812	22	C	G	PAA_bio_rep_2_contig_4455	109	G	A
HAA_bio_rep_3_contig_1147	106	C	G	PAA_bio_rep_2_contig_4455	111	T	C

Fonte: Autor

## APÊNDICE F – CONTINUAÇÃO

---

PAA_bio_rep_2_contig_4455	116	C	T
PAA_bio_rep_2_contig_4455	125	C	T
PAA_bio_rep_2_contig_4456	32	A	T
PAA_bio_rep_3_contig_12	49	C	T
PAA_bio_rep_3_contig_12	51	A	G
PAA_bio_rep_3_contig_21	174	T	C
PAA_bio_rep_3_contig_26	1052	G	A
PAA_bio_rep_3_contig_47	18	G	A
PAA_bio_rep_3_contig_52	250	C	A
PAA_bio_rep_3_contig_57	77	C	T
PAA_bio_rep_3_contig_57	79	A	G
PAA_bio_rep_3_contig_104	290	T	C
PAA_bio_rep_3_contig_112	769	T	C
PAA_bio_rep_3_contig_112	771	A	G
PAA_bio_rep_3_contig_112	794	A	G
PAA_bio_rep_3_contig_112	797	C	T
PAA_bio_rep_3_contig_123	27	G	A
PAA_bio_rep_3_contig_123	53	G	A
PAA_bio_rep_3_contig_136	51	T	G
PAA_bio_rep_3_contig_503	53	G	A
PAA_bio_rep_3_contig_537	126	A	T
PAA_bio_rep_3_contig_636	155	C	T
PAA_bio_rep_3_contig_1032	78	C	T
PAA_bio_rep_3_contig_1297	104	A	C
PAA_bio_rep_3_contig_1668	77	C	T
PAA_bio_rep_3_contig_2545	56	G	A
PAA_bio_rep_3_contig_3017	351	G	A
PAA_bio_rep_3_contig_3017	356	C	A
PAA_bio_rep_3_contig_3523	279	G	A
PAA_bio_rep_3_contig_3523	281	G	A
PAA_bio_rep_3_contig_3523	286	C	A
PAA_bio_rep_3_contig_3523	290	C	T
PAA_bio_rep_3_contig_3523	300	T	C
PAA_bio_rep_3_contig_3523	310	C	T
PAA_bio_rep_3_contig_4205	162	T	C
PAA_bio_rep_3_contig_4355	185	G	A
PAA_bio_rep_3_contig_4355	187	G	A
PAA_bio_rep_3_contig_4355	192	C	A
PAA_bio_rep_3_contig_4355	196	C	T
PAA_bio_rep_3_contig_4355	205	T	C
PAA_bio_rep_3_contig_5654	61	T	G
PAA_bio_rep_3_contig_6008	110	T	C
PAA_bio_rep_3_contig_7114	90	A	G
PAA_bio_rep_3_contig_7787	109	T	C

---

Fonte: Autor

**APÊNDICE G – Listagem de primers EST-SSR obtidos para *A. angustifolia*, seqüências, temperaturas de hibridização, tamanho de amplicon, motivo SSR e início e fim dos motivos nas seqüências dos contigs.**

