

Wendell Jacinto Pereira

**Padrões diferenciais de metilação do DNA em  
escala genômica de clones de *Eucalyptus* em  
ambientes variáveis**

Brasília

2020



Wendell Jacinto Pereira

**Padrões diferenciais de metilação do DNA em escala genômica de clones de *Eucalyptus* em ambientes variáveis**

Tese de doutorado apresentada ao programa de Pós-Graduação em Ciências Biológicas (Biologia Molecular) do Instituto de Ciências Biológicas da Universidade de Brasília.

Universidade de Brasília – UnB

Instituto de Ciências Biológicas

Programa de Pós-Graduação em Ciências Biológicas (Biologia Molecular)

Orientador: Prof. Dr. Georgios Joannis Pappas Júnior

Brasília

2020

Wendell Jacinto Pereira

# **Padrões diferenciais de metilação do DNA em escala genômica de clones de *Eucalyptus* em ambientes variáveis**

Tese de doutorado apresentada ao programa de Pós-Graduação em Ciências Biológicas (Biologia Molecular) do Instituto de Ciências Biológicas da Universidade de Brasília.

Brasília, 24 de julho de 2020

---

**Prof. Dr. Georgios Joannis Pappas  
Júnior**  
Orientador - Universidade de Brasília

---

**Prof. Dr. Alexandre Siqueira Guedes  
Coelho**  
Universidade Federal de Goiás

---

**Prof. Dr. Dario Grattapaglia**  
Embrapa Recursos Genéticos e Biotecnologia,  
Universidade Católica de Brasília

---

**Prof. Dr. David John Bertioli**  
University of Georgia

---

**Prof. Dr. Matias Kirst**  
University of Florida

Brasília  
2020

# Agradecimentos

Agradeço à Universidade de Brasília, ao Instituto de Biologia e ao Programa de Pós-Graduação em Ciências Biológicas (Biologia Molecular) pela formação propiciada durante a realização deste doutorado. Em especial, agradeço aos membros docentes e aos servidores técnico-administrativos que propiciaram o ambiente necessário para a minha formação e para a realização deste trabalho. Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão de bolsa de doutorado durante a realização deste trabalho. Agradeço também à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão de bolsa de doutorado sanduíche durante a realização deste trabalho. Agradeço ainda à Universidade da Flórida por receber-me para a realização de doutorado sanduíche.

Agradeço ao prof. Dr. Georgios Joannis Pappas Jr. pela exímia orientação durante o período do doutorado, por todos os ensinamentos e direcionamentos e, especialmente, pela amizade. Agradeço também à Dra. Marília de Castro Rodrigues Pappas que, na prática, atuou como co-orientadora durante o desenvolvimento deste trabalho, por todas as orientações e trocas de conhecimentos e pela amizade. Agradeço ao Dr. Dario Grattapaglia pelas enriquecedoras discussões que ajudaram a delinear os experimentos e resultados descritos nesta tese. Agradeço aos membros do Laboratório de Genômica Computacional da Universidade de Brasília, Rodrigo Theodoro Rocha, Felipe Marques de Almeida e Vinícius Nattan Silva Lemos pelas colaborações, troca de conhecimentos e pela amizade.

Agradeço ao doutor Matias Kirst pela supervisão durante o meu período de doutorado sanduíche na Universidade da Flórida, pelas oportunidades de integração nos múltiplos projetos de pesquisa de seu laboratório, pelo fornecimento de recursos computacionais utilizados para a geração de parte dos resultados descritos nesta tese. Agradeço aos membros do Laboratório de Genômica Florestal da Universidade da Flórida pelas trocas de conhecimento e pela amizade.

Agradeço às doutoras Kátia Flávia Fernandes e Rosana Pereira Vianello, que foram minhas orientadoras em etapas anteriores da minha formação acadêmica, por despertarem em mim o desejo de seguir carreira acadêmica, por todos os ensinamentos e experiências. Agradeço a todos os professores, professoras, pesquisadores e pesquisadoras que de alguma forma contribuíram para minha formação intelectual. Agradeço a todos os colegas e amigos pelo incentivo e suporte nesses muitos anos dedicados à minha carreira acadêmica.

Agradeço aos meus pais, Luiz Jacinto da Silva e Fátima Martins Pereira da Silva pelo apoio e amor incondicional.

Finalmente, agradeço ao povo brasileiro pelo financiamento das instituições que

propiciaram a minha formação e a execução deste trabalho.

*“All our science, measured against reality, is primitive and childlike - and  
yet is the most precious thing we have.”*

— Albert Einstein

(Citado por Banesh Hoffmann em “Albert Einstein: Creator and Rebel”, 1972)

*“With time, my eyes began to open and I really started to understand nature.  
I also learned to love it.”*

— Claude Monet

(Le Temps newspaper, 26 de novembro de 1900)





# RESUMO

Nos últimos anos, a epigenética tem ganhado destaque na pesquisa biológica por constituir um nível de regulação adicional aos mecanismos genéticos, com potencial envolvimento no estabelecimento de fenótipos. Há grande expectativa de que seu melhor entendimento possa levar a uma melhor compreensão de uma série de processos biológicos complexos, como a resposta a estresses bióticos e abióticos e processos adaptativos. Dentre as modificações epigenéticas, a metilação do DNA é a mais bem estudada, em parte pela capacidade de detecção facilitada pelas tecnologias de sequenciamento de DNA de próxima geração (NGS). Neste trabalho, é apresentada uma abordagem analítica custo-efetiva para detecção de metilações do DNA a partir da aplicação da tecnologia *Methylation Sensitive DArT Sequencing* (MS-DArT-seq). Foi desenvolvido um protocolo computacional que promove a detecção de metilações do DNA e a caracterização do contexto genômico das metilações identificadas. Como estudo de caso, o protocolo computacional foi aplicado para investigação dos perfis de metilação do DNA em diferentes amostras coletadas de árvores da espécie *Eucalyptus grandis*. Milhares de sítios metilados foram identificados e, como esperado, metilações do DNA amostra-específicas foram identificadas, embora o perfil de metilação das amostras avaliadas seja bastante similar. A metodologia foi então aplicada em um desenho experimental mais complexo, no contexto do projeto TECHS (*Tolerance of Eucalyptus Clones to the Hydrous and Thermal Stresses*), com o objetivo de avaliar a associação do padrão de metilação do DNA com fatores ambientais. Cinco clones de *Eucalyptus*, híbridos interespecíficos de *E. grandis* e *E. urophylla*, plantados em dois ambientes com características edafoclimáticas contrastantes, foram utilizados na análise. Dentre os milhares de sítios de metilação de DNA identificados verificou-se o mesmo padrão anterior para um único indivíduo, ou seja, a grande maioria das metilações de DNA comuns aos dois ambientes e metilações ambiente-específicas detectadas para cada clone. Análises foram conduzidas para verificar associação entre as metilações e caracteres fenotípicos relacionados a crescimento (altura, volume e diâmetro a altura do peito). Centenas de sítios de metilação apresentaram associação significativa, contudo, os resultados indicam que o padrão de metilação do DNA evidência forte relação com a variação genética. Ademais, através de análise de associação com o ambiente, não foi possível estabelecer uma associação clara entre metilação e os sítios experimentais amostrados.

**Palavras-chaves:** metilação do DNA, epigenética, MS-DArT-seq, *Eucalyptus*, plasticidade fenotípica, TECHS.



# ABSTRACT

In recent years, epigenetics has gained prominence among the fields of biological research. As it constitutes an additional layer of regulation of the genetic mechanisms, with the potential to interfere in an organism phenotype, there is a high expectation that the understanding of its characteristics leads to a better comprehension of important biological processes, such as the response to biotic and abiotic stress and adaptation. Among epigenetic marks, DNA methylation is the best-studied, partly due to the detection capacity made possible by the advance of the next-generation sequencing (NGS). This work introduces a cost-effective analytical approach for DNA methylation detection using the Methylation Sensitive DArT Sequencing (MS-DArT-seq) technology. A computational protocol that promotes the identification of DNA methylation as well as the characterization of their genomic context was developed. As a proof-of-concept, the computational protocol was applied to ascertain the DNA methylation profiles of distinct samples of *Eucalyptus grandis*, the genotype sequenced for the species reference genome. Thousands of methylated sites were identified and sample-specific DNA methylation was identified, as expected, although the methylation profile of the evaluated samples is mostly similar. Additionally, the methodology was applied in a more complex experimental design provided by the TECHS (Tolerance of Eucalyptus Clones to the Hydrous and Thermal Stresses) program plant material. Five clones of *Eucalyptus*, *E. grandis* X *E. urophylla* hybrids, were used and thousands of methylated sites detected. Again, it was observed that the vast majority of the methylated sites was common to both environments, nevertheless environment-specific DNA methylation was also determined for each clone. Analyses were conducted to test for association between the methylated sites and traits related to growth (height, volume and diameter at breast height). Although hundreds of methylated sites showed a significant association with the measured traits, results suggest that the detected DNA methylation holds a strong connection with underlying genetic variation. Moreover, it was not possible to establish a clear connection between the presence of DNA methylations and the different environments.

**Key-words:** DNA methylation, epigenetics, MS-DArT-seq, *Eucalyptus*, phenotypic plasticity, TECHS.



# Lista de ilustrações

Figura 1 – Contextos de sequência de DNA passíveis de metilação na citosina . . . . .	29
Figura 2 – Fluxograma simplificado representando as etapas do estudo descritas em cada capítulo desta tese. . . . .	42
Figura 3 – Página inicial do artigo publicado no periódico PLOS ONE e baseado no estudo descrito neste capítulo. . . . .	44
Figura 4 – Interpretação dos resultados obtidos a partir da utilização dos isoesquisômeros <i>MspI</i> e <i>HpaII</i> para detecção de metilações do DNA. . . . .	48
Figura 5 – Padrões de metilação possíveis para o sítio de restrição dos isoesquisômeros <i>MspI</i> e <i>HpaII</i> . . . . .	49
Figura 6 – Esquema demonstrando as etapas para detecção de metilações de DNA via MS-DArT-seq. . . . .	51
Figura 7 – Fluxograma representando as etapas do estudo conduzido para a geração dos resultados descritos neste capítulo. . . . .	53
Figura 8 – Correção das contagens de fragmentos sobrepostos. . . . .	57
Figura 9 – Distribuição dos fragmentos sequenciados em relação ao seu tamanho. . . . .	63
Figura 10 – Distribuição das sequências geradas por MS-DArT-seq em relação aos elementos do genoma de <i>E. grandis</i> . . . . .	64
Figura 11 – Análise de co-localização com genes e regiões intergênicas dos conjuntos de fragmentos gerados pela redução genômica <i>in silico</i> usando diferentes combinações de enzimas de restrição. . . . .	65
Figura 12 – Distribuição dos sítios MSD ao longo dos 11 cromossomos do genoma de <i>E. grandis</i> . . . . .	66
Figura 13 – Reprodutibilidade de contagens entre as réplicas biológicas. . . . .	67
Figura 14 – Diagrama de Venn comparando os sítios metilados entre as três amostras. . . . .	69
Figura 15 – Análise de componentes principais demonstrando a diferença entre as amostras e de acordo com a biblioteca de sequenciamento implementada. . . . .	70
Figura 16 – Distância das metilações de DNA para o gene ou TE mais próximo . . . . .	72
Figura 17 – Gráfico de associação demonstrando os resíduos (Pearson) do teste qui-quadrado aplicado para a distribuição dos sítios metilados na proximidade de genes (10 kb). . . . .	73

Figura 18 – Diagrama de Venn comparando os elementos transponíveis metilados em cada amostra. . . . .	73
Figura 19 – Classificação dos elementos transponíveis que contêm ao menos uma metilação do DNA. . . . .	74
Figura 20 – Gráfico de associação demonstrando os resíduos (Pearson) do teste qui-quadrado aplicado para a distribuição de TEs metilados em classes. . . . .	75
Figura 21 – Diagrama de Venn comparando os genes metilados em cada amostra. . . . .	76
Figura 22 – Página inicial do artigo publicado no periódico <i>Forest Ecology and Management</i> e baseado no estudo descrito neste capítulo. . . . .	86
Figura 23 – Fluxograma representando as etapas do estudo conduzido para a geração dos resultados descritos neste capítulo. . . . .	91
Figura 24 – Comparação entre o volume dos clones nos ambientes seco e chuvoso. . . . .	95
Figura 25 – Comparação entre a produtividade dos clones em um mesmo ambiente. . . . .	97
Figura 26 – Distribuição dos sítios MSD ao longo dos cromossomos do genoma de referência de ( <i>E. grandis</i> ). . . . .	98
Figura 27 – Frequência de sítios metilados por clone em cada local de plantio de acordo com o tecido avaliado (folhas e xilema). . . . .	99
Figura 28 – Comparação entre todos os sítios metilados detectados nas amostras de folha de cada clone em cada sítio experimental . . . . .	100
Figura 29 – Comparação entre todos os sítios metilados detectados nas amostras de xilema de cada clone em cada sítio experimental . . . . .	101
Figura 30 – Comparações par-a-par dos sítios metilados ambiente-específicos. . . . .	103
Figura 31 – Análise de correspondências múltiplas (MCA) dos sítios MSD identificados nas amostras de xilema. . . . .	104
Figura 32 – Análise de correspondências múltiplas (MCA) dos sítios MSD identificados nas amostras de folha. . . . .	105
Figura 33 – Sítios metilados significativamente associados com volume (A e B), altura (C e D) e DAP (E e F) considerando o tecido foliar. . . . .	106
Figura 34 – Sítios metilados significativamente associados com volume (A e B), altura (C e D) e DAP (E e F) considerando o xilema. . . . .	107
Figura 35 – Diagramas de Venn demonstrando a comparação entre os conjuntos de sítios metilados significativamente associados (p-valor < 0.01) com os três caracteres fenotípicos avaliados neste estudo (volume, altura e DAP). . . . .	108

# Lista de tabelas

Tabela 1 – Definição dos parâmetros para determinação de metilação do DNA usando dados de MSD-DArT-seq pela comparação com dados de sequenciamento bissulfito. . . . .	68
Tabela 2 – Distribuição das metilações de DNA verificadas em cada amostra de acordo com o contexto genômico. . . . .	71
Tabela 3 – Tabela de contingência demonstrando a amostragem de sítios MSD no genoma de acordo com o contexto genômico. . . . .	71
Tabela 4 – Comparação das ocorrência de metilações nos sítios MSD em regiões gênicas e intergênicas. . . . .	71
Tabela 5 – Lista de termos GO enriquecidos para o conjunto de genes metilados de cada amostra. . . . .	77
Tabela 6 – Distribuição dos sítios metilados em cada clone de acordo com o contexto genômico. . . . .	102





# Lista de abreviaturas e siglas

AGO	Proteína Argonauta
CMTs	Cromometilases
Col	Columbia
DAP	Diâmetro na Altura do Peito
DArT	<i>Diversity Arrays Technology</i>
DArT-seq	<i>Diversity Arrays Technology Sequencing</i>
DCL	<i>DICER-LIKE</i>
DDM	<i>Decrease in DNA Methylation</i>
DME	<i>Demeter</i>
DML	<i>Demeter-like</i>
DMRs	<i>Differentially methylated regions</i>
Dnmts	DNA metiltransferases
DRM	DOMAINS-REARRANGED METHYLTRANSFERASES
dp	Desvio padrão
Embrapa CENARGEN	Embrapa Recursos Genéticos e Biotecnologia
EWAS	<i>Epigenome-wide Association Study</i>
epiRILs	<i>Epigenetic Recombinant Inbred Lines</i>
FDR	<i>False Discovery Rate</i>
FLG22	<i>bacterial flagellin-derived peptide flg22</i>
GBM	Metilações em corpos gênicos ou <i>Gene body Methylation</i>
GBS	<i>Genotyping By Sequencing</i>
GEA	<i>Gene-environment Association</i>
GME	Genes maternalmente expressos

GO	<i>Gene Ontology Consortium</i>
GPE	Genes paternalmente expressos
HELP-seq	<i>HpaII-tiny fragment Enrichment by Ligation-mediated PCR sequencing</i>
HPLC	<i>High performance liquid chromatography</i>
LFMM	<i>Latent Factor Mixed Models</i>
MBD-seq	<i>methyl-binding protein sequencing</i>
MCA	<i>Multiple Correspondence Analysis</i>
MeDIP-Seq	<i>Methylated DNA immunoprecipitation sequencing</i>
MEMS	<i>DNA Methylation Monitoring Sequence</i>
MET1	DNA Metiltransferase 1
MG	Minas Gerais
miRNAs	<i>Micro RNAs</i>
MPTHs	Modificações Pós-Traducionais de Histonas
MRE-seq	<i>Methylation-sensitive restriction enzymes sequencing</i>
MethylCap	<i>Methyl-CpG binding domain-based capture and sequencing</i>
MS-DArT-seq	<i>Methylation Sensitive DArT Sequencing</i>
MSD	<i>Methylation Sensitive DArT.</i>
MS-RAPD	<i>methyl sensitive random amplified polymorphic DNA</i>
MS-RFLP	Methylation-sensitive Restriction Fragment Length Polymorphism
MSAP	<i>Methylation Sensitive Amplification Polymorphism</i>
MSCC	<i>Methyl-Sensitive Cut Counting</i>
ncRNAs	<i>non-coding RNAs</i>
NGS	<i>Next Generation Sequencing</i>
pb	pares de bases
PCA	<i>Principal Component Analysis</i>
PCR	<i>Polymerase Chain Reaction</i>

Pol	DNA polimerase
PR	Paraná
QTL	<i>Quantitative Trait Locus</i>
RdDM	<i>RNA-directed DNA methylation</i>
RDR	<i>RNA-DEPENDENT RNA POLYMERASE</i>
ROS1	<i>Repressor of silencing 1</i>
RRBS-seq	<i>Reduced-representation bisulfite sequencing</i>
RT-qPCR	<i>Reverse Transcription Quantitative Polymerase Chain Reaction</i>
SAM	S-adenosyl-L-methionine
siRNAs	<i>small interfering RNAs</i>
SRA	<i>SET AND RING ASSOCIATED Domain</i>
SSD	<i>Single seed descent</i>
TECHS	<i>Tolerance of Eucalyptus Clones to the Hydrous and Thermal Stresses</i>
TEs	<i>Transposable elements</i>
kpb	Kilo pares de bases
WGBS	<i>Whole-Genome Bisulfite Sequencing</i>



# Sumário

<b>I</b>	<b>INTRODUÇÃO GERAL</b>	<b>25</b>
<b>1</b>	<b>INTRODUÇÃO</b>	<b>26</b>
<b>1.1</b>	<b>EPIGENÉTICA</b>	<b>26</b>
1.1.1	DEFINIÇÃO DO TERMO “EPIGENÉTICA”	26
<b>1.2</b>	<b>METILAÇÃO DE DNA EM PLANTAS</b>	<b>28</b>
1.2.1	ESTABELECIMENTO DA METILAÇÃO DO DNA EM PLANTAS	30
1.2.1.1	METILAÇÃO DO DNA <i>DE NOVO</i> E O ESTABELECIMENTO DAS METILAÇÕES DE DNA NO CONTEXTO DE SEQUÊNCIA “CHH”	31
1.2.1.2	MANUTENÇÃO DA METILAÇÃO DO DNA NOS CONTEXTOS DE SEQUÊNCIA “CG” E “CHG”	31
1.2.2	METILAÇÃO DO DNA DIRECIONADA POR RNAs (RdDM)	31
<b>1.3</b>	<b>DEMETILAÇÃO DO DNA</b>	<b>32</b>
<b>1.4</b>	<b>METILAÇÃO DE DNA EM CORPOS GÊNICOS (<i>GENE BODY DNA METHYLATION</i>)</b>	<b>32</b>
<b>1.5</b>	<b>PROCESSOS BIOLÓGICOS SOB REGULAÇÃO EPIGENÉTICA EM PLANTAS</b>	<b>33</b>
1.5.1	<i>IMPRINTING</i> GENÔMICO	33
1.5.2	VARIAÇÃO SOMACLONAL	34
1.5.3	VIGOR HÍBRIDO (HETEROSE)	35
1.5.4	METILAÇÃO DO DNA NA RESPOSTA AOS ESTÍMULOS AMBIENTAIS	36
1.5.4.1	RESPOSTA EPIGENÉTICA A ESTRESSES BIÓTICOS E ABIÓTICOS	36
<b>1.6</b>	<b>DEFINIÇÃO E ESCOPO DO ESTUDO</b>	<b>39</b>
	<b>OBJETIVOS</b>	<b>41</b>
	<b>OBJETIVO GERAL</b>	<b>41</b>
	<b>OBJETIVOS ESPECÍFICOS</b>	<b>41</b>
<b>II</b>	<b>CAPÍTULO 1 - DESENVOLVIMENTO DE PROTOCOLO ANALÍTICO PARA DETECÇÃO DE METILAÇÃO DE DNA A PARTIR DE DADOS DE MS-DART-SEQ</b>	<b>43</b>
<b>2</b>	<b>INTRODUÇÃO</b>	<b>45</b>
<b>2.1</b>	<b>DETECÇÃO EXPERIMENTAL DE METILAÇÃO DE DNA</b>	<b>45</b>
2.1.1	PRINCIPAIS METODOLOGIAS PARA DETECÇÃO	45

2.1.1.1	DETECÇÃO DE METILAÇÕES POR TRATAMENTO COM BISSULFITO DE SÓDIO COMBINADO AO NGS . . . . .	46
2.1.1.2	DETECÇÃO DE METILAÇÕES NO DNA BASEADA EM ENZIMAS DE RESTRIÇÃO . . . . .	47
2.1.2	DETECÇÃO DE METILAÇÕES NO DNA ATRAVÉS DA TECNOLOGIA DE SEQUENCIAMENTO DArT-seq . . . . .	50
2.1.2.1	SEQUENCIAMENTO DArT (DArT-seq) . . . . .	50
2.1.2.2	SEQUENCIAMENTO DArT SENSÍVEL A METILAÇÃO (MS-DArT-seq) . . . . .	50
	<b>OBJETIVOS . . . . .</b>	<b>52</b>
	<b>OBJETIVO GERAL . . . . .</b>	<b>52</b>
	<b>OBJETIVOS ESPECÍFICOS . . . . .</b>	<b>52</b>
<b>3</b>	<b>MATERIAIS E MÉTODOS . . . . .</b>	<b>54</b>
<b>3.1</b>	<b>MATERIAL VEGETAL . . . . .</b>	<b>54</b>
<b>3.2</b>	<b>IDENTIFICAÇÃO DOS SÍTIOS PASSÍVEIS DE METILAÇÃO . . . . .</b>	<b>54</b>
3.2.1	OBTENÇÃO DE DADOS VIA MS-DArT-seq . . . . .	54
3.2.2	MAPEAMENTO DAS SEQUÊNCIAS NO GENOMA DE REFERÊNCIA . . . . .	55
3.2.3	IDENTIFICAÇÃO DOS SÍTIOS DE METILAÇÃO . . . . .	55
<b>3.3</b>	<b>AVALIAÇÃO DOS FRAGMENTOS SEQUENCIADOS E SÍTIOS DE METILAÇÃO OBTIDOS POR MSD-DArT-seq EM RELAÇÃO À DISTRIBUIÇÃO NO GENOMA . . . . .</b>	<b>56</b>
<b>3.4</b>	<b>GERAÇÃO DAS CONTAGENS CORRESPONDENTES A CADA SÍTIO DE METILAÇÃO . . . . .</b>	<b>56</b>
<b>3.5</b>	<b>IDENTIFICAÇÃO DAS METILAÇÕES POR COMPARAÇÃO DE <i>HpaII</i> vs <i>MspI</i> . . . . .</b>	<b>57</b>
<b>3.6</b>	<b>PADRÕES DE METILAÇÃO DE DNA DAS DIFERENTES AMOSTRAS DE BRASUZ1 . . . . .</b>	<b>58</b>
3.6.1	REMOÇÃO DE SÍTIOS MSD COM DADOS FALTANTES . . . . .	58
3.6.2	IDENTIFICAÇÃO DO CONTEXTO GENÔMICO DAS METILAÇÕES VERIFICADAS . . . . .	59
3.6.3	ANOTAÇÃO FUNCIONAL DOS GENES QUE CONTÊM METILAÇÕES DE DNA . . . . .	59
<b>3.7</b>	<b>SEQUENCIAMENTO BISSULFITO . . . . .</b>	<b>59</b>
3.7.1	CONSTRUÇÃO DA BIBLIOTECA DE SEQUENCIAMENTO . . . . .	59
3.7.2	DETERMINAÇÃO DA PROPORÇÃO DE CONVERSÃO POR BISSULFITO . . . . .	60
3.7.3	IDENTIFICAÇÃO DAS METILAÇÕES DO DNA . . . . .	60
<b>3.8</b>	<b>DISTRIBUIÇÃO E REPRODUTIBILIDADE DO PROTOCOLO COMPUTACIONAL . . . . .</b>	<b>61</b>

<b>4</b>	<b>RESULTADOS</b>	<b>62</b>
<b>4.1</b>	<b>CARACTERÍSTICAS DO SEQUENCIAMENTO VIA MS-DArT-seq</b>	<b>62</b>
4.1.1	PRODUÇÃO DE DADOS E MAPEAMENTO NO GENOMA	62
4.1.2	EFICIÊNCIA DA REDUÇÃO GENÔMICA E ABRANGÊNCIA DA COBERTURA DE AMOSTRAGEM	62
4.1.2.1	EFICIÊNCIA DA REDUÇÃO GENÔMICA PELA DUPLA DIGESTÃO COM AS ENZIMAS DE RESTRIÇÃO <i>Pst</i> I- <i>Msp</i> I/ <i>Hpa</i> II	62
4.1.2.2	O PAPEL DA ENZIMA <i>Pst</i> I NA AMOSTRAGEM DE REGIÕES GÊNICAS	62
4.1.3	AMOSTRAGEM DOS SÍTIOS DE METILAÇÃO POR MS-DArT-seq	65
4.1.4	REPRODUTIBILIDADE ENTRE RÉPLICAS BIOLÓGICAS	66
<b>4.2</b>	<b>METILAÇÕES DETECTADAS NO DNA E CONTEXTO BIOLÓGICO</b>	<b>67</b>
4.2.1	REMOÇÃO DE DADOS FALTANTES PERMITE A COMPARAÇÃO DAS METILAÇÕES DE DNA ENTRE AMOSTRAS	67
4.2.2	METILAÇÕES DETECTADAS NO DNA	67
4.2.3	COMPARAÇÃO DAS METILAÇÕES ENTRE AMOSTRAS	69
4.2.4	CONTEXTO GENÔMICO DAS METILAÇÕES IDENTIFICADAS	70
4.2.5	METILAÇÕES DO DNA EM ELEMENTOS TRANSPONÍVEIS	72
4.2.6	METILAÇÕES DO DNA EM GENES CODIFICADORES DE PROTEÍNAS	74
4.2.7	ANOTAÇÃO FUNCIONAL E ENRIQUECIMENTO DE TERMOS DO <i>GENE ONTOLOGY</i> PARA OS GENES METILADOS	75
<b>5</b>	<b>DISCUSSÃO</b>	<b>78</b>
<b>5.1</b>	<b>IMPORTÂNCIA DA UTILIZAÇÃO DE <i>E. grandis</i> COMO MODELO EXPERIMENTAL</b>	<b>78</b>
<b>5.2</b>	<b>CARACTERÍSTICAS DOS DADOS GERADOS POR MS-DArT-seq</b>	<b>79</b>
5.2.1	QUALIDADE DAS SEQUÊNCIAS GERADAS POR MS-DArT-seq	79
5.2.2	AMOSTRAGEM DOS SÍTIOS DE METILAÇÃO EM RELAÇÃO AO CONTEXTO GENÔMICO	80
5.2.3	INFLUÊNCIA DE METILAÇÕES EM CORPOS GÊNICOS NOS PADRÕES DE METILAÇÕES VERIFICADOS POR MS-DART-SEQ	81
<b>5.3</b>	<b>METILAÇÕES DO DNA IDENTIFICADAS E COMPARAÇÃO ENTRE AS AMOSTRAS AVALIADAS</b>	<b>82</b>
5.3.1	COMPARAÇÃO DOS PERFIS DE METILAÇÃO ENTRE AS AMOSTRAS AVALIADAS	82
<b>5.4</b>	<b>COMPARAÇÃO COM OUTRAS METODOLOGIAS PARA DETECÇÃO DE METILAÇÃO DO DNA USANDO NGS</b>	<b>83</b>
	<b>CONCLUSÃO</b>	<b>84</b>

<b>III</b>	<b>CAPÍTULO 2 - PADRÕES DE METILAÇÃO DE DNA EM CLONES ELITE DE <i>EUCALYPTUS</i> PLANTADOS EM AMBIENTES CONTRASTANTES.</b>	<b>85</b>
<b>6</b>	<b>INTRODUÇÃO</b>	<b>87</b>
	<b>OBJETIVOS</b>	<b>90</b>
	<b>OBJETIVO GERAL</b>	<b>90</b>
	<b>OBJETIVOS ESPECÍFICOS</b>	<b>90</b>
<b>7</b>	<b>MATERIAL E MÉTODOS</b>	<b>92</b>
<b>7.1</b>	<b>MATERIAL VEGETAL</b>	<b>92</b>
7.1.1	CLONES SELECIONADOS PARA O ESTUDO	92
7.1.2	AMBIENTES EXPERIMENTAIS	92
<b>7.2</b>	<b>MS-DArT-seq PARA DETECÇÃO DE METILAÇÕES DO DNA</b>	<b>93</b>
<b>7.3</b>	<b>ANÁLISES ESTATÍSTICAS E DE ASSOCIAÇÃO ENTRE AS METILAÇÕES DO DNA IDENTIFICADAS E CARACTERES FENOTÍPICOS RELACIONADOS COM O CRESCIMENTO</b>	<b>93</b>
<b>7.4</b>	<b>DISPONIBILIDADE DE DADOS E REPRODUTIVIDADE DAS ANÁLISES</b>	<b>94</b>
<b>8</b>	<b>RESULTADOS</b>	<b>95</b>
<b>8.1</b>	<b>CRESCIMENTO DOS CLONES NOS AMBIENTES CONTRASTANTES</b>	<b>95</b>
<b>8.2</b>	<b>GERAÇÃO DE DADOS VIA MS-DArT-seq</b>	<b>96</b>
<b>8.3</b>	<b>DISTRIBUIÇÃO DAS METILAÇÕES DO DNA ENTRE CLONES</b>	<b>96</b>
<b>8.4</b>	<b>ANOTAÇÃO FUNCIONAL DOS GENES CONTENDO SÍTIOS METILADOS</b>	<b>98</b>
<b>8.5</b>	<b>METILAÇÕES DO DNA ASSOCIADAS AO AMBIENTE</b>	<b>99</b>
<b>8.6</b>	<b>ASSOCIAÇÃO ENTRE SÍTIOS METILADOS E CARACTERES FENOTÍPICOS RELACIONADOS COM CRESCIMENTO</b>	<b>103</b>
8.6.1	ANÁLISE FUNCIONAL DOS SÍTIOS METILADOS ASSOCIADOS COM MEDIDAS DE CRESCIMENTO	105
<b>9</b>	<b>DISCUSSÃO</b>	<b>109</b>
	<b>CONCLUSÃO</b>	<b>113</b>
	<b>CONSIDERAÇÕES FINAIS</b>	<b>114</b>



**REFERÊNCIAS** ..... 115



Parte I

INTRODUÇÃO GERAL

# 1 INTRODUÇÃO

## 1.1 EPIGENÉTICA

### 1.1.1 DEFINIÇÃO DO TERMO “EPIGENÉTICA”

O termo “epigenética” foi originalmente cunhado em 1942 por Conrad Hal Waddington para descrever os processos de desenvolvimento que levam à formação do fenótipo a partir do genótipo de um indivíduo (WADDINGTON, 2012). Dessa forma, originalmente, o termo descrevia o campo de estudos que atualmente é conhecido como desenvolvimento molecular (GREALLY, 2018). Na concepção de Waddington, os processos que constituem um sistema biológico deveriam estar conectados entre si, formando uma rede onde uma perturbação nos estágios iniciais provocaria efeitos cada vez mais intensos nos órgãos e tecidos desenvolvidos em estágios posteriores (DEICHMANN, 2016). Com o avanço científico e a descoberta de processos biológicos capazes de interferir na regulação da expressão gênica independente do genótipo, o termo “epigenética” passou a ter diferentes conotações (DEICHMANN, 2016).

Em uma delas, o termo epigenética refere-se a um nível de informação além da informação contida na sequência de DNA, fazendo com que o genoma seja interpretado de forma distinta nos diferentes tipos celulares e em diferentes fases do desenvolvimento. Por essa definição, epigenética englobaria todas as modificações na cromatina, modificações químicas nas bases do DNA, além de outros reguladores da transcrição que atuam no contexto da cromatina (GREALLY, 2018). Alternativamente, a epigenética pode ser definida como o estudo das mudanças na função gênica que são herdadas mitótica ou meioticamente e que não podem ser explicadas por mudanças na sequência de DNA (DIEZ; ROESSLER; GAUT, 2014; FELSENFELD, 2014; IWASAKI; PASZKOWSKI, 2014; DEICHMANN, 2016).

Em termos gerais, pode-se considerar que a informação epigenética é mediada por três processos principais:

- I) Metilação de DNA, usualmente metilação das citosinas, que está envolvida com o silenciamento de elementos repetitivos, principalmente os elementos transponíveis (TEs) e associada à repressão da expressão gênica;
- II) Modificações Pós-traducionais de Histonas (MPTHs) por adição de diferentes grupos químicos, como a metilação ou acetilação de resíduos específicos nas porções N-terminal das histonas, os quais determinam o estado conformacional da cromatina, podendo reprimir ou viabilizar a expressão gênica;

III) Processos mediados por RNAs não codificadores (ncRNAs, *non-coding RNAs*) que podem promover regulação da expressão gênica, além de estarem envolvidos nos processos de direcionamento de metilação de DNA (RdDM, do inglês *RNA directed DNA methylation*).

Os mecanismos epigenéticos são bastante comuns nos diferentes reinos, estando presentes de bactérias a mamíferos, passando por plantas, fungos e outros animais (DIEZ; ROESSLER; GAUT, 2014). Uma hipótese para sua proliferação seria que estes evoluíram para lidar com o conflito genômico resultante da expansão de TEs sendo, posteriormente, adaptados para outros usos (MAURER-ALCALÁ; KATZ, 2015). Embora as modificações epigenéticas sejam conservadas entre espécies, diferentes organismos possuem sistemas epigenéticos únicos (ZHONG, 2016) e o contínuo avanço nas pesquisas tem propiciado evidências do papel desses mecanismos em diversos processos.

Em função de serem organismos sésseis, as plantas precisam, obrigatoriamente, se adaptar para conviver com condições ambientais variáveis durante seu desenvolvimento. Nesse contexto, os mecanismos epigenéticos possuem grande importância para as espécies vegetais ao abrir uma janela para a regulação da expressão gênica a partir da transdução de sinais ambientais, podendo mediar alguns processos de adaptação (PIKAARD; SCHEID, 2014; SEYMOUR; BECKER, 2017). Talvez por isso, os mecanismos epigenéticos são altamente complexos em plantas. Por exemplo, em comparação aos mamíferos, as famílias gênicas que codificam modificadores epigenéticos são maiores em espécies vegetais (PIKAARD; SCHEID, 2014).

A descoberta de diversos mecanismos e processos biológicos, que atualmente fazem parte da epigenética, foi realizada em estudos com plantas. Dentre estas estão a distinção entre eucromatina e heterocromatina, a existência do fenômeno conhecido como paramutação, o *imprinting* parental e a dominância nuclear. Adicionalmente, levaram ao conhecimento dos processos de silenciamento de transgênicos, da variação somaclonal e da descoberta dos mecanismos de metilação do DNA direcionada por RNAs (RdDM, *RNA-directed DNA methylation*) (PIKAARD; SCHEID, 2014).

Nas próximas seções, serão descritos alguns exemplos do importante papel da epigenética na regulação biológica. Como o enfoque deste trabalho é o estudo da metilação de DNA em *Eucalyptus*, os exemplos descritos também se concentram nesta marca epigenética. Modificações pós-traducionais de histonas e RNAs não-codificadores, que *per se* são amplas áreas de estudo, serão mencionados apenas quando relevantes para o entendimento dos processos relacionados à metilação de DNA.

Até o momento, a variação epigenética existente tem sido classificada em três grupos: a) Variação epigenética obrigatória, cujas variantes estão completamente associadas com a variação genética existente; b) Variação epigenética facilitada, cuja ocorrência é

determinada probabilisticamente pelo genótipo e c) Variação epigenética pura (também conhecida como variação epigenética espontânea (JOHANNES; SCHMITZ, 2019)) a qual é atualmente demonstrável para metilação de DNA, onde essas variantes são determinadas por processos estocásticos que independem do genótipo (O'DEA et al., 2016).

As variantes da metilação de DNA que representam a variação epigenética pura são denominadas como “epialelos puros”, ou epimutações espontâneas. Este tipo de epimutação acontece devido a falhas nos processos de manutenção da metilação do DNA, ocasionando a perda ou ganho de metilação do DNA de maneira estocástica (JOHANNES; SCHMITZ, 2019). Adicionalmente, essas modificações podem ser responsivas à modulação ambiental e passadas pelos gametas para as gerações seguintes, originando variação epigenética herdável (WALKER et al., 2018; JOHANNES; SCHMITZ, 2019). Em teoria, a ocorrência de epimutações espontâneas pode fornecer vias alternativas para resposta aos estímulos ambientais de forma independente do conteúdo genético (SEYMOUR; BECKER, 2017). Contudo, uma vez que essas epimutações podem ser neutras, essa hipótese carece ainda de avaliação, principalmente, para o entendimento da capacidade adaptativa.

## 1.2 METILAÇÃO DE DNA EM PLANTAS

A metilação do DNA é uma das modificações epigenéticas mais estudadas, em particular a adição de um grupo metil ao quinto carbono da citosina, gerando a 5-metilcitosina (YONG; HSU; CHEN, 2016). Essa reação é realizada pela transferência de um grupo metil do cofator S-adenosyl-L-methionine (SAM), catalisada pelas enzimas DNA metiltransferases (Dnmts) (CHEN et al., 1991). Essa modificação epigenética possui papel central no controle da atividade genômica em plantas e mamíferos, estando envolvida em diversos processos biológicos como a regulação da expressão gênica, inativação do cromossomo X, regulação do desenvolvimento e diferenciação celular, *imprinting* genômico, silenciamento de elementos transponíveis e resposta a estresses bióticos e abióticos (DEICHMANN, 2016; ZHONG, 2016; BOUYER et al., 2017; SCHMITZ; LEWIS; GOLL, 2019).

Em plantas, a metilação do DNA ocorre em padrões não aleatórios, com grande concentração em regiões de elementos transponíveis e outros elementos repetitivos, como os presentes nas regiões centroméricas e nos aglomerados de genes que codificam para os RNAs ribossômicos 5S e 45S (PIKAARD; SCHEID, 2014). Nestes casos, a metilação é persistente e promove o fechamento da cromatina (inativação). No entanto, metilações podem ocorrer também em regiões promotoras e gênicas, podendo ser reversíveis e resultando em ativação ou repressão da expressão gênica.

A metilação das citosinas pode ocorrer em três contextos distintos de sequência – “CG”, “CHG”, ambos simétricos, e “CHH”, assimétrico, onde H representa qualquer base exceto a guanina – assim como apresentado na Figura 1 (NIEDERHUTH; SCHMITZ,

2014; PIKAARD; SCHEID, 2014; DELERIS; HALTER; NAVARRO, 2016; JOHANNES; SCHMITZ, 2019). Contudo, os padrões de metilação do DNA podem variar amplamente entre os organismos. Por exemplo, genomas de mamíferos e plantas são, normalmente, altamente metilados no contexto de sequência “CG” (GOUIL; BAULCOMBE, 2016). Além disso, enquanto em mamíferos existe a formação das regiões conhecidas como ilhas CpGs, caracterizadas pela alta concentração de sítios “CG” não metilados em regiões promotoras de genes, essas regiões não são comuns em plantas (PIKAARD; SCHEID, 2014).

Em plantas, os níveis de metilação variam entre os contextos de sequência, com níveis mais elevados no contexto “CG”. Frequentemente, apenas 10% ou menos das citosinas localizadas nos contextos de sequência “CHG” e “CHH” estão metiladas nos genomas de espécies vegetais (DELERIS; HALTER; NAVARRO, 2016; YONG; HSU; CHEN, 2016). Não obstante, as metilações de DNA que ocorrem nos contextos não-CG desempenham papel fundamental nas plantas, sendo relacionadas ao silenciamento de elementos repetitivos (GOUIL; BAULCOMBE, 2016).