Contig	Seqüência iniciador direto (F)	TM °C (F)	Seqüência iniciador reverso (R)	TM °C (R)	Tamanho	Motivo SSR	Início	Fim
FAA_bio_rep_1_contig_3	AGAAGAGGTTGGAGAAAATAC	50,082	AGTAAAGCTTCATAGGGTCT	50,03	563	(AGGT)3	1700	1711
FAA_bio_rep_1_contig_6	CTATCTTAAGAGTGAGAGCG	49,593	AAAGAAATAAGGTTCCGTCTT	50,16	489	(GAGAG)3	643	657
FAA_bio_rep_1_contig_26	TTGTGTCTATTGGCTATTCT	49,982	CAAACAAAAGTACTAGGGAA	49,82	267	(T)13	951	963
FAA_bio_rep_1_contig_76	CTCTATTTCAAGCTCACTG	50,3	CCAACTACCTAATACTACGC	49,48	243	(GCTTTC)4-(GAG)4	1060	1113
FAA_bio_rep_1_contig_141	GTTCAAAGACTCGATGATT	49,789	GCAGAAAAAGTTTGAGAATA	49,98	144	(G)17	152	168
FAA_bio_rep_1_contig_168	TTACATTGGCTCTCTTTTC	49,977	AGATTTCGATCGTAGTTTACA	49,98	327	(TCTT)3	1439	1450
FAA_bio_rep_1_contig_294	AAGAACGAAAGTTAGAGGAT	49,986	CAGTCTCAGCCTCTCTATG	50,6	162	(T)12	582	593
FAA_bio_rep_1_contig_382	GAATGTTTTAGAGTTAGCCA	49,84	AAAGAAAGGAAGCTCTAATC	49,9	260	(GTT)4	361	372
FAA_bio_rep_1_contig_616	CATATCGTATCGACTTTTTTC	50,033	GCAGTCTCGCCTCTCTA	52,06	341	(T)14	571	584
FAA_bio_rep_2_contig_83	CGTGAGTTCTGACTTAGAG	50,571	TACTTTACCCATATTTCCAAA	50,06	298	(T)22	125	146
FAA_bio_rep_2_contig_102	GAATAATCAACCCTGTATCA	49,94	ATGTTGCAGTACCAAGTAAC	50,17	202	(ACAA)3	531	542
FAA_bio_rep_2_contig_554	AATCTGATAATGGTGACAAC	49,65	AACAATCTAAAATTCGCTG	49,84	507	(TCT)5	517	531
FAA_bio_rep_2_contig_738	GAAAGACTAATCTGGTCTCC	50,375	CATTTCTCATGGATTCTTC	50,19	361	(GTT)3	56	67
HAA_bio_rep_1_contig_342	GTTGGTTCGTTGTAGTAGT	49,36	GTGCCCTGGCAGTCT	50,16	428	(G)17	576	592
HAA_bio_rep_1_contig_449	AATATAACATGATCCATCG	50,041	GTTTCATATTCACCATCTCAG	50,19	267	(GATA)3	358	369
HAA_bio_rep_2_contig_255	TTAGACTAGAGGTGCTCTGA	50,117	GTGCCCTGGCAGTCTA	50,65	514	(G)14	528	541
HAA_bio_rep_3_contig_31	TCTCTTTTAAAACAAAGGG	49,816	AAGTCTGTCTGCTTACTGAA	50,02	470	(GAGAG)3	418	432
HAA_bio_rep_3_contig_62	CACGTAGGAATGTTTTAGAG	50,147	TTTTAAGTCGTCTCAAAAGG	50	281	(GTT)4	472	483
HAA_bio_rep_3_contig_288	AAGGACTGTGAGGTAGG	47,25	GATATATGGTCGCTGTGAA	49,99	332	(TC)6	23	34
HAA_bio_rep_3_contig_432	GTTTCCATTGTTAGGTAATC	48,839	TACTTTACCCATATTTCCAAA	50,06	486	(TA)8	43	58
PAA_bio_rep_1_contig_25	TGCGTTATACATACAGAACA	50,116	CTGTAATCTTTAACACCAGC	50,12	250	(ATCA)3	335	346
PAA_bio_rep_1_contig_84	TAACAGAACTAGCAGGTGTT	50,065	CAGTCTCAGCCTCTCTATG	50,6	470	(A)13	989	1001
PAA_bio_rep_1_contig_139	CTTTAATGACTTGAGAGGTG	49,912	TTAGATCTCACCTGAATCAC	49,83	565	(T)12-(AAT)4	347	443
PAA_bio_rep_1_contig_363	CCAACGAGAATACTTTTATG	50,193	TATCCCTGTGTGCC	50,81	326	(C)16	650	665
PAA_bio_rep_1_contig_1809	GTCTGAGTTTCTGTAATGGA	49,977	CAGTCTCAGCCTCTCTATG	50,6	453	(C)13	448	460
PAA_bio_rep_2_contig_149	CTACTCTTCATATTCACCG	50,016	GAAGCTACAGTACGAGAC	48,69	282	(G)12	1050	1061
PAA_bio_rep_3_contig_148	ATGTTTGACTCTCAAGTG	49,92	TTGGATCAGTGATATAGTT	49,98	432	(GCTT)3	605	616