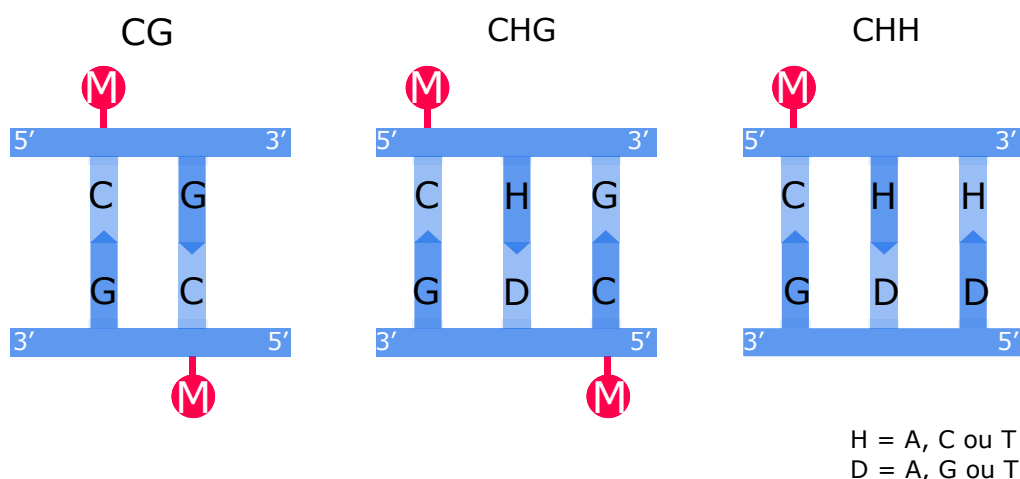


Figura 1 – Contextos de sequência de DNA passíveis de metilação na citosina. Os contextos de sequência “CG” e “CHG” são simétricos, de forma que a metilação do DNA pode ocorrer em ambas as fitas da molécula de DNA. Contudo, no contexto de sequência “CHH”, apenas uma das fitas pode ser metilada na região, devido à assimetria entre as sequências das duas fitas complementares. Imagem adaptada de *Plants’ Epigenetic Secrets*<sup>1</sup>.

Outra diferença marcante entre mamíferos e plantas é a transmissão das metilações de DNA durante a reprodução. Em mamíferos, após a fecundação e ainda nos estágios iniciais de desenvolvimento, ocorre um massivo processo de demetilação no genoma. Durante essa etapa, praticamente todas as metilações no DNA são removidas, permitindo que essas células retornem ao estado de pluripotência. Apenas algumas exceções são mantidas, como

<sup>1</sup> [www.the-scientist.com](http://www.the-scientist.com)

as metilações em genes que sofreram *imprinting*. As células destinadas à produção de gametas passam ainda por uma segunda etapa de demetilação (DEICHMANN, 2016).

Portanto, esta verdadeira cascata de demetilação aponta que a metilação de citosinas é uma marca epigenética não herdável em mamíferos (DEICHMANN, 2016). Contudo, é possível que mecanismos de restabelecimento dessas metilações possam ser influenciados por outros reguladores, como ncRNAs ou MPTs, que poderiam ser transmitidos durante a reprodução (PROKOPUK; WESTERN; STRINGER, 2015). Partindo dessa premissa, as modificações no perfil de metilação de DNA poderiam ser “herdadas” mesmo que indiretamente.

Em plantas, até o momento, não há evidências de que ocorram etapas de ampla demetilação do genoma. Pelo contrário, acumulam-se evidências de que, após a fecundação, na verdade, ocorra um reforço das metilações de DNA em elementos repetitivos em um processo mediado por pequenos RNAs via RdDM. Sendo assim, em plantas, considera-se a possibilidade de que as metilações sejam herdadas de forma estável, mesmo através da reprodução sexuada (YONG; HSU; CHEN, 2016; GALLUSCI et al., 2017; SPRINGER; SCHMITZ, 2017).

É importante ressaltar que muitas plantas podem se reproduzir de forma assexuada. Nesse tipo de reprodução, uma nova planta se desenvolve a partir das células meristemais presentes no chamado “botão vegetativo”. Por esse mecanismo, mudanças nos padrões de metilação de DNA nas células do meristema poderiam ser propagadas e mantidas na nova planta (RICHARDS et al., 2017a). De forma similar, os padrões de metilação podem ser transmitidos por propagação em cultura de tecidos. Adicionalmente, o próprio processo de propagação por cultura de tecidos representa uma condição adversa que, sabidamente, pode induzir modificações nos perfis de metilação do DNA. Alterações em decorrência de propagação *in vitro*, causando a hipometilação de um loco derivado da inserção de um TE (Karma), foram identificadas como causadoras de variação somaclonal responsável por um grande declínio da produtividade de dendezeiro (*Elaeis guineensis*), como demonstrado por Ong-Abdullah et al. (2015).

### 1.2.1 ESTABELECIMENTO DA METILAÇÃO DO DNA EM PLANTAS

Tanto em plantas quanto em mamíferos, a metilação do DNA ocorre pela atividade das enzimas DNA Metiltransferases. Os dois principais mecanismos de ação são o estabelecimento de metilações em citosinas previamente no estado não-metilado (metilação do DNA *de novo*) e a manutenção de metilações já estabelecidas (PIKAARD; SCHEID, 2014; LYKO, 2018; ZHANG; LANG; ZHU, 2018). Todas as DNA Metiltransferases utilizam um mecanismo similar para promover a adição do grupo metil ao substrato e seus motivos catalíticos são altamente conservados permitindo a identificação a partir de suas sequências de DNA (LYKO, 2018).



### 1.2.1.1 METILAÇÃO DO DNA *DE NOVO* E O ESTABELECIMENTO DAS METILAÇÕES DE DNA NO CONTEXTO DE SEQUÊNCIA “CHH”

A metilação de DNA *de novo* é direcionada pela ação de pequenos RNAs de interferência (siRNAs, *small interfering RNAs*) e ocorre em todos os três contextos de sequência (Figura 1) (MATZKE; MOSHER, 2014; PIKAARD; SCHEID, 2014; WENDTE; PIKAARD, 2017). Esse é o principal mecanismo responsável pelo estabelecimento da metilação de DNA, embora estas possam também ser estabelecidas por CMT2 (ZHANG; LANG; ZHU, 2018). Dos três contextos de sequência, apenas “CHH” não apresenta um mecanismo de manutenção descrito, principalmente, devido à sua assimetria em relação as fitas do DNA (ver Figura 1).

Supõe-se que a manutenção das metilações nesse contexto de sequência se dá pelo estabelecimento *de novo* a cada ciclo de replicação do DNA via RdDM. Adicionalmente, em transposons longos e normalmente localizados em regiões de heterocromatina, o estabelecimento pode ocorrer de forma independente da ação de pequenos RNAs (PIKAARD; SCHEID, 2014; MATZKE; MOSHER, 2014; SPRINGER; SCHMITZ, 2017; WENDTE; PIKAARD, 2017).

### 1.2.1.2 MANUTENÇÃO DA METILAÇÃO DO DNA NOS CONTEXTOS DE SEQUÊNCIA “CG” E “CHG”

A manutenção da metilação do DNA ocorre nas citosinas presentes nos contextos de sequência “CG” ou “CHG”. Esse processo baseia-se na simetria verificada nos sítios de metilação entre as fitas de DNA (SPRINGER; SCHMITZ, 2017; ZHANG; LANG; ZHU, 2018), conforme demonstrado na Figura 1. Esse processo é determinante para o restabelecimento dos padrões de metilação após a replicação, onde a fita mãe atua como molde para o estabelecimento das metilações de DNA na fita filha pela enzima Metiltransferase 1 (MET 1). Já a manutenção da metilação do DNA em citosinas no contexto de sequência “CHG” é realizada por outra metiltransferase (Cromometilase 3 – CMT3) em um ciclo de retroalimentação com modificações nas histonas (PIKAARD; SCHEID, 2014; SPRINGER; SCHMITZ, 2017).

## 1.2.2 METILAÇÃO DO DNA DIRECIONADA POR RNAs (RdDM)

Outra característica de grande importância para plantas é a ampla participação dos pequenos RNAs na regulação dos processos epigenéticos, principalmente, no estabelecimento das metilações de DNA para o silenciamento de elementos repetitivos, através de RdDM. Na RdDM, pequenos RNAs funcionam como guias para a maquinaria que promove as metilações do DNA, podendo levar ao silenciamento de regiões genômicas específicas, definidas por meio da complementariedade de sequência entre os pequenos RNAs e as regiões alvo (PIKAARD; SCHEID, 2014).

Além do silenciamento de TEs, RdDM também está envolvida na resposta a patógenos e estresses em geral. Finalmente, RdDM parece estar envolvida na reprodução sexual de angiospermas, especialmente na manutenção do balanço parental, e sua disrupção reduz drasticamente a viabilidade das sementes (CHOW; CHAKRABORTY; MOSHER, 2020).

### 1.3 DEMETILAÇÃO DO DNA

Os níveis de metilação do DNA verificados para uma espécie refletem a dinâmica entre o estabelecimento e a remoção dessa marcação epigenética (LI; KUMAR; QIAN, 2018), sendo a perda de metilações um componente importante na geração de epimutações espontâneas (JOHANNES; SCHMITZ, 2019). Tem sido demonstrado que essa dinâmica depende de diversas enzimas e vias biológicas distintas, podendo ainda ser alterada de acordo com o tecido, bem como com a fase de desenvolvimento em que a planta se encontra.

A remoção das metilações do DNA (demetilação) ocorre em plantas por dois processos distintos, remoção passiva e remoção ativa (PARRILLA-DOBLAS et al., 2019; LIU; LANG, 2020). A remoção passiva ocorre quando falhas nos mecanismos de manutenção da metilação do DNA, seja por inatividade ou por perda de eficiência, impedem a completa restauração após os ciclos de replicação do DNA (LI; KUMAR; QIAN, 2018). Dessa forma, a cada ciclo de replicação os níveis de metilação serão reduzidos, exceto se o processo for contrabalanceado pela instauração de metilações do DNA *de novo*.

A via de demetilação ativa, como o próprio nome indica, é desencadeada por ação enzimática e, em plantas, é desempenhada pelas enzimas conhecidas como DNA glicosilases, em conjunto com um mecanismo de reparo de DNA por excisão de bases (PIKAARD; SCHEID, 2014; LI; KUMAR; QIAN, 2018; PARRILLA-DOBLAS et al., 2019).

### 1.4 METILAÇÃO DE DNA EM CORPOS GÊNICOS (*GENE BODY DNA METHYLATION*)

Em plantas, a metilação de DNA é um processo normalmente associado com a repressão da transcrição e formação de heterocromatina. Contudo, nas angiospermas, existe uma classe de metilações que ocorre em genes altamente expressos, as chamadas metilações de DNA em corpos gênicos (*gene body DNA methylation*; GBM). GBM refere-se a genes que possuem um enriquecimento de metilações do DNA no contexto “CG” em suas regiões transcritas, acompanhado pela depleção dessas metilações nos sítios de início e término da transcrição (BEWICK; SCHMITZ, 2017; ZHANG; LANG; ZHU, 2018). Considera-se que um gene contém GBM quando há metilação em mais de 90% das citosinas no contexto “CG” e menos de 5% das citosinas no contexto “CHG” (VIDALIS et al., 2016). Nesse contexto, é importante ressaltar que genes podem ser enriquecidos em metilações de DNA em outras

configurações, que não podem ser classificadas como GBM. Por exemplo, genes que são enriquecidos em metilações do DNA no contexto “CG” em toda a sua extensão, incluindo os sítios de iniciação da transcrição, normalmente são silenciados e não se enquadram no conjunto de genes que contém GBM.

Os genomas das angiospermas possuem muitos genes com esse tipo de metilação. A característica distintiva destes genes é que são evolutivamente conservados, relacionados com processos biológicos basais e costumam ser maiores do que genes não metilados (BEWICK; SCHMITZ, 2017; SPRINGER; SCHMITZ, 2017; BEWICK et al., 2019). Contudo, o papel funcional das metilações de DNA em corpos gênicos não é bem esclarecido. Inicialmente, hipotizou-se que essas metilações poderiam atuar na regulação da expressão gênica. Contudo, estudos avaliando-se populações naturais de *Arabidopsis* (KAWAKATSU et al., 2016) não encontraram correlação entre GBM e expressão gênica. Outras hipóteses são que as GBM atuam impedindo transcrição aberrante a partir de promotores crípticos internos (TAKUNO; GAUT, 2012) ou durante os processos de *splicing*, aumentando a eficiência no processamento de pré-mRNAs (ZILBERMAN et al., 2008). Também foi demonstrado que as espécies *Conringia planisiliqua* e *Eutrema salsugineum* perderam totalmente GBM, apesar de manterem as metilações em “CG” nas regiões de elementos repetitivos, indicando que este não é um processo vital em plantas (BEWICK; SCHMITZ, 2017).

Uma hipótese é que GBM forneça variabilidade epigenética, o que poderia ter função relevante em casos de mudança na pressão seletiva gerada pelo ambiente. Assim, sendo uma marca epigenética herdável nas plantas, GBM poderia contribuir para adaptabilidade. Em estudo com mais de 1000 acessos de *A. thaliana*, foram observadas variações quanto à formação de GBM. Evidenciou-se que a presença de GBM correlacionava-se com a latitude de origem dos acessos, sugerindo uma relação com a adaptação de cada acesso ao seu ambiente de origem (KAWAKATSU et al., 2016). São ainda necessários novos estudos que consigam determinar se a ocorrência de GBM desempenha algum papel funcional relevante em espécies vegetais (BEWICK; SCHMITZ, 2017).

## 1.5 PROCESSOS BIOLÓGICOS SOB REGULAÇÃO EPIGENÉTICA EM PLANTAS

### 1.5.1 IMPRINTING GENÔMICO

O *imprinting* genômico, também conhecido como *imprinting* parental, foi um dos primeiros fenômenos epigenéticos relatados, sendo descoberto no começo da década de 70 por Kermicle (1970). Pode ser definido como a expressão gênica enviesada dependente do sexo do parental doador do alelo (BAI; SETTLES, 2015; RODRIGUES; ZILBERMAN,

2015; BATISTA; KÖHLER, 2020). O que caracteriza o *imprinting* genômico como um fenômeno epigenético é o fato de que o viés de expressão pode ocorrer mesmo quando ambos os alelos do gene são idênticos em relação à sequência de DNA (BAI; SETTLES, 2015; RODRIGUES; ZILBERMAN, 2015). Muitos dos genes sob *imprinting* parental foram relacionados com o desenvolvimento do endosperma e tamanho da semente (BAI; SETTLES, 2015; BATISTA; KÖHLER, 2020).

### 1.5.2 VARIAÇÃO SOMACLONAL

A propagação de plantas *in vitro* é uma importante ferramenta biotecnológica para a produção maciça de plantas, tanto para o plantio comercial quanto nas rotinas dos programas de melhoramento (US-CAMAS et al., 2014). Basicamente, a partir de um tecido vegetal já diferenciado, oriundo da planta com as características que se deseja propagar, é induzida a desdiferenciação do tecido para um conjunto de células pluripotentes para depois, a partir de estímulos controlados, promover a diferenciação dos tecidos para a formação de uma nova planta.

Através da propagação *in vitro*, é esperada a obtenção de clones geneticamente idênticos, facilitando a propagação das características de interesse. Contudo, diversos níveis de variação morfológica, bioquímica, genética e epigenética são observados em plantas obtidas por essa tecnologia, podendo ocasionar o surgimento de novas características ou a ruptura das características selecionadas, levando à perdas de produção (US-CAMAS et al., 2014). A ocorrência desse tipo de variação em plantas que foram expostas ao crescimento *in vitro* é recorrente e conhecida como variação somaclonal. Nos últimos anos, esta tem sido diretamente relacionada com mecanismos epigenéticos, principalmente, com a redução nos níveis de metilação do DNA mas também com o estabelecimento de MPTs (US-CAMAS et al., 2014; LEE; SEO, 2018).

Além do inerente estresse decorrente da indução da desdiferenciação e rediferenciação celular, diversos fatores podem induzir a modificação dos perfis de metilação do DNA durante o cultivo *in vitro* como a temperatura, estresse biótico e o tecido utilizado como doador (US-CAMAS et al., 2014). Como resultado, os perfis de silenciamento de genes e TEs podem ser alterados, levando à modificações na expressão gênica e à mobilização de TEs.

De fato, um exemplo marcante do papel epigenético no desenvolvimento de variação somaclonal foi relatado para o dendezeiro (*Elaeis guineensis*). A propagação das variedades de dendezeiro com alta produtividade do óleo é realizada em larga escala através da micropropagação clonal. Contudo, em diversas plantas, é observada uma variação somaclonal bastante recorrente com modificações morfológicas no fruto, detectada apenas na fase adulta e de produção, o que leva à grandes perdas de produtividade. Durante três décadas esse fenômeno foi estudado por diversos métodos de genética quantitativa, assim como por

análise genômica, sem que fosse possível a identificação da causa do fenótipo observado. No entanto, [Ong-Abdullah et al. \(2015\)](#) demonstraram, a partir da técnica conhecida como EWAS (*Epigenome-wide Association Study*), o papel da metilação do DNA no controle desse fenótipo.

[Ong-Abdullah et al. \(2015\)](#) determinaram que a demetilação de um TE (homólogo ao TE *karma* de arroz) é responsável por esse fenótipo. No dendezeiro, esse TE encontra-se inserido em um íntron do gene de um fator de transcrição (*DEF1*) e, quando demetilado, causa a geração de isoformas truncadas de mRNAs oriundos desse gene. Os autores geraram um teste de detecção de metilação nesse loco, permitindo o *screening* rápido e precoce para seleção das plantas com fenótipo produtivo para o cultivo, potencialmente, eliminando a ocorrência dessa variação somaclonal em campo e, portanto, as perdas de produção decorrentes. Conforme demonstrado nesse exemplo, para as espécies vegetais dependentes da propagação *in vitro*, é bastante relevante o entendimento das variações epigenéticas para o controle da variação somaclonal seja em casos prejudiciais ou em eventuais variantes benéficas.

### 1.5.3 VIGOR HÍBRIDO (HETEROSE)

Uma das principais estratégias utilizadas para aumentar a produção em espécies vegetais é a utilização de híbridos, derivados do cruzamento entre parentais de espécies diferentes, diferentes grupos heteróticos ou, ainda, de parentais de variedades com características diversas. Os híbridos, principalmente os utilizados comercialmente, apresentam um fenômeno denominado como “vigor híbrido” ou “heterose”, caracterizado pela mudança fenotípica de alguns caracteres na progênie em relação aos seus parentais, normalmente, levando ao aumento da biomassa e da produtividade ([GROSZMANN et al., 2013](#); [GREAVES et al., 2015](#)). Praticamente todas as espécies largamente cultiváveis apresentam algum nível de heterose que pode ser medido pela diferença entre a performance do híbrido e a média da performance dos dois parentais para um mesmo caractere ([HOCHHOLDINGER; BALDAUF, 2018](#)).

Além da influência da variação genética, as variantes epigenéticas, provavelmente, desempenham papel importante para o estabelecimento do vigor híbrido em plantas ([GROSZMANN et al., 2013](#); [GREAVES et al., 2015](#)). Alterações na metilação do DNA são verificadas em vários locos da geração F1 de cruzamentos intraespecíficos. Essas alterações são dependentes dos estados de metilação herdados de cada parental, ocorrendo em regiões que são diferencialmente metiladas (DMRs) entre os parentais ([GROSZMANN et al., 2013](#); [GREAVES et al., 2015](#)). Contudo, é extremamente complexo separar a variação epigenética da variação genética. Dessa forma, desenhos experimentais capazes de reduzir ou eliminar a variação genética, utilizando clones ou populações isogênicas, por exemplo, se fazem necessários. Nesse sentido, uma das abordagens mais promissoras constitui-se na utilização

de linhas puras recombinantes quase isogênicas mas epigeneticamente divergentes (epiRILs, *isogenic epigenetic recombinant inbred lines*).

Em um estudo recente, [Lauss et al. \(2018\)](#) utilizaram populações de epiRILs para demonstrar que a variação nos perfis de metilação do DNA é capaz de promover heterose, mesmo quando os parentais utilizados apresentam *backgrounds* genéticos quase isogênicos. Os autores verificaram que heterose existe para diversos dos fenótipos avaliados, como área foliar, altura total e dias para florescimento. Interessantemente, para diversos fenótipos, foram verificados híbridos que não só desviam da média dos parentais como apresentam média maior (ou menor) do que a média do parental com valor mais extremo. Adicionalmente, a partir de centenas de regiões diferencialmente metiladas, foi também possível a identificação de QTLs (*Quantitative Trait Locus*) associados com os fenótipos para os quais os híbridos apresentam heterose.

#### 1.5.4 METILAÇÃO DO DNA NA RESPOSTA AOS ESTÍMULOS AMBIENTAIS

Um dos aspectos mais interessantes da epigenética é a possibilidade de que ela esteja envolvida na resposta a estímulos ambientais devido à maleabilidade de seus mecanismos e efeitos. Sendo a modificação epigenética melhor estudada, diversos estudos têm sido conduzidos para o entendimento do papel da metilação de DNA na resposta a estresses bióticos e abióticos, bem como nos processos de adaptação e na ocorrência de plasticidade fenotípica. Embora evidências concretas do papel da metilação do DNA nesses processos permanecem escassas, nos últimos anos alguns exemplos foram reportados, conforme descrito nas sessões abaixo.

##### 1.5.4.1 RESPOSTA EPIGENÉTICA A ESTRESSES BIÓTICOS E ABIÓTICOS

Sendo organismos sésseis, as plantas precisam interagir e suportar ambientes variáveis e, em muitos casos, divergentes das condições favoráveis, ocasionando a ocorrência de estresse. Estes, podem ser classificados em duas categorias principais: estresses abióticos, em decorrência da interação com as condições edafoclimáticas locais, como seca, salinidade e temperaturas extremas, e bióticos, em decorrência das interações com outros organismos biológicos, como herbívoros, patógenos e outras plantas ([ALONSO; RAMOS-CRUZ; BECKER, 2019](#); [CHANG et al., 2020](#)).

Em relação ao papel epigenético na resposta aos fatores bióticos, prevalecem os estudos da interação planta-patógeno, os quais têm indicado um papel importante da metilação do DNA nos mecanismos de defesa ([ALONSO; RAMOS-CRUZ; BECKER, 2019](#)). Por exemplo, [Downen et al. \(2012\)](#) detectaram diferenças nos perfis de metilação do DNA em *A. thaliana* em resposta a duas cepas do patógeno *Pseudomonas syringae*, uma virulenta

e a outra avirulenta, e ao tratamento com ácido salicílico. Adicionalmente, os autores reportaram que o tratamento com ácido salicílico desencadeia modificações nas metilações em TEs localizados próximos a genes de resistência à doença. De forma similar, ainda em *A. thaliana*, Yu et al. (2013) identificaram demetilação do DNA em TEs próximos aos genes de resistência em resposta ao estresse biótico causado pelo estímulo com FLG22 (*bacterial flagellin-derived peptide flg22*). Downen et al. (2012), demonstraram ainda que mutantes deficientes em metilação do DNA são menos suscetíveis a *P. syringae*. Similarmente, Sánchez et al. (2016) demonstraram que mutantes de *A. thaliana* apresentando hipermetilação do DNA são mais suscetíveis à infecção por *Hyaloperonospora arabidopsidis*, enquanto mutantes hipometilados apresentam maior resistência. Le et al. (2014), chegaram a conclusões parecidas, demonstrando que triplo mutantes para DNA demetilases apresentam maior suscetibilidade ao fungo *Fusarium oxysporum* em *A. thaliana*.

A exposição de plantas a estresse abiótico tem sido, frequentemente, relacionada com modificações na estrutura da cromatina e com metilações no DNA. As alterações nos níveis de metilação do DNA podem acontecer tanto em escala genômica, quanto em locos específicos (SANTOS; FERREIRA; OLIVEIRA, 2017). Por exemplo, a exposição do milho ao frio promove a redução da metilação do DNA de maneira global (HU et al., 2012), enquanto a exposição de cultivares de arroz resistentes ao frio promove a demetilação de locos associados com genes de resposta ao frio (GUO et al., 2019). Ainda em arroz, o déficit de fosfato inorgânico está associado com a redução das metilações do DNA, principalmente, no contexto CHH e em TEs próximos a genes conhecidos pela resposta ao estresse em decorrência da falta deste nutriente (SECCO et al., 2015). Na mesma espécie, foi demonstrado que a exposição ao estresse salino promove demetilação global do genoma (FERREIRA et al., 2015). Por outro lado, a exposição de tabaco aos estresse salino, oxidativo, de frio ou por excesso de alumínio promove a diminuição dos níveis de metilação do DNA em sítios específicos como, por exemplo, no loco da proteína *glycerophosphodiesterase-like*, onde essa redução é correlacionada com um grande aumento transcricional (CHOI; SANO, 2007). Outras evidências para o papel da metilação do DNA na resposta ao estresse são observadas pela constatação que há diferença nos perfis de metilação do DNA entre indivíduos tolerantes e suscetíveis a um determinado estresse, conforme foi observado para diferentes espécies e estresses abióticos (SANTOS; FERREIRA; OLIVEIRA, 2017).

Evidências apontam para a herdabilidade dessas modificações constituindo, assim, o que vem sendo denominado como “memória epigenética” (ANNA CONDIA; MAGERØY; MARTINEZ, 2018). Contudo, acredita-se que essa memória seja apagada após uma ou poucas gerações, dependendo do organismo, se as condições estressantes não forem mantidas (ZHANG; LANG; ZHU, 2018). A perda dessa memória pode ser alcançada de forma passiva ou ativa, embora mecanismos regulatórios que atuem na remoção ativa ainda não tenham sido identificados.

Em plantas, parte da discussão em torno do papel da epigenética na resposta ambiental é guiada por evidências indiretas. Um exemplo é o observado para o cultivo de coníferas (*Picea abies*) na Escandinávia. Ao longo dos anos, os pomares de sementes foram sendo migrados ao sul, visando o aumento da produção devido ao plantio em condições climáticas mais amenas. Contudo, quando as sementes produzidas nesses pomares foram utilizadas para o plantio ao norte, onde as condições climáticas são mais severas, demonstraram menor resistência ao frio, a despeito do seu *background* genético ter origem nas coníferas do norte e, portanto, de genótipos adaptados ao clima mais severo (JOHNSEN; DIETRICHSON; SKARET, 1989). Esse fato foi replicado em cruzamentos controlados, eliminando assim a possibilidade de contaminação por pólen de árvores do sul (JOHNSEN; DIETRICHSON; SKARET, 1989). Dessa forma, especula-se que esse efeito seja uma memória do clima de plantio das árvores parentais dessas sementes, a qual provavelmente está relacionada com uma “memória epigenética” (GöMöRY et al., 2017).

Outro exemplo, também utilizando *Picea abies* mas em condições experimentais melhor controladas, é o estudo publicado por Carneros et al. (2017). Neste estudo, os autores demonstraram que plantas geradas por embriogênese, sendo este procedimento realizado em diferentes temperaturas (18 °C e 28 °C), apresentam diferenças de mais de duas semanas no início do brotamento, mensurado quando as plantas já tinham oito anos de idade, ou seja, oito anos após o tratamento.

Usando híbridos de *Populus*, Gourcilleau et al. (2010) demonstraram que o nível global das metilações do DNA, mensurado por HPLC, correlaciona-se positivamente com a performance dos híbridos sob déficit hídrico, enquanto também verificaram que existe interação entre os perfis de metilação e os genótipos utilizados. Ainda em *Populus*, Sow et al. (2018), investigaram amostras de três populações de *P. nigra* após aclimatização e déficit hídrico. Enquanto as populações não apresentaram diferenças no perfil de metilação, as amostras submetidas à seca tiveram aumento na porcentagem de citosinas metiladas. Adicionalmente, os níveis de metilação se correlacionam positivamente com a altura e biomassa apenas nas amostras submetidas à redução na irrigação.

Song et al. (2016), testaram a resposta de *Populus simonii* em diferentes condições de estresse abiótico, frio, calor, estresse salino e osmótico. Usando HPLC, eles identificaram o aumento da metilação de DNA global em todos os tratamentos (após 3 horas da aplicação do tratamento), sendo que o estresse por calor desencadeia o aumento mais pronunciado. Além disso, houve aumento na expressão de genes codificando diferentes enzimas relacionadas com o estabelecimento e manutenção de metilação de DNA e os grupos de genes diferencialmente expressos variaram de acordo com o tipo de estresse abiótico. Foram também verificadas metilações ocorrendo em miRNAs e lncRNAs que podem estar relacionadas com a resposta aos estresses avaliados. Interessantemente, os autores demonstraram ainda que, aproximadamente, 15% das metilações *de novo* em



reposta ao estresse são detectadas seis meses após os tratamentos. Também utilizando *P. simonii*, [Ci et al. \(2016\)](#) demonstraram, em populações naturais, que a variação na metilação de DNA é correlacionada com a distância geográfica das populações, enquanto a mesma correlação não é identificada para a variação genética, sugerindo papel das condições ambientais no estabelecimento dos perfis de metilação nessas populações.

Em conjunto, estes estudos apontam para uma clara conexão entre a metilação de DNA e os processos biológicos relacionados à resposta ambiental, motivando a ampliação dessas investigações para outras espécies perenes. Além disso, como demonstrado por [Champigny et al. \(2020\)](#), existe também relação entre a metilação do DNA e os fenótipos observados em *Populus*. Considerando-se que essa relação exista também em outras espécies florestais, faz-se relevante a expansão da investigação dessa modificação epigenética em outras espécies de interesse comercial, como as espécies do gênero *Eucalyptus*.

## 1.6 DEFINIÇÃO E ESCOPO DO ESTUDO

Diante da relevância da epigenética, o presente estudo utilizou o gênero *Eucalyptus* para estudar diferenças nos padrões de metilação de DNA seja em amostras do mesmo genótipo ou em genótipos diferentes em diferentes localidades.

Em função dos custos de sequenciamento para se obter um perfil completo das citosinas metiladas em um genoma, estudos amplos para sua detecção em espécies perenes são escassos, sendo circunscritos principalmente ao gênero *Populus* ([BRÄUTIGAM et al., 2017](#); [CHAMPIGNY et al., 2020](#)).

No intuito de explorar alternativas com custo acessível e de média escala, ou seja detecção direcionada de milhares de citosinas metiladas, o presente trabalho utilizou uma nova abordagem metodológica baseada em digestão por enzimas de restrição, denominada MS-DArT-seq, para a detecção dos perfis de metilação do DNA de três amostras da árvore BRASUZ1, a mesma sequenciada para a obtenção do genoma de referência da espécie *Eucalyptus grandis*.

Para a análise dos dados, foi desenvolvido um protocolo computacional combinando ferramentas de bioinformática de forma que todos os resultados descritos possam ser reproduzidos de forma automatizada. A combinação da técnica MS-DArT-seq com o protocolo de análise desenvolvido possibilitou a detecção de milhares de sítios de metilação no genoma de *E. grandis*. Adicionalmente, foram realizadas a avaliação do contexto genômico de cada sítio de metilação de DNA identificado, a classificação e anotação de genes e elementos transponíveis que contêm essas metilações.

A partir da validação realizada com as amostras de BRASUZ1, o protocolo de análise foi aplicado para identificação e análise comparativa dos perfis de metilação de DNA

de diferentes genótipos do gênero *Eucalyptus*. Foram utilizados cinco clones integrantes do programa TECHS (*Tolerance of Eucalyptus Clones to the Hydrous and Thermal Stresses*), sendo quatro híbridos entre *E. grandis* e *E. urophylla* e um clone pertencente à espécie *E. urophylla*, os quais são alguns dos clones comerciais mais plantados no Brasil. Os cinco clones foram plantados em dois sítios com características ambientais contrastantes e a metodologia MS-DArT-seq utilizada para avaliar a extensão dos efeitos ambientais nos seus perfis de metilação.

Finalmente, os resultados contidos neste estudo demonstram a viabilidade de aplicação da metodologia MS-DArT-seq para a detecção de metilação do DNA em larga escala a um custo acessível. A partir da metodologia de análise desenvolvida, é possível analisar os dados de forma rápida e automatizada, propiciando ampla exploração em relação aos aspectos biológicos relacionados com as metilações do DNA detectadas. Os resultados apresentados sugerem que tanto a metodologia MS-DArT-seq quanto o protocolo de análise desenvolvido possam ser adotados para outras espécies, principalmente, para estudos em organismos não-modelo e/ou estudos em escala populacional, para os quais a obtenção do metiloma completo se torna inviável.

# OBJETIVOS

## OBJETIVO GERAL

Investigar fenômenos epigenéticos em espécies do gênero *Eucalyptus* através da utilização de técnica de detecção de metilação em citosinas, visando catalogar o contexto genômico das metilações além de estudar padrões diferenciais de metilação em resposta ao ambiente de crescimento.

## OBJETIVOS ESPECÍFICOS

- Geração e disponibilização de protocolo computacional para identificação de metilação de DNA a partir de dados de MS-DArT-seq.
- Avaliação do perfil de metilação de DNA em diferentes amostras, xilema e folhas, do clone BRASUZ1, genótipo de referência da espécie *Eucalyptus grandis*.
- Estudar alterações nos padrões de metilação do DNA em clones plantados em ambientes com marcantes diferenças edafoclimáticas.
- Explorar a influência de fenômenos epigenéticos no crescimento de clones de *Eucalyptus* buscando correlacionar padrões diferenciais de metilação com caracteres fenotípicos associados ao crescimento.

# FLUXOGRAMA DO CAPÍTULO

Abaixo encontra-se uma visão global das etapas do presente trabalho.

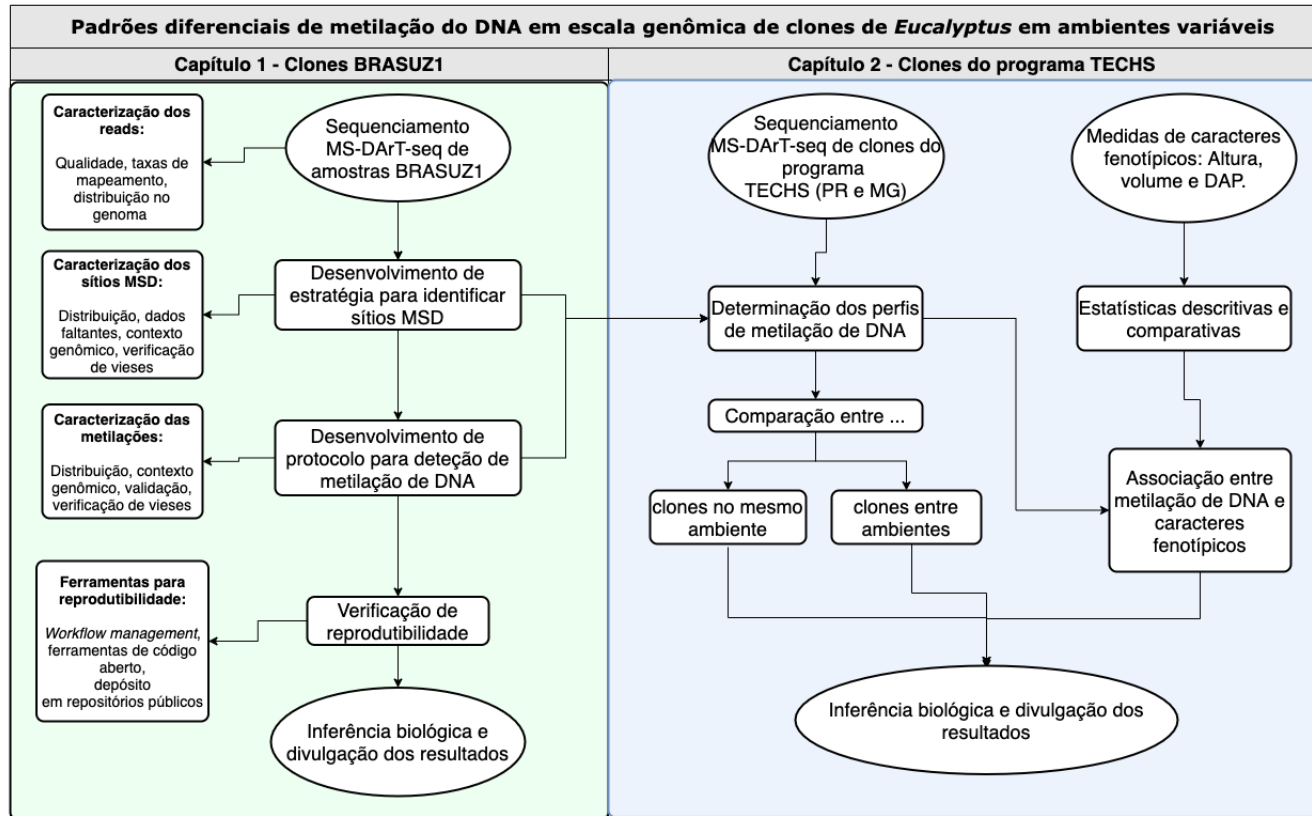


Figura 2 – Fluxograma simplificado representando a etapas do estudo descritas em cada capítulo desta tese.

## Parte II

# CAPÍTULO 1<sup>2</sup> - DESENVOLVIMENTO DE PROTOCOLO ANALÍTICO PARA DETECÇÃO DE METILAÇÃO DE DNA A PARTIR DE DADOS DE MS-DArT-seq.

---

<sup>2</sup> Os resultados descritos neste capítulo encontram-se também publicados na forma de artigo científico no periódico PLOS One (<https://doi.org/10.1371/journal.pone.0233800>), como demonstrado na Figura 3.

PLOS ONE

[PUBLISH](#)
[ABOUT](#)
[BROWSE](#)

OPEN ACCESS
PEER-REVIEWED

RESEARCH ARTICLE

## A cost-effective approach to DNA methylation detection by Methyl Sensitive DArT sequencing

Wendell Jacinto Pereira, Marília de Castro Rodrigues Pappas, Dario Grattapaglia, Georgios Joannis Pappas Jr

Published: June 4, 2020 • <https://doi.org/10.1371/journal.pone.0233800>

**Article**

**Authors**

**Metrics**

**Comments**

**Media Coverage**

**Abstract**

- 1. Introduction
- 2. Results
- 3. Discussion
- 4. Materials and methods

Supporting information

Acknowledgments

References

---

Reader Comments (0)

Media Coverage (0)

Figures

**Abstract**

Several studies suggest the relation of DNA methylation to diseases in humans and important phenotypes in plants drawing attention to this epigenetic mark as an important source of variability. In the last decades, several methodologies were developed to assess the methylation state of a genome. However, there is still a lack of affordable and precise methods for genome wide analysis in large sample size studies. Methyl sensitive double digestion MS-DArT sequencing method emerges as a promising alternative for methylation profiling. We developed a computational pipeline for the identification of DNA methylation using MS-DArT-seq data and carried out a pilot study using the *Eucalyptus grandis* tree sequenced for the species reference genome. Using a statistic framework as in differential expression analysis, 72,515 genomic sites were investigated and 5,846 methylated sites identified, several tissue specific, distributed along the species 11 chromosomes. We highlight a bias towards identification of DNA methylation in genic regions and the identification of 2,783 genes and 842 transposons containing methylated sites. Comparison with WGBS, DNA sequencing after treatment with bisulfite, data demonstrated a precision rate higher than 95% for our approach. The availability of a reference genome is useful for determining the genomic context of methylated sites but not imperative, making this approach suitable for any species. Our approach provides a cost effective, broad and reliable examination of DNA methylation profile on *MspI/HpaII* restriction sites, is fully reproducible and the source code is available on GitHub (<https://github.com/wendelljpereira/ms-dart-seq>).

Figura 3 – Página inicial do artigo publicado no periódico PLOS ONE e baseado no estudo descrito neste capítulo.

## 2 INTRODUÇÃO

### 2.1 DETECÇÃO EXPERIMENTAL DE METILAÇÃO DE DNA

#### 2.1.1 PRINCIPAIS METODOLOGIAS PARA DETECÇÃO

As metilações do DNA podem ser utilizadas como um sinal para a detecção de alterações na estrutura da cromatina que, por sua vez, podem ser relacionadas com o estágio de desenvolvimento, sinais ambientais ou estado fisiológico de células e indivíduos. Devido à importância biológica das metilações de DNA e a sua alta prevalência na maioria dos eucariotos, ao longo das últimas décadas houve um contínuo aumento no interesse pelo entendimento dos padrões de ocorrência dessas marcações epigenéticas nos genomas de diferentes espécies, em diferentes estágios de desenvolvimento e condições ambientais. Dessa forma, há grande interesse pelo desenvolvimento de metodologias capazes de detectar metilações no DNA. Idealmente, o método deveria ser capaz de amostrar sítios de metilação em escala genômica, de fácil aplicabilidade, escalabilidade e baixo custo. Contudo, até o momento, não há exemplos capazes de atender a todos os principais quesitos mencionados. Dessa forma, múltiplas abordagens foram criadas ou adaptadas recentemente, principalmente após o surgimento do sequenciamento de próxima geração (NGS - *Next Generation Sequencing*), permitindo a detecção das metilações do DNA com diferentes níveis de resolução (KURDYUKOV; BULLOCK, 2016; UMER; HERCEG, 2013; ZENG; HE; YI, 2019).