## APÊNDICE G – CONTINUAÇÃO

PAA_bio_rep_3_contig_364	GAGTCAATAGGAGGGTACT	49,942	TGATCTATTCCTTATGGTCG	50,35	256	(GGGAGAG)3	1024	1044
PAA_bio_rep_3_contig_366	GACTATACTATTGTGGAGCG	50,032	TACAGAGAGGGTTTGATTTA	49,97	136	(AATG)3	192	203
PAA_bio_rep_3_contig_392	GAAAACGAATTTCTAGCC	49,94	TATTTTGCACCTTTTAATC	50,07	443	(CGCCTC)3	77	94
PAA_bio_rep_3_contig_524	GGAAGTGGCGTAAGAG	49,192	CTTGTAAGGAAATGAATCAG	49,97	172	(AGC)5	26	40
PAA_bio_rep_3_contig_968	AATACATTGGAATTGAGTTG	50,073	TATCCCCTGTGTGCC	50,81	458	(A)14	1091	1104
PAA_bio_rep_3_contig_1174	AACGTTTTGATTTCTAGTGA	50,065	ATCCAGACTTTTCITTTTCT	49,94	436	(GAAAAG)3	275	292
PAA_bio_rep_3_contig_1211	TGCAAGTTAGAATAGAGAGC	50,032	GTGCCTTGGCAGTCT	50,16	113	(G)20	686	705
PAA_bio_rep_3_contig_1785	AACAGCAGTATGAAAACAGT	49,836	TGTTTCITTCAGTGAAGACT	49,77	286	(CAA)5-(GAG)4-CAG)4	292	434

Fonte: Autor

**APÊNDICE H - Listagem de *primers* EST-SSR obtidos para *P. lambertii*, sequências, temperaturas de hibridização, tamanho de *amplicon*, motivo SSR e início e fim dos motivos nas sequências dos *contigs*.**

<i>Contig</i>	Sequência iniciador direto (F)	TM °C (F)	Sequência iniciador reverso (R)	TM °C (R)	Tamanho	Motivo SSR	Início	Fim
FPL_bio_rep_1_contig_8	CGAGTACTACGGGAATAATA	49,8	CTGTGTTTTGGTAAACAGT	50,3	251	(AGGT)3	1577	1588
FPL_bio_rep_1_contig_14	TAATAGCTCACTGGTCTAGC	49,9	ATTCTCGTTACTCATGTCAG	50	240	(GCTT)5-(GTGA)4	1677	1721
FPL_bio_rep_1_contig_28	TGGAACAAGTTGATAGAATC	50	ATGAGCACATTGAACATCT	49,6	544	(T)13-(AAT)4	1739	1836
FPL_bio_rep_1_contig_165	CTATGTCTTTGACACTAGGC	49,9	GCCCCTAGGTATTCTCT	48,7	167	(AGGT)3	501	512
FPL_bio_rep_1_contig_166	ACGTTATCCGTAAGATACAA	50	TCTTACTTCAGAACGCTTAC	50	558	(AT)6	897	908
FPL_bio_rep_1_contig_183	GTAAACTTTCTTCATGCAC	50	AGTTACTTCTGTTCCCTATTT	50	429	(T)12	269	280
FPL_bio_rep_1_contig_216	TCATTTTTATTGATCTCGT	50	AGTTCTGCTACCAAATG	50,4	475	(A)13	1079	1091
FPL_bio_rep_1_contig_296	GTCTAGTAGCTGGTTTCCTC	50,7	CGATGACCAATATATGAACT	50	224	(AATG)3	57	68
FPL_bio_rep_1_contig_360	ATTAGAACCGGATTATCA	49,7	CGTCTCTGGTTTTACTTA	50	173	(AT)7	59	72
FPL_bio_rep_1_contig_388	AAATCATGCTCATTCTCAT	50,2	AAATTAAGTACGATGACAC	50	453	(GGC)4	109	120
FPL_bio_rep_1_contig_954	CTGAAAAGGGTAGTGTAAAA	49,8	AGATAGTGCTTCGATTATCA	50	371	(GAA)4	247	258
FPL_bio_rep_2_contig_81	AAAGTATAGCGTTGAACAAG	50	GATTC AATCATAGAAAAACG	50	539	(A)12	2270	2281
FPL_bio_rep_2_contig_95	TGGAACAAGTTGATAGAATC	50	TCAATTCCTTTAAGTTTCAG	49,8	444	(T)12	815	826
FPL_bio_rep_2_contig_152	CATTTCTCAGAAGCTATGAC	49,9	CTGCCCATAGAGACAAA	49,8	258	(G)14	1732	1745
FPL_bio_rep_2_contig_152	GTTATTTTGACTGATCTGGA	50	GACCTGCCCATAGAGAC	50,7	368	(T)18	1746	1763
FPL_bio_rep_2_contig_162	AATGGTGGTCAAATAAAAA	50,5	GCAGTCTCAGCCTCTCTA	51,1	142	(G)22	506	527
FPL_bio_rep_2_contig_177	TATTCTGGATGTTGAATACC	49,9	GCAGTCTCAGCCTCTCTAT	52	430	(A)12	2364	2375
FPL_bio_rep_2_contig_243	ATATCCTTATCTATGAGGGG	50	ATTAGAACCGGATTATCA	50,7	211	(A)12	845	856
FPL_bio_rep_3_contig_47	ATCATTACTCGACAGTGAAG	50	CCGGTCAATAGATAGAGA	50,6	116	(C)13	1463	1475
FPL_bio_rep_3_contig_48	AGAAGAGGTGGAGAAAAATAC	50,1	AGTAAAGCTTCATAGGGTCT	50	564	(AGGT)3	436	447
FPL_bio_rep_3_contig_56	CTCTATTTCAAGCTCACTG	50,3	TTGTTCCTATCTTCCATA	50	267	(GCTTTC)4(GAG)4	276	329