De maneira geral, esses métodos podem ser classificados em três classes: conversão por tratamento com bissulfito de sódio, enriquecimento por afinidade ou com enzimas de restrição (KIM; BAIDOURI; JACKSON, 2014). Adicionalmente, métodos de detecção baseados em sequenciamento de moléculas únicas, como as tecnologias da Pacific Biosciences e Oxford Nanopore, são capazes de identificar diretamente modificações epigenéticas presentes no DNA sequenciado, como as metilações de DNA (YONG; HSU; CHEN, 2016). Uma das grandes vantagens dessas metodologias é o sequenciamento sem a necessidade da amplificação do DNA por métodos baseados em PCR (*Polymerase Chain Reaction*). Ao permitirem o sequenciamento de moléculas únicas, além de propiciarem a obtenção de sequências muito maiores, essas metodologias permitem a leitura de modificações epigenéticas presentes no DNA sequenciado, como as metilações de DNA (YONG; HSU; CHEN, 2016). Contudo, devido seu recente estabelecimento e exiguidade de trabalhos científicos, estas não serão abordadas no presente estudo.

### 2.1.1.1 DETECÇÃO DE METILAÇÕES POR TRATAMENTO COM BISSULFITO DE SÓDIO COMBINADO AO NGS

A partir do advento do sequenciamento de próxima geração (NGS) os custos do sequenciamento das quatro bases canônicas do DNA em escala genômica se tornaram cada vez mais acessíveis.

Uma engenhosa derivação metodológica foi criada para permitir a identificação do estado de metilação da citosina por meio do tratamento do DNA com o composto bissulfito de sódio. O princípio para a identificação da metilação do DNA por essa metodologia está fundamentado no fato de que, quando tratadas com bissulfito de sódio, as citosinas não metiladas serão convertidas para uracila por deaminação, enquanto que as citosinas metiladas não são afetadas. A posterior amplificação por PCR faz com que as uracilas sejam transformadas em timinas (T), diferenciando-as das citosinas (C) metiladas. Assim, uma análise comparativa com o genoma de referência da espécie permite determinar as bases metiladas através do padrão de diferenças C/T (BIRNBAUM; ROUDIER, 2017). O sequenciamento de bissulfito em escala genômica (*Whole-genome bisulfite sequencing*), ou WGBS, é considerado como o padrão ouro da detecção de metilação de DNA, o método que propicia a melhor resolução gerando o mapa completo das metilações do DNA de uma amostra (conhecido como “metiloma”) (LISTER; ECKER, 2009; RIVERA; REN, 2013).

Contudo, existem algumas restrições relacionadas a essa técnica como a conversão incompleta das citosinas, fragmentação do DNA, a necessidade de um genoma de referência, além de vieses intrínsecos das tecnologias NGS, como a baixa amostragem de regiões enriquecidas em conteúdo GC. Adicionalmente, apesar do custo cada vez menor do sequenciamento, o custo total para a obtenção de um metiloma completo, incluindo a estrutura necessária para as análises de bioinformática, ainda é inacessível para muitos grupos de pesquisa.

Por isso, foram criadas alternativas como o sequenciamento bissulfito a partir de uma representação reduzida do genoma. A redução genômica pode ser realizada por diferentes métodos, incluindo abordagens para captura de regiões genômicas de interesse. A abordagem mais empregada é a redução da complexidade genômica através da digestão do DNA genômico com enzimas de restrição, seguido pela seleção de fragmentos, tratamento com bissulfito e sequenciamento NGS (GU et al., 2011; YONG; HSU; CHEN, 2016), usualmente nomeada como *Reduced-representation bisulfite sequencing*, ou RRBS-seq (GU et al., 2011). Mesmo com a redução de complexidade do genoma, métodos baseados em sequenciamento bissulfito continuam a ter custos bastante elevados para estudos com grande número de amostras.

Alternativas ao sequenciamento bissulfito foram desenvolvidas visando a viabilização de estudos com múltiplas amostras, detecção direcionada a regiões específicas do genoma



ou para a investigação em espécies com poucos recursos genômicos disponíveis. Estas metodologias se enquadram em duas categorias amplas: metodologias baseadas na seleção de fragmentos por afinidade e na digestão usando enzimas de restrição sensíveis à metilação.

Como revisado por Zeng, He e Yi (2019), métodos baseados em afinidade promovem a captura de fragmentos oriundos de regiões metiladas do genoma pela ligação de anticorpos (MeDIP-seq (DOWN et al., 2008)) ou proteínas que se ligam a sítios CG simetricamente metilados (MDB-seq (LI et al., 2010); MethylCap-seq (BRINKMAN et al., 2010)), seguido pelo sequenciamento desses fragmentos. As principais desvantagens dessas metodologias são o viés de captura para regiões hipermetiladas do genoma (geralmente associadas a heterocromatina) e a incapacidade de determinar exatamente quais citosinas do fragmento sequenciado apresentam metilação (KIM; BAIDOURI; JACKSON, 2014; BARROS-SILVA et al., 2018).

#### 2.1.1.2 DETECÇÃO DE METILAÇÕES NO DNA BASEADA EM ENZIMAS DE RESTRIÇÃO

Métodos para detecção da metilação do DNA a partir da utilização de enzimas de restrição têm sido utilizados por décadas (LINDSAY; BIRD, 1987) e são baseados na sensibilidade diferencial à metilação do DNA. Algumas enzimas de restrição conseguem clivar o DNA apenas se o sítio de restrição estiver livre de metilações, enquanto outras clivam o mesmo sítio de restrição, independente do estado de metilação. Adicionalmente, enzimas podem ter diferentes níveis de sensibilidade de acordo com o número de metilações e com a posição onde elas se encontram no sítio de restrição, bem como se as metilações estão presentes em ambas as fitas de DNA ou apenas em uma delas (YONG; HSU; CHEN, 2016; BIRNBAUM; ROUDIER, 2017).

Como revisado por Birnbaum e Roudier (2017), o par de isoesquisômeros *HpaII* e *MspI* foi o primeiro exemplo de utilização de enzimas de restrição para detecção de metilações no DNA. Essas enzimas reconhecem e clivam o mesmo sítio de restrição (5'-CCGG-3') diferindo entre si devido a sensibilidade da enzima *HpaII* à metilação na citosina interna do sítio de restrição, ou seja, caso a citosina esteja metilada a enzima não consegue clivar o sítio, enquanto *MspI* não apresenta sensibilidade a essa metilação. Adicionalmente, ambas as enzimas não conseguem clivar o sítio de restrição caso a citosina externa esteja metilada. A escolha desse par de enzimas é vantajosa porque, tanto em animais quanto em plantas, as metilações de DNA ocorrem mais frequentemente no contexto de sequência "CG" (FELSENFELD, 2014).

Resumidamente, o procedimento para avaliação da metilação de DNA a partir desses isoesquisômeros consiste na utilização de cada enzima para a geração de duas bibliotecas de fragmentos de DNA em paralelo a partir de uma mesma amostra de DNA. Dessa forma, duas populações de fragmentos são obtidas para cada amostra avaliada, uma digerida com *HpaII* (sensível à metilação) e outra com *MspI* (insensível à metilação). Em

seguida, as populações de fragmentos de cada biblioteca são comparadas pela observação do perfil de bandas formadas para a biblioteca de cada enzima após eletroforese de DNA em gel de agarose, conforme Figura 4-a (FULNEČEK; KOVAŘÍK, 2014). Conhecendo-se o perfil de sensibilidade à metilação das enzimas, é então possível relacionar a presença/ausência de bandas no gel com a presença de metilações nos sítios de restrição dos isoesquisômeros, como demonstrado na Figura 4-b (BEDNAREK; ORŁOWSKA; NIEDZIELA, 2017).

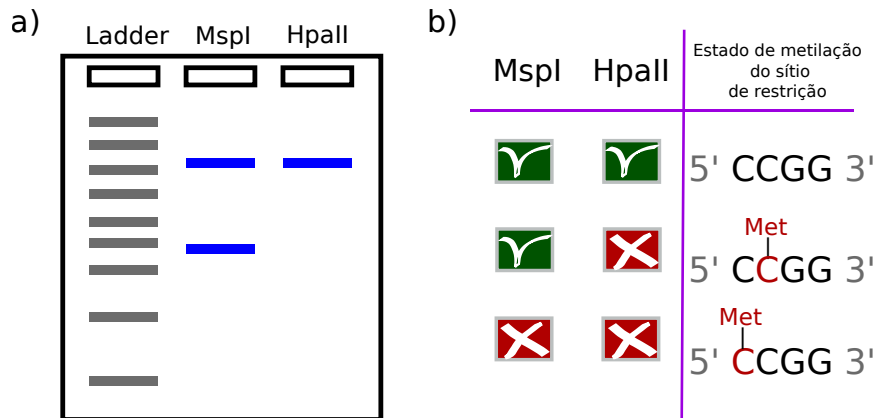


Figura 4 – **Interpretação dos resultados obtidos a partir da utilização dos isoesquisômeros *MspI* e *HpaII* para detecção de metilações do DNA.** a) Representação esquemática dos resultados possíveis ao comparar-se os fragmentos obtidos pela digestão com as enzimas *MspI* e *HpaII* após eletroforese de DNA em gel de agarose. b) Inferência sobre o estado de metilação do sítio de restrição que originou os fragmentos representados pelas bandas no gel.

Considerando-se os sítios de restrição do par de isoesquisômeros *HpaII* e *MspI*, em cada fita de DNA existem 2 citosinas passíveis de metilação, localizadas em dois contextos de sequência distintos e simétricos, “CG” e “CHG” (onde H=A, C ou T). Adicionalmente, cada contexto de sequência dentro do sítio de restrição pode ser verificado em três estados: a) completamente metilado, quando as citosinas em ambas as fitas estão metiladas, b) não metilado e c) hemi-metilado, quando somente a citosina de uma das fitas de DNA está metilada (Figura 5).

A utilização de metodologias baseadas em enzimas de restrição para detecção de metilações do DNA apresenta como grande vantagem o menor custo e facilidade de execução. Contudo, são limitadas devido ao pequeno número de sítios de restrição amostrados. Além disso, a menos que se realize o sequenciamento das bandas, é impossível determinar o contexto genômico no qual o sítio avaliado se insere, permitindo apenas uma avaliação do estado de metilação global de um indivíduo. No entanto, ainda hoje, a despeito dos avanços tecnológicos que permitiram a detecção em escala genômica e com uma precisão muito maior, a utilização dessa classe de métodos continua sendo uma opção bastante frequente, especialmente MSAP (*Methylation Sensitive Amplification Polymorphism*) e MS-RFLP (*Methylation-sensitive Restriction Fragment Length Polymorphism*) para os

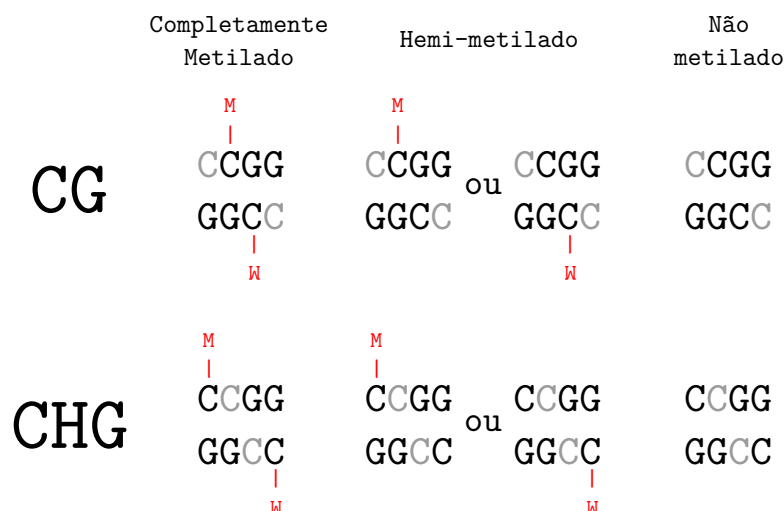


Figura 5 – Padrões de metilação possíveis para o sítio de restrição dos isoesquisômeros *MspI* e *HpaII*. As citosinas representadas em cinza não dizem respeito ao contexto de sequência demonstrado. Em adição aos estados de metilação representados, há situações onde ambas as citosinas estão metiladas ou não metiladas, além de diversas combinações das formas hemi-metiladas de cada citosina.

estudos de epigenética para as espécies não-modelo, que não contam com um genoma de referência, e, principalmente, devido ao seu baixo custo.

Uma alternativa às metodologias mencionadas surge a partir da combinação dos métodos baseados em enzimas de restrição com o sequenciamento de próxima geração, conhecidos como MRE-seq (*Methylation-sensitive restriction enzymes sequencing*). Nesses métodos, após a construção das bibliotecas de fragmentos para cada enzima do par de isoesquisômeros, os fragmentos são sequenciados por tecnologias NGS. A comparação de fragmentos é então realizada pela presença/ausência das sequências originadas em cada biblioteca. Assim, caso o sítio de restrição esteja metilado, somente haverá a presença de sequências para a biblioteca de fragmentos gerada com a enzima insensível à metilação, uma vez que a metilação bloqueia a ação da enzima sensível, impedindo a clivagem do DNA e, conseqüentemente, a geração do fragmento correspondente para o sequenciamento. Quando existe um genoma de referência para o organismo, é possível identificar o contexto genômico no qual a metilação se insere, através do mapeamento das sequências, o que é importante para o processo de inferência biológica.

Diversas metodologias como *methyl-sensitive cut counting* (MSCC, (BALL et al., 2009)), Methyl-seq (BRUNNER et al., 2009) e HELP-seq (ODA et al., 2009) são baseadas nesta abordagem e amplamente aplicadas. Nestes casos, é relatado o viés de amostragem para regiões menos metiladas do genoma. Por exemplo, os sítios amostrados por essas metodologias localizam-se em regiões distintas dos sítios amostrados por métodos baseados no enriquecimento por afinidade que, em geral, são enviesados para regiões hipermetiladas do genoma (MAUNAKEA et al., 2010). Uma das metodologias de MRE-seq visando a

detecção de metilações no DNA é o MS-DArT-seq (*Diversity Arrays Technology*), utilizada no contexto deste trabalho e descrita na próxima sessão.

## 2.1.2 DETECÇÃO DE METILAÇÕES NO DNA ATRAVÉS DA TECNOLOGIA DE SEQUENCIAMENTO DArT-seq

### 2.1.2.1 SEQUENCIAMENTO DArT (DArT-seq)

A estratégia adotada para o sequenciamento de DNA pela empresa *Diversity Arrays Technology* (DArT-seq) baseia-se na redução da complexidade genômica, através da fragmentação do genoma pela combinação de duas enzimas de restrição, uma de corte raro (*PstI*, que reconhece e cliva o sítio 5'CTGCAG3') e uma de corte frequente, seguida pela ligação de adaptadores na extremidade clivada pela *PstI*, usados para selecionar fragmentos gerados pela dupla digestão, amplificação e sequenciamento dos fragmentos resultantes em plataformas de sequenciamento de próxima geração (KILIAN et al., 2012; COURTOIS et al., 2013; MARK; KILIAN; DIERIG, 2013).

A escolha das enzimas de corte raro utilizadas para a digestão do DNA é realizada para cada espécie visando otimizar o perfil de tamanhos de fragmentos para o sequenciamento NGS. As sequências geradas são então processadas por protocolos de análise proprietários para descoberta de SNPs, bem como marcadores moleculares do tipo presença/ausência, chamados de marcadores DArT.

### 2.1.2.2 SEQUENCIAMENTO DArT SENSÍVEL A METILAÇÃO (MS-DArT-seq)

Visando possibilitar a identificação de metilações do DNA, a metodologia foi adaptada para o uso de enzimas de restrição de corte raro que apresentam sensibilidade a metilação. A técnica, denominada como *methyl-sensitive DArT sequencing* (MS-DArT-seq), baseia na comparação entre fragmentos gerados para as bibliotecas *MspI* e *HpaII*, de forma similar ao descrito na seção 2.1.1.2. A principal diferença é a utilização da enzima *PstI* em combinação com esses isoesquisômeros para a geração das duas bibliotecas de fragmentos (*PstI-MspI*) e (*PstI-HpaII*). As bibliotecas são sequenciadas por NGS, gerando duas populações de sequências que são então comparadas para a identificação das metilações do DNA (Figura 6).

Para cada um desses fragmentos, é feita a contagem de quantas vezes a sequência foi amostrada. Assim, ao comparar-se as contagens de cada leitura é possível inferir sobre os estados de presença/ausência de um determinado fragmento e, conseqüentemente, sobre o estado de metilação do sítio de restrição correspondente. A Figura 6 representa o procedimento de amostragem para um loco.

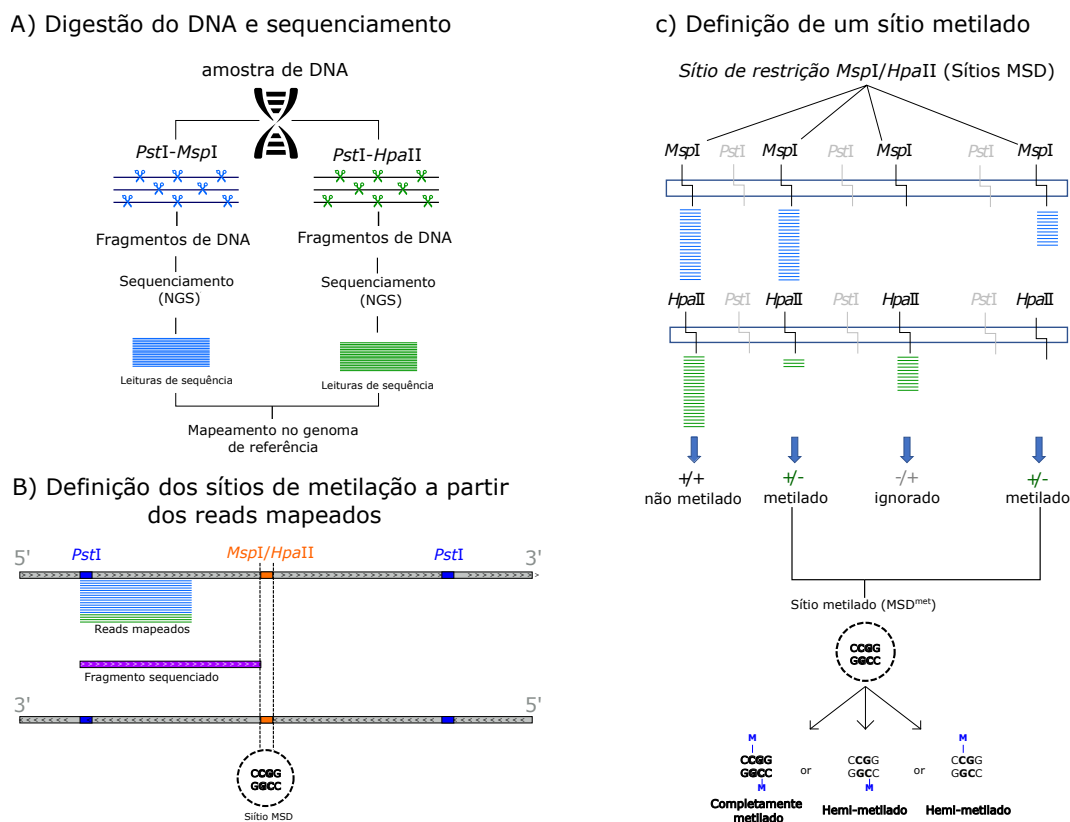


Figura 6 – Esquema demonstrando as etapas para detecção de metilações de DNA via MS-DArT-seq. Inicialmente, cada amostra de DNA é utilizada para o preparo de duas bibliotecas de sequenciamento, a primeira baseada na dupla digestão pelas enzimas de restrição *PstI* and *MspI* e a segunda pela combinação das enzimas *PstI* e *HpaII*. Ambas as bibliotecas são então sequenciadas gerando leituras de seqüências curtas (77pb) que são mapeadas em um genoma de referência (A). A partir das posições de mapeamento das seqüências, que sempre se iniciam no sítio de restrição *PstI*, é possível a identificação dos sítios onde será averiguada a presença de metilação, nomeado como sítio MSD (B). A comparação das contagens associadas a cada sítio de metilação nas duas bibliotecas de sequenciamento determina se um sítio está metilado. Contagens significativamente maiores na biblioteca *MspI* (Fold change  $\geq 2$  e FDR  $\leq 0.05$  após análise com DEseq2) indicam presença de metilação, enquanto contagens indistinguíveis estatisticamente indicam ausência de metilação. Para alguns poucos sítios MSD, foram verificadas contagens maiores para a biblioteca *HpaII*. Contudo, a interpretação deste perfil é controversa na literatura e, neste estudo, optou-se por considerá-los como sítios não metilados (C). Os sítios MSD que contêm metilações representam dois estados de metilação: completamente metilados, onde ambas as citosinas internas das duas fitas de DNA estão metiladas, ou hemi-metilado, onde apenas uma delas está metilada, não sendo possível distinguir dentre eles, uma vez que ambos bloqueiam a ação da enzima *HpaII* (C).

# OBJETIVOS

## OBJETIVO GERAL

Desenvolvimento de protocolo analítico para detecção de metilação de DNA a partir do sequenciamento MS-DArT-seq. Avaliação do protocolo desenvolvido e caracterização dos dados gerados através de prova de conceito utilizando amostras do clone de referência (BRASUZ1) da espécie *E. grandis*.

## OBJETIVOS ESPECÍFICOS

- Desenvolvimento de protocolo computacional para identificação de metilação de DNA utilizando ferramentas de código aberto.
- Caracterização dos dados gerados por MS-DArT-seq através de análises estatísticas e/ou representação visual. Comparação da metodologia com métodos bem estabelecidas e amplamente adotados para a detecção de metilação de DNA.
- Utilização do protocolo desenvolvido para determinação dos perfis de metilação de DNA de três amostras vegetais coletadas do clone BRASUZ1, genótipo utilizado para a geração do genoma referência da espécie *Eucalyptus grandis*.
- Caracterização do contexto genômico das metilações de DNA identificadas.

# FLUXOGRAMA DO CAPÍTULO

Abaixo encontra-se uma visão global das etapas do presente capítulo.

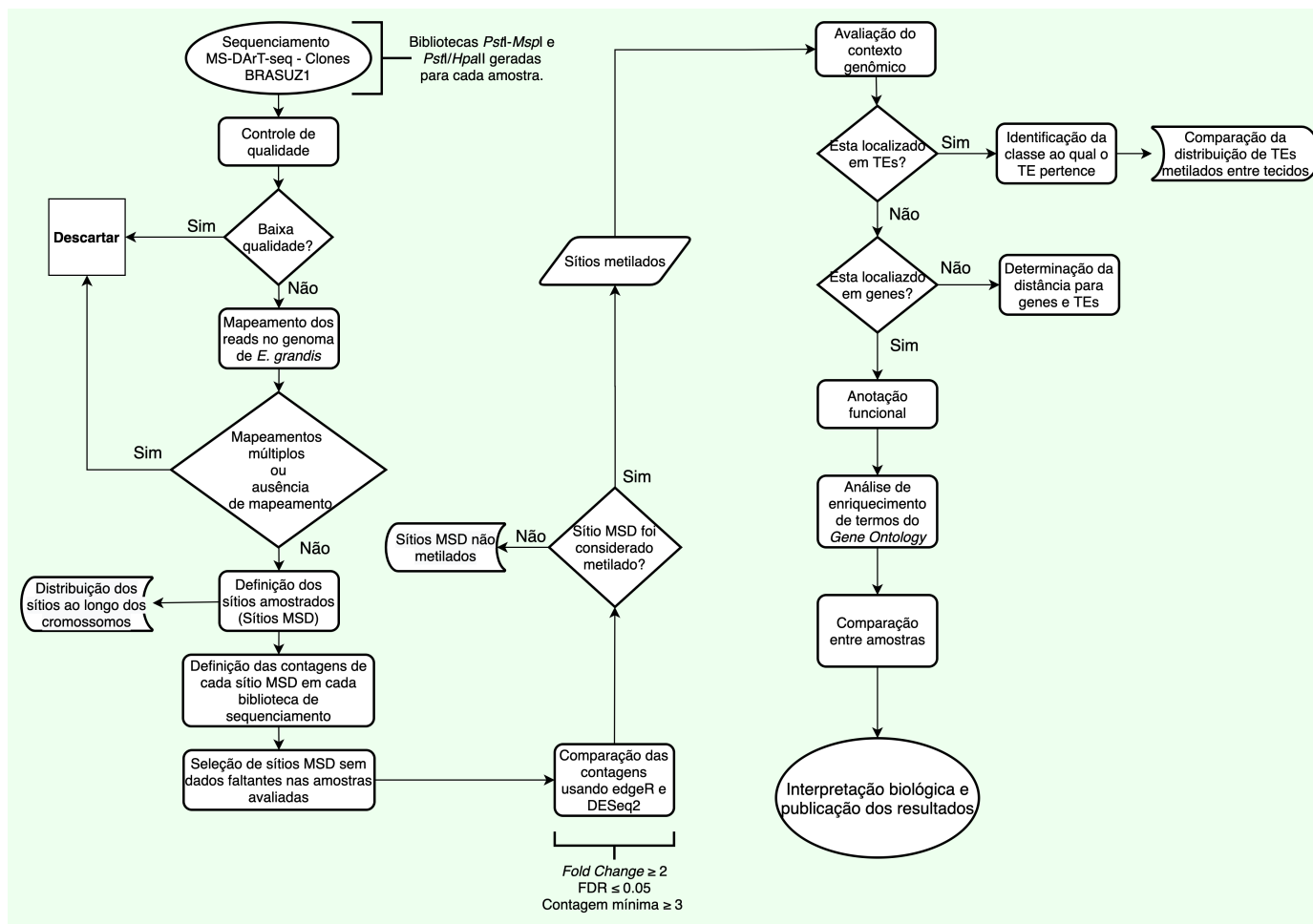


Figura 7 – Fluxograma representando as etapas do estudo conduzido para a geração dos resultados descritos neste capítulo.

# 3 MATERIAIS E MÉTODOS

## 3.1 MATERIAL VEGETAL

As amostras utilizadas neste estudo foram coletadas de clones da árvore BRASUZ1, a doadora do DNA sequenciado para a obtenção do genoma de referência para a espécie *Eucalyptus grandis* (MYBURG et al., 2014). Foram coletadas amostras de xilema (câmbio em desenvolvimento) e folhas em desenvolvimento e completamente desenvolvidas, aqui denominadas juvenis e adultas. Dessa forma, estabelecendo três amostras experimentais: “xilema”, “folhas adultas” e “folhas juvenis”.

Para cada grupo experimental, foram coletadas amostras de três réplicas biológicas (três árvores do mesmo genótipo BRASUZ1) com 4 anos de idade, plantadas na unidade da Embrapa Recursos Genéticos e Biotecnologia (Embrapa CENARGEN), Brasília - DF, Brasil. Em cada árvore (clone), a amostra de xilema foi coletada a partir da parte interna de um fragmento de casca do tronco e as amostras de folha foram coletadas do mesmo galho, localizado no lado de maior exposição à luz solar. A extração do DNA genômico de cada amostra foi realizada usando protocolo adaptado baseado no método CTAB (DOYLE; DOYLE, 1987).

## 3.2 IDENTIFICAÇÃO DOS SÍTIOS PASSÍVEIS DE METILAÇÃO

### 3.2.1 OBTENÇÃO DE DADOS VIA MS-DArT-seq

O DNA extraído foi enviado para a realização do sequenciamento MS-DArT-seq pela empresa Diversity Arrays Technology<sup>1</sup>, Canberra, Austrália, conforme descrito no tópico 2.1.2. Para cada amostra, foram geradas duas bibliotecas de sequenciamento, derivadas dos fragmentos gerados pela digestão com os pares de enzimas *Pst*I-*Msp*I e *Pst*I-*Hpa*II. O sequenciamento foi realizado em plataforma Illumina HiSeq 2000, *single-end*, produzindo leituras de sequência que iniciam-se no sítio de restrição da enzima *Pst*I e se prolongam por 77 pares de bases em direção ao sítio de restrição das enzimas *Msp*I ou *Hpa*II.

As sequências de todas as amostras foram recebidas e inspecionadas para controle de qualidade utilizando o software FASTQC<sup>2</sup>. Adaptadores e bases com baixa qualidade foram removidos com o software Trimmomatic<sup>3</sup> v.0.36 (BOLGER; LOHSE; USADEL,

---

<sup>1</sup> [www.diversityarrays.com](http://www.diversityarrays.com)

<sup>2</sup> [www.bioinformatics.babraham.ac.uk](http://www.bioinformatics.babraham.ac.uk)

<sup>3</sup> [www.usadellab.org](http://www.usadellab.org)



2014), aplicando os parâmetros pré-definidos, exceto pela utilização do parâmetro “SLIDINGWINDOW:5:25”.

### 3.2.2 MAPEAMENTO DAS SEQUÊNCIAS NO GENOMA DE REFERÊNCIA

Para a identificação dos sítios passíveis de metilação e detectáveis por MS-DArT-seq, doravante chamados sítios MSD, as sequências de cada leitura de sequenciamento foram mapeadas no genoma de *E. grandis* v.2.0 (BARTHOLOMÉ et al., 2015), disponível na base de dados Phytozome<sup>4</sup> (GOODSTEIN et al., 2012). Para o mapeamento, foi utilizado o software bowtie2<sup>5</sup> v.2.3.2 (LANGMEAD; SALZBERG, 2012). O mapeamento foi realizado controlando-se o número de *mismatches* permitidos. Dessa forma, foram executadas 4 etapas de mapeamento, variando-se o número de *mismatches* permitidos de 0 a 3, de forma crescente a cada etapa. Em todas as etapas, foi exigido o mapeamento da sequência completa e não foi permitida a inserção de *gaps* na sequência mapeada, nem na sequência correspondente no genoma. A cada etapa de mapeamento, somente as sequências não mapeadas na etapa anterior foram utilizadas. As sequências com múltiplos alinhamentos no genoma e as não mapeadas foram desconsideradas nas etapas seguintes do estudo.

### 3.2.3 IDENTIFICAÇÃO DOS SÍTIOS DE METILAÇÃO

Utilizando bedtools<sup>6</sup> v.2.27.1 (QUINLAN; HALL, 2010), as coordenadas das sequências mapeadas foram identificadas e a localização dos sítios de restrição investigados para metilação foi recuperada através de *scripts* elaborados no contexto desse estudo. De forma simplificada, a partir da posição e sentido (*strand*) das leituras mapeadas e conhecendo-se as posições de todos os sítios de restrição *MspI/HpaII* no genoma, é possível identificar o sítio onde as citosinas serão avaliadas para metilação (citosinas internas dos sítios *MspI/HpaII*). A determinação exata do sítio é realizada pela busca do sítio de restrição *MspI/HpaII*, a jusante, mais próximo da posição de mapeamento das sequências (Figura 6-B), não sendo determinada uma distância máxima para um sítio ser incluído. Neste estudo, os sítios de restrição *MspI/HpaII* avaliados para a presença de metilação são nomeados como sítios MSD.

Uma pequena parcela das sequências mapearam em regiões onde não foi possível encontrar um sítio de restrição *PstI* e foram desconsideradas na determinação dos sítios MSD. Uma característica importante dessa abordagem é o fato de que um mesmo sítio de restrição *MspI/HpaII* pode ser representado por mais de um fragmento sequenciado, se houver sítios *PstI* a montante em ambas as fitas da molécula de DNA. Essa redundância experimental foi considerada, conforme descrito abaixo.

<sup>4</sup> [phytozome.jgi.doe.gov](http://phytozome.jgi.doe.gov)

<sup>5</sup> [bowtie-bio.sourceforge.net](http://bowtie-bio.sourceforge.net)

<sup>6</sup> [bedtools.readthedocs.io](http://bedtools.readthedocs.io)

### 3.3 AVALIAÇÃO DOS FRAGMENTOS SEQUENCIADOS E SÍTIOS DE METILAÇÃO OBTIDOS POR MSD-DArT-seq EM RELAÇÃO À DISTRIBUIÇÃO NO GENOMA

A distribuição dos fragmentos sequenciados foi comparada com a distribuição dos fragmentos gerados por digestão *in silico* da sequência do genoma usando as enzimas (*PstI-MspI/HpaII*). Para essa finalidade, foi utilizado o teste de Kolmogorov-Smirnov para duas amostras. Uma vez que essas distribuições são discretas, o teste Kolmogorov-Smirnov foi aplicado usando 1000 simulações (Monte Carlo) para determinar o p-valor mais apropriado, conforme implementação do pacote `Matching`<sup>7</sup> (SEKHON, 2008), implementado na linguagem de programação R<sup>8</sup>.

O software ALFA<sup>9</sup> (BAHIN *et al.*, 2019) foi aplicado para averiguar a distribuição das sequências em relação aos elementos gênicos (UTRs, CDS e íntrons) e regiões intergênicas.

Para investigar se a utilização da enzima *PstI*, que também apresenta sensibilidade à metilação, na dupla digestão produz vieses na amostragem de fragmentos obtidos, o software Ologram<sup>10</sup> (FERRÉ *et al.*, 2020) foi aplicado. O Ologram avalia a significância estatística da co-localização dos fragmentos gerados *in silico* por *PstI-MspI/HpaII* com genes ou regiões intergênicas do genoma de *E. grandis*. Como comparação, a mesma análise foi realizada usando os fragmentos gerados pela dupla digestão *in silico* com as enzimas *EcoRI* e *BamHI* em substituição a *PstI*. Em cada análise, apenas fragmentos maiores do que 100 bases e menores do que 600 bases foram mantidos.

Para avaliar se os sítios MSD amostrados são uniformemente distribuídos ao longo do genoma, cada cromossomo foi dividido em janelas de diferentes tamanhos (1 Mb, 500 kb, 250 kb e 100 kb) e, para cada tamanho, um teste de chi-quadrado foi aplicado usando o número de sítios contidos em cada janela.

### 3.4 GERAÇÃO DAS CONTAGENS CORRESPONDENTES A CADA SÍTIO DE METILAÇÃO

A contagem das leituras de sequência mapeadas correspondentes a cada sítio MSD foi realizada utilizando o software `featureCounts`<sup>11</sup> v.1.6.2 (LIAO; SMYTH; SHI, 2014), considerando-se as bibliotecas *PstI-MspI* e *PstI-HpaII* separadamente. Nesta etapa, a

<sup>7</sup> [cran.r-project.org/web/packages/Matching](http://cran.r-project.org/web/packages/Matching)

<sup>8</sup> [cran.r-project.org](http://cran.r-project.org)

<sup>9</sup> [github.com/biocompibens/ALFA](https://github.com/biocompibens/ALFA)

<sup>10</sup> [dpuhier.github.io/pygtftk/ologram.html](https://dpuhier.github.io/pygtftk/ologram.html)

<sup>11</sup> <http://subread.sourceforge.net>

partir do mapeamento das sequências no genoma de referência, constatou-se a presença de sobreposição de fragmentos sequenciados, sendo este um fator confundidor na geração das contagens. Esses fragmentos são gerados por fatores diversos que podem interferir na clivagem de um ou mais sítios de restrição localizados a menos de 77 pares de bases um do outro. Devido à essas sobreposições, as sequências de todos os fragmentos se acumulam em uma mesma região sendo somadas e atribuídas a cada sítio MSD, superestimando as contagens reais conforme exemplificado na Figura 8. Essas contagens foram corrigidas aproveitando-se da informação das posições de mapeamento das sequências (Figura 8).

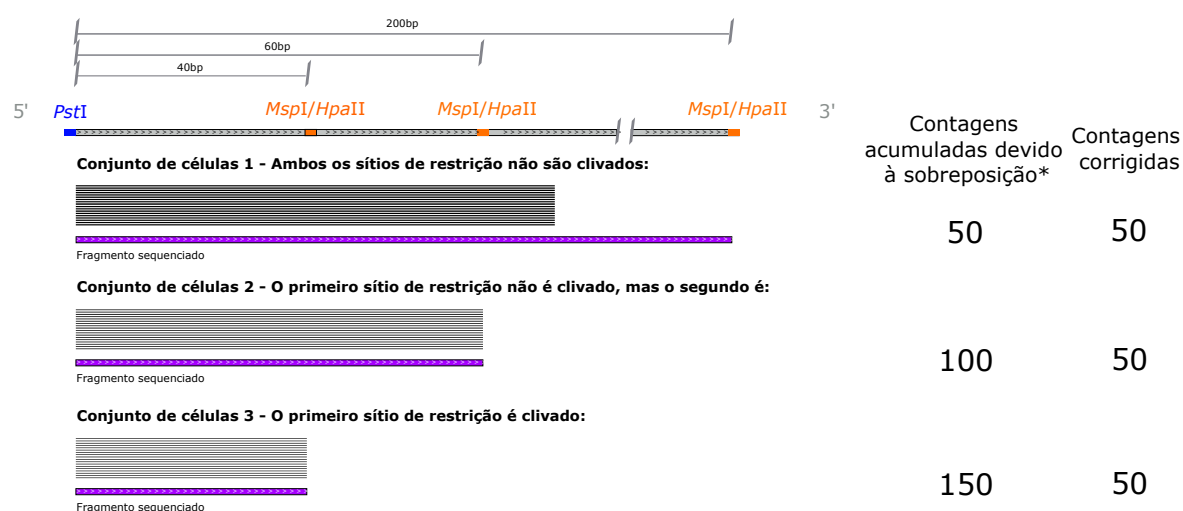


Figura 8 – **Correção das contagens de fragmentos sobrepostos.** A comparação entre os fragmentos sequenciados e os sítios de restrição identificados no genoma permite identificar que a formação de fragmentos que se sobrepõem. Neste exemplo, é ilustrada a correção das contagens que é realizada pela subtração entre as contagens dos fragmentos maiores e as contagens dos fragmentos menores.

### 3.5 IDENTIFICAÇÃO DAS METILAÇÕES POR COMPARAÇÃO DE *HpaII* vs *MspI*

Para a identificação dos sítios amostrados que contêm metilação do DNA, foi realizada a comparação das contagens geradas para cada leitura de sequenciamento nas duas bibliotecas de sequenciamento (“*PstI-MspI*” e “*PstI-HpaII*”). Essa etapa foi realizada utilizando-se os pacotes `edgeR`<sup>12</sup> (ROBINSON; MCCARTHY; SMYTH, 2010) e `DESeq2`<sup>13</sup> (LOVE; HUBER; ANDERS, 2014), ambos implementados na linguagem R e disponibilizados pelo Bioconductor<sup>14</sup>. A comparação das contagens e, portanto, a determinação das metilações foi realizada separadamente para cada grupo amostral.

<sup>12</sup> [bioconductor.org/packages/release/bioc/html/edgeR.html](https://bioconductor.org/packages/release/bioc/html/edgeR.html)

<sup>13</sup> [bioconductor.org/packages/release/bioc/html/DESeq2.html](https://bioconductor.org/packages/release/bioc/html/DESeq2.html)

<sup>14</sup> [bioconductor.org](https://bioconductor.org)

Visando garantir que contagens espúrias não seriam identificadas como uma diferença real entre as bibliotecas de fragmentos, levando à incorreta determinação da presença de metilações, um filtro foi implementado para remover fragmentos com contagens muito diminutas.

Para determinar o limite mínimo de contagem, bem como os valores de outros parâmetros necessários para determinar a presença de metilações, foi usado o sequenciamento bissulfito (BS-seq) da amostra de folha adulta como referência para validar o método MS-DArT-seq por uma análise comparativa das metilações identificadas. Foram usados três valores de *false discovery rate* (FDR;  $< 0,05$ ,  $< 0,01$  e  $< 0,001$ ), dois valores de *fold change* ( $\geq 1$  e  $\geq 2$ ) e três valores de contagem mínima ( $\geq 3$ ,  $\geq 5$  e  $\geq 10$ ). Para cada combinação de parâmetros, os sítios metilados detectados por MSD-DArT-seq foram comparados com os detectados por BS-seq.

A combinação de parâmetros que produziu o maior número de metilações e a maior taxa de validação com os dados de bissulfito foi utilizada ao longo desse estudo, conforme descrito a seguir. Apenas sequências que continham contagem  $\geq 3$  nas bibliotecas geradas com a enzima de restrição *MspI* (não sensível à metilação) foram mantidas na análise. Foram considerados como sítios metilados apenas os sítios cujas contagens entre *MspI* e *HpaII* foram determinadas como significativamente diferentes pelos algoritmos de ambos *edgeR* e *DESeq2* usando *fold change*  $\geq 2$  e  $FDR < 0,05$  como parâmetros.

Quando existe mais de um fragmento sequenciado representando um mesmo sítio MSD, a seguinte abordagem foi aplicada. Se a diferença entre as contagens das bibliotecas *PstI-MspI* e *PstI-HpaII* de, ao menos um dos fragmentos, foi considerada significativa, o sítio MSD foi considerado metilado.

## 3.6 PADRÕES DE METILAÇÃO DE DNA DAS DIFERENTES AMOSTRAS DE BRASUZ1

### 3.6.1 REMOÇÃO DE SÍTIOS MSD COM DADOS FALTANTES

MS-DArT-seq produz milhares de sítios MSD que são exclusivos para cada amostra. Contudo, para que a comparação entre as diferentes amostras da árvore BRASUZ1 fosse fidedigna, foi necessária a exclusão desses sítios da análise. Dessa forma, apenas sítios MSD sem dados faltantes em nenhuma das amostras foram mantidos neste estudo. Como, para a biblioteca *PstI-HpaII*, são esperadas contagens iguais a zero, apenas a biblioteca *PstI-MspI* foi utilizada para a verificação de dados faltantes (sítios com contagem igual a zero foram considerados como dados faltantes).