Fonte: Autor



**APÊNDICE I – Resultados do BLASTX dos contigs de *P. lambertii* contendo motivos SSRs, *E-value*, n° de acesso, espécie referente, proteína e origem dos transcritos.**

SSR	<i>E-value</i>	Acesso n°	Espécie	Proteína	Origem
FPL_bio_rep_1_contig_8	3.23E-36	ABR26094.1	<i>Oryza sativa</i>	Retrotransposon protein	Nuclear
FPL_bio_rep_1_contig_14	2.65E-29	YP_173374	<i>Nicotiniana tabacum</i>	NitaMp027	Mt
FPL_bio_rep_1_contig_28	6.36E-66	CDY19671	<i>Brassica napus</i>	BnaC09g29270D	Nuclear
FPL_bio_rep_1_contig_165	2.7E-8	KJB11774	<i>Gossypium raimondii</i>	B456_001G276600	Nuclear
FPL_bio_rep_1_contig_166	0	YP_009020363	<i>Podocarpus lambertii</i> *	Ribosomal protein L2	Cp
FPL_bio_rep_1_contig_183	2.10E-7	ACG27180	<i>Zea mays</i>	Hypotetical protein	Nuclear
FPL_bio_rep_1_contig_216	0	YP_009020357	<i>Podocarpus lambertii</i> *	Photosystem I subunit IX	Cp
FPL_bio_rep_1_contig_296	0	YP_009020385	<i>Podocarpus lambertii</i> *	NADH-plastoquinone Oxi.S.I	Cp
FPL_bio_rep_1_contig_360	0	YP_009020399	<i>Podocarpus lambertii</i> *	Maturase K	Cp
FPL_bio_rep_1_contig_388	1.87E-49	P84338	<i>Picea sitchensis</i>	Plastocyanin	Cp
FPL_bio_rep_1_contig_954	9.16E-8	ADQ42102	<i>Elymus virginicus</i>	Photosystem II protein D1	Cp
FPL_bio_rep_2_contig_81	0	YP_009020379	<i>Podocarpus lambertii</i> *	Photosystem II CP47	Cp
FPL_bio_rep_2_contig_95	1.8E-58	YP_001152206	<i>Pinus koraiensis</i>	ORF137	Nuclear
FPL_bio_rep_2_contig_152	0	AJP33420	<i>Araucaria heterophylla</i>	ATPase subunit 1	Mt
FPL_bio_rep_2_contig_162	1.21E-7	XP_003599574	<i>Medicago trunculata</i>	Hypotetical protein MTR	Nuclear
FPL_bio_rep_2_contig_177	0	YP_009020336	<i>Podocarpus lambertii</i> *	P700 apoprotein A2	Cp
FPL_bio_rep_2_contig_243	4.99E-45	BAH11228	<i>Welwitschia mirabilis</i>	ATP syntase CFO C chain	Cp
FPL_bio_rep_3_contig_47	2.22E-40	EPS74531	<i>Genlisea aurea</i>	Hypotetical protein M569	Nuclear
FPL_bio_rep_3_contig_48	3.50E-60	KMS65245	<i>Beta vulgaris</i>	Hypotetical protein BVRB	Nuclear
FPL_bio_rep_3_contig_56	4.08E-31	AJP33420	<i>Araucaria heterophylla</i>	ATPase subunit 1	Mt

\*Top BLASTX hits no plastoma da própria espécie (VIEIRA et al. 2014).