Após a identificação das metilações de DNA pelos programas *edgeR* e *DESeq2*, o teste Q de Cochran foi aplicado para verificar se a probabilidade de ocorrência de metilação

do DNA em um sítio MSD difere de acordo com a amostra avaliada.

### 3.6.2 IDENTIFICAÇÃO DO CONTEXTO GENÔMICO DAS METILAÇÕES VERIFICADAS

Para cada metilação do DNA verificada, foi feita a identificação do contexto genômico na qual ela está inserida. A posição genômica da metilação do DNA foi comparada com a a anotação de genes e elementos transponíveis no genoma de *E. grandis* usando bedtools v2.27.0. De forma similar ao realizado para os sítios metilados, o teste Q de Cochran foi aplicado para verificar se a probabilidade de um gene ou TE ser metilado, de acordo com os dados de MS-DArT-seq, difere de acordo com a amostra avaliada.

### 3.6.3 ANOTAÇÃO FUNCIONAL DOS GENES QUE CONTÊM METILAÇÕES DE DNA

A anotação funcional dos genes contendo metilações foi realizada por duas ferramentas, blast2GO<sup>15</sup> v4.1 e BioMart<sup>16</sup> (disponível no pacote biomaRt<sup>17</sup>), a partir de diferentes bases de dados. Inicialmente, todos os genes do genoma de *E. grandis* foram anotados com ambas as ferramentas. Em seguida, as anotações foram combinadas em um único conjunto de dados, utilizado para a anotação dos subconjuntos de genes de interesse. Adicionalmente, a anotação do conjunto total de genes do genoma foi utilizada para gerar a referência para a análise de enriquecimento de termos do *Gene Ontology Consortium*. Nesse ponto, visando evitar a ocorrência de falsos positivos na determinação dos termos enriquecidos, apenas os genes que continham ao menos um sítio de metilação amostrado por MS-DArT-seq foram utilizados como grupo de referência na análise de enriquecimento de termos, realizada com o pacote clusterProfiler<sup>18</sup>.

## 3.7 SEQUENCIAMENTO BISSULFITO

### 3.7.1 CONSTRUÇÃO DA BIBLIOTECA DE SEQUENCIAMENTO

Sequenciamento bissulfito foi realizado para a amostra de folhas adultas. A amostra foi enviada para a empresa *Zymo Research*<sup>19</sup> que realizou o sequenciamento e análise inicial dos dados, conforme descrito a seguir. Inicialmente, 500 ng de DNA foram usados para digestão com 2 unidades do kit *Zymo Research's (ZR) dsDNA Shearase TM Plus*. Os fragmentos gerados foram então purificados usando o kit *Zymo Research (ZR) DNA Clean*

<sup>15</sup> [www.blast2go.com](http://www.blast2go.com)

<sup>16</sup> [www.biomart.org](http://www.biomart.org)

<sup>17</sup> [bioconductor.org/packages/release/bioc/html/biomaRt.html](http://bioconductor.org/packages/release/bioc/html/biomaRt.html)

<sup>18</sup> [bioconductor.org/packages/release/bioc/html/clusterProfiler.html](http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html)

<sup>19</sup> [www.zymoresearch.com](http://www.zymoresearch.com)

and Concentrator TM. O tratamento bissulfito e a ligação de adaptadores foi realizada usando o kit *EZ DNA Methylation – Lightning kit*. O sequenciamento foi realizado em plataforma Illumina HiSeq 2000, usando Illumina TruSeq.

### 3.7.2 DETERMINAÇÃO DA PROPORÇÃO DE CONVERSÃO POR BISSULFITO

A identificação dos níveis de metilação de cada citosina no genoma foi também realizada pela empresa *Zymo Research*. As sequências geradas foram mapeadas no genoma de referência de *E. grandis* usando um *software* proprietário, baseado no *software* Bismark<sup>20</sup> (KRUEGER; ANDREWS, 2011). Os níveis de metilação de cada citosina foram definidos como o número de sequências reportando uma citosina (C), dividido pelo total de sequências reportando uma citosina (C) ou timina (T). Desta forma, os valores variam entre 0 e 1, sendo 0 quando todas as sequências possuem T (todas as citosinas não metiladas e convertidas) e 1 quando todas as sequências reportam C (todos os fragmentos metilados nessa posição, impedindo a conversão). Contudo, na prática, outros fatores interferem nesses valores, como a taxa de eficiência da conversão por bissulfito e a presença de múltiplos tipos celulares na amostra que, com frequência, apresentam diferentes perfis de metilação.

### 3.7.3 IDENTIFICAÇÃO DAS METILAÇÕES DO DNA

Devido aos fatores mencionados acima, apenas a proporção de conversão não é suficiente para a identificação de metilações do DNA, sendo necessária a aplicação de métodos estatísticos para determinar se uma citosina é, de fato, metilada. Após o recebimento dos dados contendo a proporção de conversão de cada citosina, a determinação das metilações foi realizada de acordo com método baseado em Schultz et al. (2015). Inicialmente, a taxa de conversão por bissulfito foi calculada usando o genoma do bacteriófago lambda como controle durante o preparo da biblioteca de sequenciamento. Como as moléculas de DNA representando esse genoma são sintéticas, elas não contêm metilação de DNA e, portanto, espera-se que todas as citosinas sejam convertidas. A taxa de não conversão verificada para o DNA controle é então extrapolada para a amostra sequenciada.

Em seguida, para cada citosina amostrada por BS-seq, um teste binomial foi aplicado usando a soma do número de sequências suportando a presença da metilação e a soma do número de leituras indicando a ausência de metilação do DNA. Em cada teste, a probabilidade de sucesso foi igual a taxa de não conversão por bissulfito. Os p-valores foram corrigidos para múltiplos testes aplicando o procedimento de Benjamini-Hochberg para o cálculo do FDR (*false discovery rate*).

<sup>20</sup> [www.bioinformatics.babraham.ac.uk/projects/bismark](http://www.bioinformatics.babraham.ac.uk/projects/bismark)

## 3.8 DISTRIBUIÇÃO E REPRODUTIBILIDADE DO PROTOCOLO COMPUTACIONAL

Todas as etapas descritas para a análise foram realizadas pela elaboração de um protocolo de análise computacional automatizado, integrando diversos programas que foram criados para atender às etapas de análise do estudo, além de programas já consolidados para tarefas específicas, como mapeamento de sequências de DNA, por exemplo. Os programas criados no contexto desse estudo foram escritos nas linguagens R e Python. Alguns dos programas elaborados e/ou utilizados fazem uso de pacotes pré-existentes e disponíveis para instalação nos repositórios oficiais de cada uma das linguagens.

Todas as ferramentas utilizadas no estudo possuem código aberto. A única exceção é o *software* Blast2GO v.4.1, utilizado para a anotação dos genes do genoma de *E. grandis* e que não foi integrado ao *workflow* automatizado, uma vez que é um software proprietário com código fechado. Dessa forma, após a anotação com Blast2GO, a tabela de resultados foi incorporada ao protocolo computacional e utilizada para a anotação do subconjunto de genes. A anotação do Blast2GO é utilizada como entrada para o protocolo de análise desenvolvido, podendo ser obtida pelo usuário através da versão gratuita do programa. Adicionalmente, a anotação pode ser realizada usando o banco de dados BioMart através do pacote `biomaRt`, esta etapa também foi incorporada à ferramenta desenvolvida.

Para o controle e automação das tarefas foi utilizado o gerenciador de *workflows* Snakemake<sup>21</sup> (KÖSTER; RAHMANN, 2012). Os parâmetros de controle e execução são fornecidos em um arquivo de configuração, lido pelo Snakemake, flexibilizando assim a modificação dos critérios da análise, caso necessário. Todos os *scripts* utilizados pela ferramenta estão disponíveis para acesso no site GitHub<sup>22</sup>: <<https://github.com/wendelljpereira/ms-dart-seq>>. Adicionalmente, os dados necessários para a execução da análise estão disponíveis na base de dados figshare<sup>23</sup>: <<https://doi.org/10.6084/m9.figshare.10305431>>.

---

<sup>21</sup> [bitbucket.org/snakemake/snakemake](https://bitbucket.org/snakemake/snakemake)

<sup>22</sup> [github.com](https://github.com)

<sup>23</sup> [figshare.com](https://figshare.com)

## 4 RESULTADOS

### 4.1 CARACTERÍSTICAS DO SEQUENCIAMENTO VIA MS-DArT-seq

#### 4.1.1 PRODUÇÃO DE DADOS E MAPEAMENTO NO GENOMA

Ao todo foram obtidas mais de 45 milhões de sequências geradas a partir das três amostras utilizadas neste estudo (folha adulta, folha juvenil e xilema) e considerando-se as sequências das duas bibliotecas geradas para cada amostra (*PstI-HpaII* e *PstI-MspI*). Após o controle de qualidade e mapeamento no genoma, identificou-se que 7.869.635 (17,40%) sequências mapearam em múltiplas posições e 5.282.139 (11,68%) não mapearam no genoma com os critérios utilizados. Essas leituras foram excluídas das etapas posteriores, restando 32.086.893 (70,92%) sequências que mapearam unicamente no genoma de *E. grandis* e foram utilizadas para a determinação dos sítios MSD e detecção das metilações do DNA. Embora a cobertura de sequenciamento varie para cada região amostrada e entre amostras, foi observada uma alta cobertura para a maioria das regiões sequenciadas por essa metodologia, com média de aproximadamente 80x e mediana de aproximadamente 30x para as regiões sequenciadas.

#### 4.1.2 EFICIÊNCIA DA REDUÇÃO GENÔMICA E ABRANGÊNCIA DA COBERTURA DE AMOSTRAGEM

##### 4.1.2.1 EFICIÊNCIA DA REDUÇÃO GENÔMICA PELA DUPLA DIGESTÃO COM AS ENZIMAS DE RESTRIÇÃO *PstI-MspI/HpaII*

Visando avaliar o grau de redução genômica atingido e a capacidade de amostragem da metodologia MSD-DArT-seq, foi realizada a comparação entre os fragmentos sequenciados e fragmentos gerados através da dupla digestão *in silico* do genoma usando as enzimas *PstI-MspI/HpaII*. Conforme demonstrado na Figura 9, mais de 80% dos fragmentos sequenciados possuem tamanho entre 48 e 574 pares de bases com média de 289 pb (d.p.  $\sim$  711,3; CV  $\sim$  2,46). Comparando-se com o conjunto de fragmentos produzidos *in silico*, estes representam 47,13% de todos os fragmentos possíveis nesta mesma faixa de tamanho.

##### 4.1.2.2 O PAPEL DA ENZIMA *PstI* NA AMOSTRAGEM DE REGIÕES GÊNICAS

No método MS-DArT-seq, a enzima *PstI*, sensível à metilações do DNA no contexto CHG, é utilizada visando o direcionamento da produção de fragmentos de DNA a partir



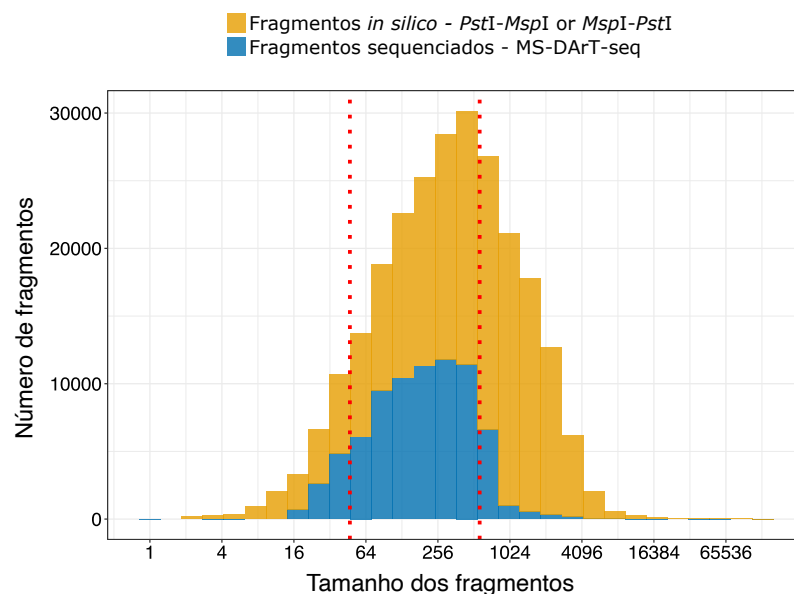


Figura 9 – **Distribuição dos fragmentos sequenciados em relação ao seu tamanho.** Em amarelo, encontram-se representados todos os fragmentos gerados pela digestão *in silico* do genoma de *E. grandis* e que contém em suas extremidades os sítios de restrição das enzimas *MspI* e *PstI*. Em azul, estão representados os fragmentos sequenciados. As linhas tracejadas em vermelho representam o 1º e o 9º decil. O tamanho dos fragmentos está representado em escala logarítmica de base 2 para facilitar a visualização.

de regiões hipo-metiladas do genoma, como a eucromatina, enriquecida em genes. De fato, é possível verificar a maior abundância de sequências oriundas de regiões gênicas, como demonstrado na Figura 10. Em números absolutos, a maior parte das sequências é mapeada em regiões intergênicas. Contudo, após a normalização das contagens de acordo com o conteúdo genômico, é clara a seleção preferencial de regiões gênicas pelo sequenciamento MS-DArT-seq (Figura 10).

Para verificar se a utilização da enzima *PstI* é o principal fator nesse direcionamento, o programa Ologram foi aplicado para testar se existe co-localização entre os fragmentos sequenciados e elementos gênicos (éxons, CDS e UTR). Adicionalmente, a análise de co-localização foi também realizada entre os fragmentos sequenciados e regiões intergênicas.

Os resultados demonstram que os fragmentos gerados pela dupla digestão com as enzimas *PstI-MspI/HpaII* são fortemente associados com regiões gênicas e que não há associação com as regiões intergênicas do genoma (Figura 11). Como comparação, a mesma análise foi realizada para fragmentos gerados *in silico* pela digestão genômica usando apenas a enzima *MspI*, que possui o mesmo sítio de restrição da *HpaII*, e os resultados demonstram uma igualmente significativa associação com genes (Figura 11). Como a adição de uma enzima de corte raro na dupla digestão genômica é importante para que a redução no custo de sequenciamento seja significativa, o uso da enzima *PstI* é justificável. Contudo, seria importante avaliar se a utilização de outras enzimas de restrição geraria o

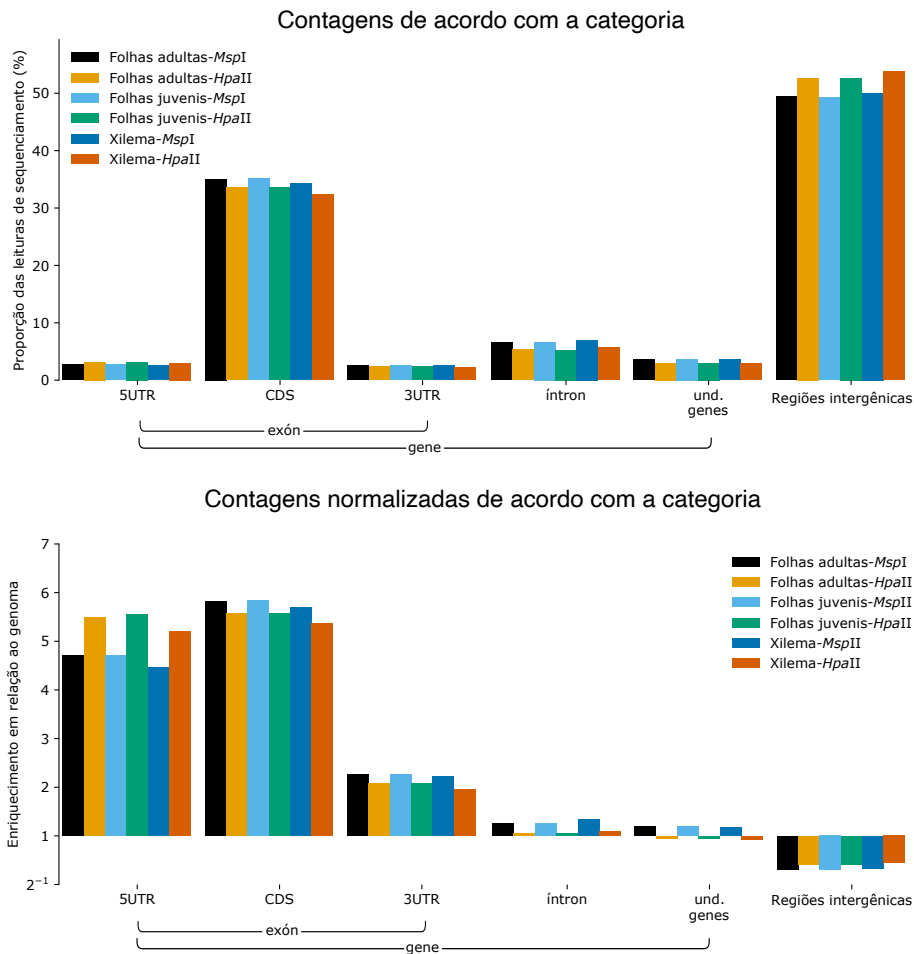


Figura 10 – Distribuição das sequências geradas por MS-DarT-seq em relação aos elementos do genoma de *E. grandis*. Na parte superior, são demonstradas as proporções das sequências mapeadas no genoma de acordo com o contexto genômico. Na parte inferior é demonstrada a abundância de sequências em cada elemento genômico de acordo com o esperado para o genoma. Como as contagens das réplicas de cada amostra são similares, a soma delas é representada nos gráficos para facilitar a visualização.

mesmo resultado.

Por isso, dois conjuntos adicionais de fragmentos foram gerados a partir da dupla digestão *in silico* do genoma com as combinações de enzimas *EcoRI-MspI/HpaII* e *BamHI-MspI/HpaII* e a análise de co-localização foi implementada. Para todos os conjuntos, incluindo o de fragmentos *PstI-MspI/HpaII* sequenciados, foi observada significativa co-localização apenas com genes. Contudo, o uso das enzimas de restrição *EcoRI* ou *BamHI* em substituição à *PstI* diminui a magnitude do efeito da associação com genes, com menos intersecções sendo identificadas (Figura 11).

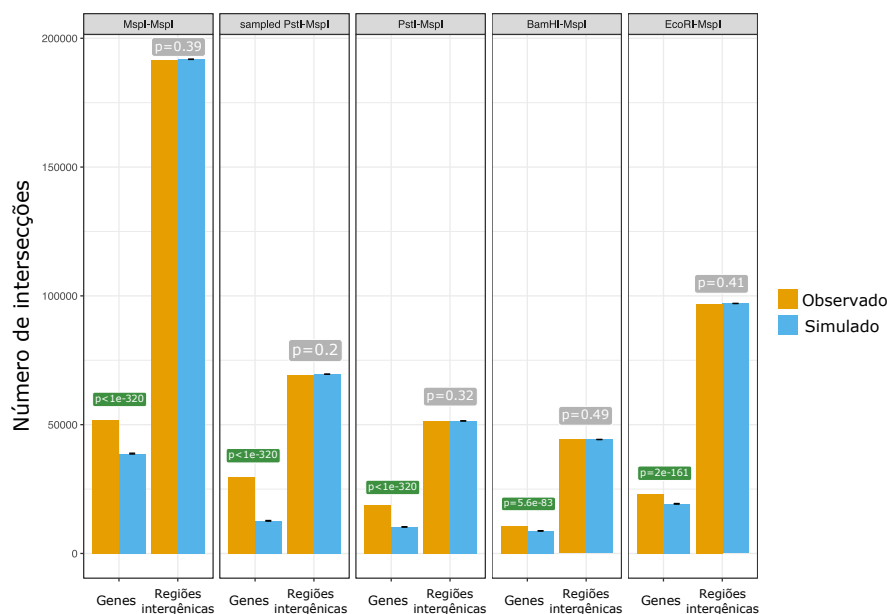


Figura 11 – Análise de co-localização com genes e regiões intergênicas dos conjuntos de fragmentos gerados pela redução genômica *in silico* usando diferentes combinações de enzimas de restrição. Para cada combinação de enzimas, é representado o número de intersecções entre os fragmentos *in silico* e genes ou regiões intergênicas e o p-valor com a significância associada à co-localização. Em amarelo, estão representadas as intersecções obtidas com o conjunto de fragmentos gerados pela digestão *in silico* e em azul, as intersecções obtidas pelas simulações gerando fragmentos obtidos aleatoriamente. Quando não há diferença entre o número de intersecções entre o conjunto de fragmentos aleatórios e as regiões investigadas (genes ou regiões intergênicas) e o número de intersecções entre o conjunto de fragmentos obtidos pela digestão *in silico* e as mesmas regiões, os resultados indicam que a combinação de enzimas utilizadas para a digestão *in silico* produz fragmentos que se comportam como fragmentos gerados aleatoriamente e, portanto, o uso das enzimas não produz co-localização com as regiões examinadas. Se o número de intersecções é maior para os fragmentos obtidos pela digestão *in silico* do que para os fragmentos gerados aleatoriamente, isso significa que o uso da combinação de enzimas produz fragmentos que co-localizam com as regiões investigadas. Barras de erros representam o desvio padrão das medidas obtidas pelas múltiplas simulações.