Fonte: Autor.

**APÊNDICE J – Resultados do BLASTX dos contigs de *A. angustifolia* contendo motivos SSRs, *E-value*, n° de acesso, espécie referente, proteína e origem dos transcritos.**

SSR	<i>E-value</i>	Acesso n°	Espécie	Proteína	Origem
FAA_bio_rep_1_contig_3	2.57E-61	KRH38400	<i>Glycine max</i>	Hypotetical protein	Nuclear
FAA_bio_rep_1_contig_6	1.57E-44	CDY45505	<i>Brassica napus</i>	BnaCang12642640D	Nuclear
FAA_bio_rep_1_contig_26	2.37E-41	XP_003604154	<i>Medicago trunculata</i>	Ycf68 protein	Nuclear
FAA_bio_rep_1_contig_76	7.83E-76	EPS73652	<i>Genlisea aurea</i>	Hypotetical protein M569	Nuclear
FAA_bio_rep_1_contig_141	3.59E-16	XP_006423776	<i>Citrus clementina</i>	Hypotetical protein CICLE	Nuclear
FAA_bio_rep_1_contig_168	0	AAV84713	<i>Araucaria bidwillii</i>	Photosystem Cp 47 protein	Cp
FAA_bio_rep_1_contig_294	1.08E-18	KGN58106.1	<i>Cucumis sativus</i>	Hypotetical protein Csa	Nuclear
FAA_bio_rep_1_contig_382	2.36E-45	ACN39871	<i>Picea sitchensis</i>	rbcS	Cp
FAA_bio_rep_1_contig_616	1.30E-108	ABY83920	<i>Chamaedaphne calyculata</i>	COI I	Mt
FAA_bio_rep_2_contig_83	1.53E-153	YP_009121330	<i>Araucaria heterophylla</i>	Cytochrome b6	Cp
FAA_bio_rep_2_contig_102	0.02	XP_002986693	<i>Salagnela moellendorffii</i>	Hypotetical protein	Nuclear
FAA_bio_rep_2_contig_554	4.29E-61	AF59223	<i>Picea wilsonii</i>	Chloroplast photosystem I	Cp
FAA_bio_rep_2_contig_738	3.44E-46	BAE92287	<i>Croptomeria japonica</i>	Glycine rich RNA binding protein	Nuclear
HAA_bio_rep_1_contig_449	1.17E-5	XP_003616777	<i>Medicago trunculata</i>	Hypotetical protein	Nuclear
HAA_bio_rep_3_contig_31	3.81E-37	ACG78890	<i>Vicia faba</i>	Hypotetical protein	Mt
HAA_bio_rep_3_contig_62	8.30E-60	CAA38346	<i>Larix laricina</i>	rbc	Cp
HAA_bio_rep_3_contig_388	5.80E-28	KDP26411	<i>Jatropha curcas</i>	Hypotetical protein	Nuclear
HAA_bio_rep_3_contig_432	3.13E-35	YP_008082439	<i>Taiwania fluosiana</i>	Photosystem II protein H	Cp
PAA_bio_rep_1_contig_25	0	YP_007475535	<i>Podocarpus totara</i>	rbcL	Cp

Fonte: Autor.

SSR	<i>E-value</i>	Acesso n°	Espécie	Proteína	Origem
PAA_bio_rep_1_contig_84	0	ABK21449	<i>Picea sitchensis</i>	Glutamine synthetase	Nuclear
PAA_bio_rep_1_contig_139	1.02E-21	YP_009159802	<i>Carnegia gigantea</i>	RF68 protein	Cp
PAA_bio_rep_2_contig_363	1.77E-49	YP_009020363	<i>Podocarpus lambertii</i>	Ribosomal protein L2	Cp
PAA_bio_rep_2_contig_1809	2.90E-57	ABK22253	<i>Picea sitchensis</i>	Thioredoxin CXXS1 protein	Nuclear
PAA_bio_rep_3_contig_149	0	AAC83427	<i>Araucaria angustifolia</i> *	rbcL	Cp
PAA_bio_rep_3_contig_148	1.85E-38	ADQ37299	<i>Pinus sylvestris</i>	Arginine decarboxilase	Nuclear
PAA_bio_rep_3_contig_364	1.23E-104	AJP33504	<i>Gnatum gnemon</i>	COI II	Cp
PAA_bio_rep_3_contig_366	8.03E-26	AJP33445	<i>Araucaria heterophylla</i>	Ribosomal protein L5	Mt
PAA_bio_rep_3_contig_392	9.30E-92	XP_002535312	<i>Ricinus communis</i>	60S Ribosomal protein L16	Nuclear
PAA_bio_rep_3_contig_524	1.99E-66	ABK22509	<i>Picea sitchensis</i>	Hypotetical protein	Nuclear
PAA_bio_rep_3_contig_968	1.47E-144	XP_010268013	<i>Nelumbo nucifera</i>	Calreticulum protein	Nuclear
PAA_bio_rep_3_contig_1174	5.12E-77	ACN41018	<i>Picea sitchensis</i>	Calnexin homolog	Nuclear
PAA_bio_rep_3_contig_1211	2.70E-17	CAN71464	<i>Vitis vinifera</i>	Hypotetical protein	Nuclear
PAA_bio_rep_3_contig_1785	1.20E-20	XP_001771148	<i>Physcomitrella patens</i>	Predicted protein	Nuclear

\*Top BLASTX hit no gene rbcL previamente sequenciado em *A. angustifolia* por Setoguchi et al. 1998.

Fonte: Autor.

## **ANEXO A – Protocolo de isolamento de RNA total de pólen utilizando o kit comercial Nucleo Spin Plant®, com modificações.**

### **1. Homogeneização das amostras**

Cerca de 100 mg de pólen foram rapidamente submersas em nitrogênio líquido visando a formação de cristais de gelo e ruptura das estruturas celulares dos grãos de pólen.

### **2. Lise celular**

Após a total evaporação do nitrogênio líquido o material foi transferido para um microtubo (2,0 ml) contendo 3,5 µl de β-mercaptoetanol e 350 µl de tampão de lise celular (RAP, cloridato de guanidina). A solução foi brevemente agitada em vórtex.

### **3. Filtragem do lisado**

A amostra lisada foi transferida para uma coluna contendo membrana de sílica (anel violeta) e centrifugada a 11.000 xg por 1 min. A solução filtrada foi transferida para um novo microtubo de 1,5 ml.

### **4. Ajuste das condições de ligação do RNA**

Foram adicionados 350 µl de etanol 70%, em seguida a solução foi homogeneizada por meio de 5 pipetagens (up and down).

### **5. Ligação do RNA**

A solução foi transferida para outra coluna contendo membrana de sílica (anel azul) e centrifugada a 11.000 xg por 1 min.

### **6. Dessalinização da membrana de sílica**

Foram adicionados 350 µl do tampão de dessalinização de membrana (MDB) visando maior eficiência na digestão do DNA contaminante. Em seguida as colunas foram centrifugadas a 11.000 xg por 1 min.

### **7. Digestão do DNA**

Na membrana de sílica, foram pipetados 10 µl de rDNase reconstituída e 90 µl de tampão de digestão. O tempo de digestão do DNA foi de 15 min, em temperatura ambiente.

### **8. Lavagem e secagem da membrana de sílica**

Foram realizadas 3 lavagens nas membranas de sílica. A primeira (200 µl RAW2, 11.000 xg por 30 s) visou a inativação da rDNase. As duas lavagens posteriores, com utilização do tampão RA3, tiveram como objetivo a remoção principalmente de proteínas e compostos fenólicos. Lavagem n°2 (600 µl RAW2, 11.000 xg por 30 s) e n°3 (250 µl RAW2, 11.000 xg por 2 min).

### **9. Eluição do RNA total**

O RNA total foi eluído em 30 µl de água DEPC (fornecida pelo kit). Os tubos foram centrifugados a 11.000 xg por 2 min e a solução filtrada foi transferida para microtubo (0,5 ml) e congelada a -80 °C.

## **ANEXO B – Protocolo de depleção de rRNA em amostras de plantas. Ribominus Plant Kit for RNA-Seq (Invitrogen, Carlsbad, EUA), com modificações.**

### **1. Hibridização do rRNA**

O RNA total de cada replicata (concentrado a 10 µl) foi transferido para um microtubo contendo 100 µl de tampão e 10 µl de sondas de hibridização. As amostras foram brevemente agitadas e incubadas em banho seco a 75 °C durante 5 min para a desnaturação do RNA total. Após esse período, a temperatura foi gradativamente diminuída, durante 30 min, até 37 °C.

### **2. Preparo das beads**

As beads magnéticas foram ressuspensas no próprio frasco, por agitação em vórtex. Para cada replicata, foram pipetados 750 µl de beads em microtubo de 1,5 ml. Os tubos foram colocados em rack magnético durante 1 min. Após esse período, o sobrenadante foi aspirado e descartado. Para cada tubo, foram adicionados 750 µl de água DEPC. Os tubos foram agitados em vórtex e novamente colocados em rack magnético. Após 1 min, o sobrenadante foi aspirado e descartado. Esta etapa de lavagem com água DEPC foi novamente repetida. As beads foram ressuspensas em 750 µl de tampão de hibridização. Deste volume, foram transferidos 250 µl para um novo tubo (round de depleção #1) que foi mantido a 37 °C para uso posterior. Os tubos com os 500 µl restantes foram colocados no rack magnético, aspirados, ressuspensos em 200 µl de tampão de hibridização e mantidos a 37 °C para uso posterior (round de depleção #2).

### **3. Remoção do rRNA**

Após a etapa de hibridização, os tubos foram brevemente centrifugados (spin) e o volume (120 µl) foi transferido para o tubo contendo 250 µl de beads magnéticas (round de depleção #1). Os volumes foram gentilmente agitados e os tubos foram incubados a 37 °C durante 15 min (com breve agitação a cada 2 min, finalizado por spin rápido). Os tubos foram colocados no rack magnético por 2 min. O sobrenadante contendo o mRNA foi transferido para um novo microtubo (1,5 ml). A este microtubo foram adicionados 200 µl de beads magnéticas (round de depleção #2). Os tubos foram brevemente agitados e incubados a 37 °C durante 15 min (com breve agitação a cada 2 min, finalizado por spin rápido). Os tubos foram colocados no rack magnético por 2 min. O sobrenadante contendo o mRNA (~320 µl) foi transferido para um novo microtubo (1,5 ml).

## **ANEXO B - Continuação.**

### **4. Concentração do mRNA**

Foram adicionados aos mRNA 1 µl de glicogênio (20 µg/µl), 3 µl de acetato de sódio 3 M e 750 µl de etanol absoluto. Os tubos foram brevemente agitados e mantidos por 12 h a -20 °C para a precipitação do mRNA. Após esta etapa, os tubos foram centrifugados a 11.000 xg durante 30 min para a formação do pellet de mRNA. O sobrenadante foi cuidadosamente descartado, sendo adicionados 500 µl de etanol 70% gelado (4 °C).

Os tubos foram centrifugados a 11.000 xg durante 5 min. Após isso, o sobrenadante foi aspirado e descartado. A lavagem com etanol 70% gelado foi novamente repetida. Os pellets de RNA foram secados em bomba de vácuo por 5 min para a remoção do etanol residual.

O mRNA foi ressuscitado em 10 µl de água DEPC e mantidos a -80 °C até o início da construção das bibliotecas.

### **5. Análise de pureza, integridade e concentração do mRNA**

A eficiência da depleção de rRNA, degradação e concentração do mRNA foram verificadas com a utilização do equipamento 2100 Bioanalyzer (Agilent Technologies), com o uso do kit Agilent RNA 6000 Pico (Agilent Technologies). A classe de ensaio utilizada foi a RNA total de eucariotos.

**ANEXO C – Tipos de RNAs ribossomais e coeficientes de sedimentação encontrados em plantas.**

<b>RNA ribossomal</b>	<b>Coefficiente de Sedimentação</b>
<b>Ribossomos citoplasmáticos</b>	
Sub-unidade maior	25S, 8S, 5S
Sub-unidade menor	18S
<b>Ribossomos cloroplastidiais</b>	
Sub-unidade maior	23S, 5S, 4.5S
Sub-unidade menor	16S
<b>Ribossomos mitocondriais</b>	
Sub-unidade maior	26S, 5S
Sub-unidade menor	18S, 5S

Fonte: Babu e Grassman, 2011. Assessing integrity of plant RNA with the Agilent 2100 Bioanalyzer.