#### 4.1.3 AMOSTRAGEM DOS SÍTIOS DE METILAÇÃO POR MS-DArT-seq

Utilizando as posições de mapeamento de todas as sequências e de todos os sítios *MspI/HpaII*, foram identificados 76.106 fragmentos sequenciados, representando 72.515 sítios MSD. A discrepância entre o número de fragmentos sequenciados e o número de sítios MSD se deve ao fato de que um mesmo sítio MSD pode ser representado por mais de um fragmento, como descrito na metodologia (seção 3.2.3) e exemplificado na Figura 6.

Os sítios MSD identificados estão amplamente distribuídos ao longo dos 11 cromossomos do genoma de *E. grandis*, conforme representado na Figura 12. Contudo, essa distribuição não pode ser considerada uniforme, conforme averiguado pela aplicação de testes qui-quadrado (*goodness-of-fit*) para verificar a equivalência do número de sítios

MSD em seções geradas pela divisão do genoma em intervalos de múltiplos tamanhos (ver métodos, seção 3.3). Em todos os casos avaliados, a uniformidade da distribuição foi rejeitada ( $p$ -valor  $< 0,001$ ). Na Figura 12 são representadas as contagens por intervalo de 250 kb.

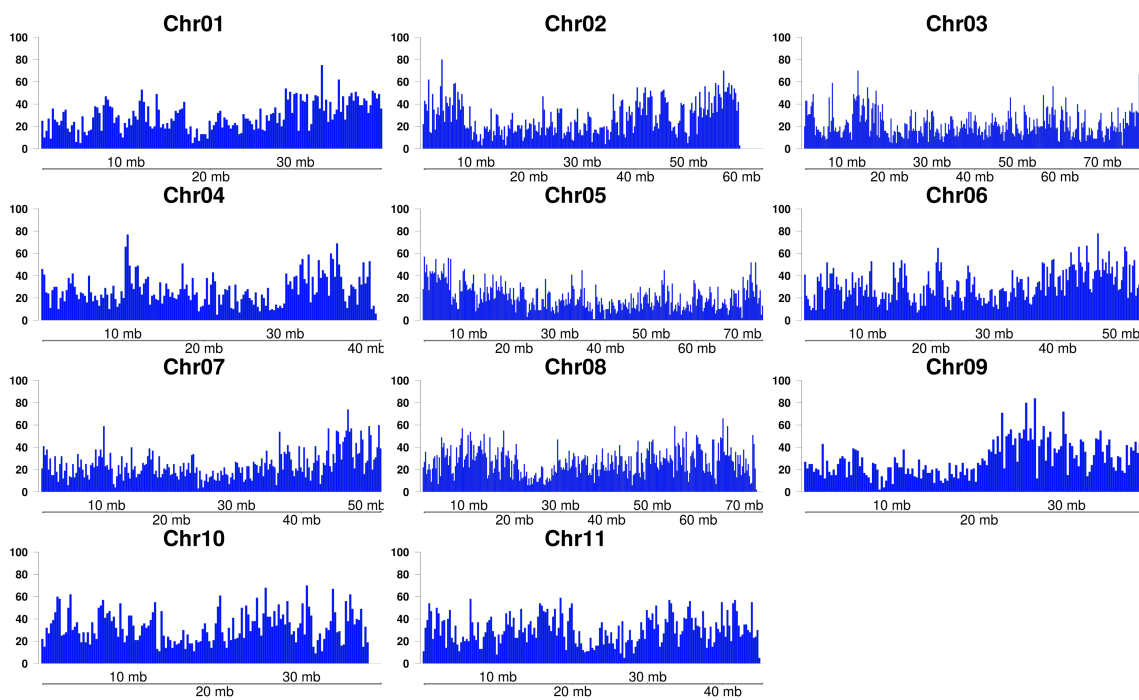


Figura 12 – Distribuição dos sítios MSD ao longo dos 11 cromossomos do genoma de *E. grandis*. Cada barra representa um intervalo de 250 mil bases (kb).

#### 4.1.4 REPRODUTIBILIDADE ENTRE RÉPLICAS BIOLÓGICAS

Após a determinação dos sítios MSD, foi realizada a contagem das sequências correspondentes a cada sítio nas bibliotecas *PstI-MspI* e *PstI-HpaII* de cada amostra, separadamente, para as três réplicas biológicas.

A reprodutibilidade entre as réplicas biológicas foi avaliada a partir da correlação entre as contagens obtidas para sítios MSD em cada amostra. Por esse quesito, foi possível observar que há grande reprodutibilidade entre as réplicas biológicas e mesmo entre as duas amostras de folha (folhas juvenis e adultas) ( $r^2 \geq 0,97$ ). Também foram observadas correlações ( $r^2$ ) acima de 0,52 entre o xilema e as amostras de folha. Desta forma, pode-se afirmar que existe moderada correlação entre as contagens dessas amostras (Figure 13).

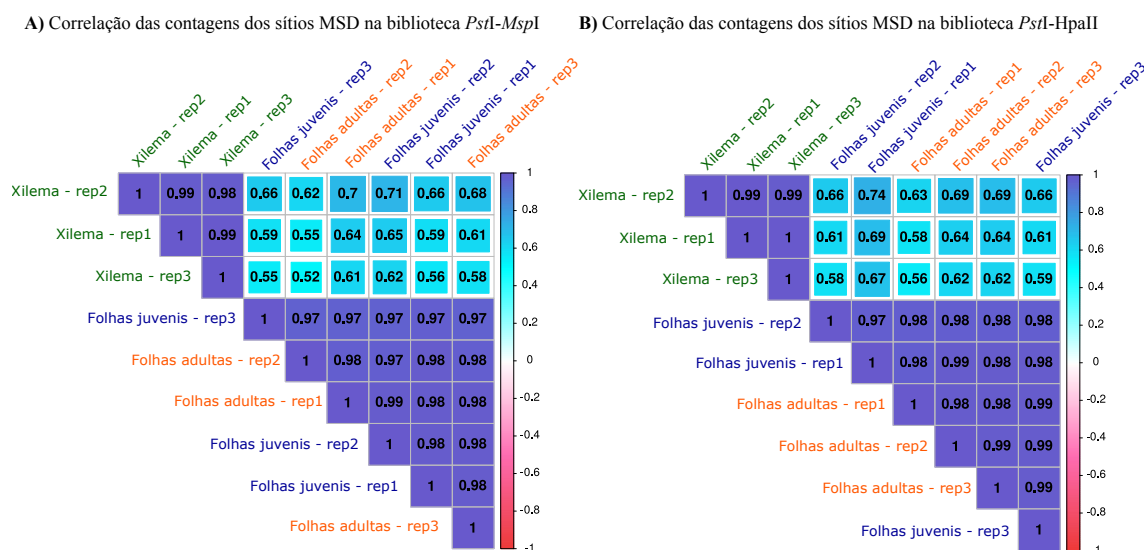


Figura 13 – **Reprodutibilidade de contagens entre as réplicas biológicas.** A) Gráfico de correlações entre as contagens obtidas para as réplicas biológicas de cada amostra nas bibliotecas *PstI-MspI*. B) Gráfico de correlações entre as contagens obtidas para as réplicas biológicas de cada amostra nas bibliotecas *PstI-HpaII*.

## 4.2 METILAÇÕES DETECTADAS NO DNA E CONTEXTO BIOLÓGICO

### 4.2.1 REMOÇÃO DE DADOS FALTANTES PERMITE A COMPARAÇÃO DAS METILAÇÕES DE DNA ENTRE AMOSTRAS

Para possibilitar a comparação entre amostras, é necessário garantir que todos os sítios avaliados para metilação tenham sido sequenciados em todas as amostras. Isto é importante para evitar a ocorrência de falso negativos, determinando a ausência de metilações quando de fato a região correspondente no genoma não foi sequenciada por fatores alheios aos padrões de metilação. Para esse fim, foram considerados apenas sítios MSD amostrados em todas as amostras, como verificado pela presença de contagens na biblioteca *PstI-MspI* que é insensível à metilação no contexto avaliado (contagem  $\geq 1$ ).

Considerando o conjunto formado apenas pelos sítios de metilação amostrados em todas as amostras e após a aplicação do filtro estabelecido para eliminar contagens espúrias (média das contagens na biblioteca gerada por *MspI*  $\geq 3$ ), um total de 30.987 sítios MSD foram identificados.

### 4.2.2 METILAÇÕES DETECTADAS NO DNA

Uma vez estabelecidas contagens para todos os sítios MSD e amostras, foi realizada a verificação se a diferença nas contagens das leituras para cada loco entre as bibliotecas *MspI* e *HpaII* eram estatisticamente significativas. Para tanto, utilizou-se o arcabouço

estatístico adaptado de análise de expressão diferencial de mRNAs pela técnica RNA-seq, conforme implementado nos programas `edgeR` (ROBINSON; MCCARTHY; SMYTH, 2010) e `DEseq2` (LOVE; HUBER; ANDERS, 2014).

Para determinação dos valores dos parâmetros a serem utilizados por `edgeR` e `DEseq2`, múltiplos testes foram realizados usando a amostra de folhas adultas para a qual dados de BS-seq foram também gerados. Em cada teste, as metilações detectadas por MS-DArT-seq foram comparados com os resultados de BS-seq, considerados como o conjunto verdadeiro de metilações do DNA durante a validação da metodologia MS-DArT-seq (ver métodos, seções 3.5 e 3.7). Através dessa comparação, considerou-se que um sítio metilado detectado por MS-DArT-seq foi validado se o mesmo sítio foi também identificado como metilado pela análise de BS-seq. Conforme apresentado na Tabela 1, os parâmetros que propiciaram a maior taxa de validação (95,47% dos sítios metilados detectados por MS-DArT-seq, e que foram também amostrados e considerados como metilados por BS-seq), alinhada com o maior número de metilações identificadas por MS-DArT-seq (4.759), foram o *fold change*  $\geq 2$ , FDR  $\leq 0,05$  com filtro de contagem mínimas igual a 3 sequências.

Tabela 1 – Definição dos parâmetros para determinação de metilação do DNA usando dados de MSD-DArT-seq pela comparação com dados de sequenciamento bissulfito.

FDR (<)	Fold Change ( $\geq$ )	Contagem mínima (MspI $\geq$ )	Número de sítios metilados	Número de sítios metilados também testados por BS-seq	% de sítios metilados testados em BS-seq	N de sítios validados por BS-seq	% de validação
0,001	1	3	4.051	3.679	90,82%	3.186	86,60%
0,001	1	5	4.081	3.707	90,84%	3.211	86,62%
0,001	1	10	3.741	3.407	91,07%	2.905	85,27%
0,001	2	3	3.320	3.023	91,05%	2.934	97,06%
0,001	2	5	3.348	3.049	91,07%	2.958	97,02%
0,001	2	10	2.993	2.734	91,35%	2.646	96,78%
0,01	1	3	5.064	4.552	89,89%	3.898	85,63%
0,01	1	5	4.955	4.467	90,15%	3.814	85,38%
0,01	1	10	4.142	3.745	90,42%	3.090	82,51%
0,01	2	3	4.086	3.688	90,26%	3.561	96,56%
0,01	2	5	3.974	3.600	90,59%	3.476	96,56%
0,01	2	10	3.137	2.855	91,01%	2.746	96,18%
0,05	1	3	6.042	5.409	89,52%	4.534	83,82%
0,05	1	5	5.509	4.943	89,73%	4.077	82,48%
0,05	1	10	4.463	4.024	90,16%	3.207	79,70%
0,05	2	3	4.759	4.280	89,93%	4.086	95,47%
0,05	2	5	4.221	3.809	90,24%	3.629	95,27%
0,05	2	10	3.160	2.875	90,98%	2.757	95,90%

\*Em azul estão destacados os parâmetros que maximizam o número de metilações de DNA identificadas e a taxa de validação com dados de BS-seq.

É importante ressaltar que embora a taxa de confirmação dos sítios metilados seja elevada, os sítios investigados por MS-DArT-seq representam uma pequena fração dos sítios investigados por BS-seq. Para folhas adultas, mais de 29 milhões de citosinas no contexto CG foram investigadas por BS-seq, enquanto 145.030, considerando-se que cada sítios MSD corresponde à duas citosinas CG, uma em cada fita da molécula de DNA, citosinas foram investigadas por MS-DArT-seq. Dessa forma, neste estudo, MS-DArT-seq amostrou 0.49% das citosinas amostradas por BS-seq. A partir do conjunto de citosinas amostradas por ambas as técnicas e pertencentes ao conjunto de 30.987 citosinas presentes

em todos os tecidos (26.164), e considerando-se que os resultados de BS-seq refletem o estado de metilação verdadeiro da amostra, determinou-se que MS-DArT-seq possui sensibilidade de 60.81% e especificidade de 99.01%.

Em cada amostra, mais de 93% das sequências com contagens estatisticamente diferentes entre as bibliotecas *MspI* e *HpaII* foram confirmadas pelos dois algoritmos utilizados na análise, (edgeR e DEseq2). Ao todo, 4.759 metilações foram detectadas para a amostra de folhas adultas, 5.026 para folhas juvenis e 4.634 para xilema. Como representado na Figura 14, há uma grande similaridade entre os perfis de metilação das três amostras avaliadas.

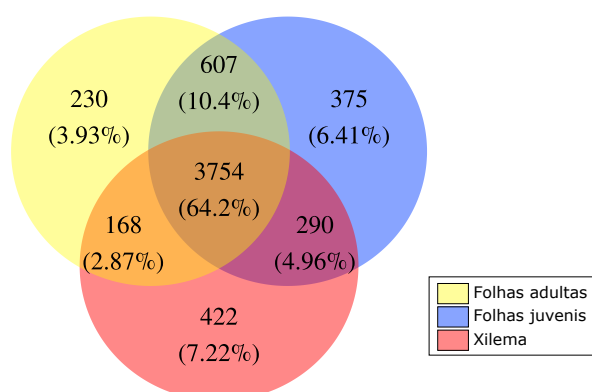


Figura 14 – Diagrama de Venn comparando os sítios metilados entre as três amostras.

Considerando-se em conjunto todas as metilações detectadas a partir da combinação da metodologia MS-DArT-seq e do protocolo de análise implementado, foram identificadas 5.846 citosinas metiladas no genoma de *E. grandis*.

### 4.2.3 COMPARAÇÃO DAS METILAÇÕES ENTRE AMOSTRAS

A partir da comparação das metilações de DNA detectadas, foi possível observar que existe grande similaridade entre os perfis de metilação de cada amostra, com aproximadamente 64,2% das metilações detectadas presentes nas três amostras amostradas. Como esperado, há um maior número de metilações comuns às duas amostras de folhas (adultas e juvenis) do que entre estas e o xilema (Figura 14). A maior diferenciação entre os perfis de metilação das folhas e do xilema pode também ser verificada a partir do resultado da análise de componentes principais (PCA, *Principal Component Analysis*) onde, utilizando as contagens obtidas para cada sítio metilado, é possível observar claramente a diferença entre as amostras do xilema e as duas amostras de folhas (Figura 15).

Apesar da grande similaridade entre os perfis de metilação das amostras avaliadas, é possível constatar a ocorrência de 1.027 metilações de DNA amostra-específicas. Para cada amostra, ao menos 230 metilações exclusivas foram verificadas (230 em folhas adultas,

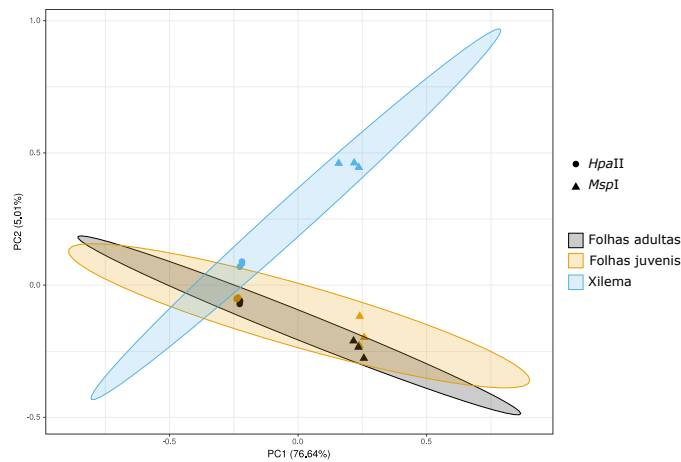


Figura 15 – Análise de componentes principais demonstrando a diferença entre as amostras e de acordo com a biblioteca de sequenciamento implementada. As PCAs foram realizadas a partir das contagens das sequências de cada sítio de metilação avaliado pela técnica MS-DaT-seq e considerados como metilados.

375 em folhas juvenis e 422 no xilema; Figura 14). Conforme observado pelo teste Q de Cochran, a probabilidade de um sítio MSD ser metilado é dependente da amostra avaliada ( $p$ -valor  $< 0,001$ ), com cada amostra sendo significativamente diferente dos demais (confirmado por comparações par-a-par usando o teste de Wilcoxon ( $p$ -valor  $< 0,001$ ).

#### 4.2.4 CONTEXTO GENÔMICO DAS METILAÇÕES IDENTIFICADAS

A caracterização do contexto genômico das metilações de DNA detectadas revelou que a maioria destas estão localizadas em genes, principalmente em éxons. TEs contém apenas 16,97% dos sítios metilados. Um conjunto total de 1.121 (19,2%) metilações estão em regiões sem anotação (Tabela 2). A distribuição das metilações das amostras em relação ao contexto genômico se assemelha à distribuição verificada para o conjunto total de metilações, como apresentado na Tabela 2.

Para uma pequena parcela das metilações identificadas, não foi possível estabelecer o contexto genômico devido à sobreposição das anotações de genes e TEs. Isso deve-se a possíveis problemas na anotação do genoma de *E. grandis*. Além disso, existe a possibilidade de inserção de TEs dentro de transposons e genes previamente existentes no genoma. Como não é possível diferenciar esses cenários, estes foram nomeados como “genes aninhados” e “TEs aninhados” todos os casos onde há sobreposição da anotação de um gene e de TE, respectivamente. Ao todo, 307 metilações ( $\sim 5,20\%$  do total) foram identificadas nestas regiões (Tabela 2).

Para determinar se a maior frequência de metilações em regiões gênicas se deve a vieses intrínsecos da tecnologia, como indicado pela prevalência de sequências (Figura 10) e fragmentos *in silico* (Figura 11) em regiões gênicas, o teste exato de Fisher foi



Tabela 2 – Distribuição das metilações de DNA verificadas em cada amostra de acordo com o contexto genômico.

Categoria	Subcategoria	Folha adultas	Folhas juvenis	Xilema	Total
Genes	Íntron ou UTRs	803	826	817	923
	Éxon	2.108	2.229	2.128	2.503
	<b>Total</b>	2.911	3.055	2.945	3.426
TEs	Íntron ou UTRs	88	91	88	106
	Regiões intergênicas	689	737	565	886
	<b>Total</b>	777	828	653	992
Elementos aninhados	Genes aninhados	23	21	21	26
	TEs aninhados	209	232	239	281
	<b>Total</b>	232	253	260	307
Regiões intergênicas	Regiões intergênicas	839	890	776	1.121
<b>Total</b>		4.759	5.026	4.634	5.846

\* Genes e TEs aninhados são elementos cuja anotação se sobrepõe à anotação de outros elementos genômicos. Por exemplo, um TE que foi inserido na sequência de outro TE.

aplicado. Inicialmente, constatou-se que os sítios MSD (amostrados por MS-DArT-seq) são significativamente mais frequentes ( $p$ -valor  $< 0,001$ ) em regiões gênicas do que o esperado, dada a distribuição de todos os sítios de restrição das enzimas *MspI/HpaII* no genoma (Tabela 3).

Tabela 3 – Tabela de contingência demonstrando a amostragem de sítios MSD no genoma de acordo com o contexto genômico.

Amostragem de sítios <i>MspI/HpaII</i> em genes e regiões intergênicas			
	Sítios MSD amostrados	Sítios <i>MspI/HpaII</i> não amostrados	Total
Genes	29.552	165.890	195.442
Regiões intergênicas	45.963	741.481	787.444

Em seguida, avaliou-se se as metilações do DNA ocorrem mais frequentemente nos sítios MSD amostrados em todas as amostras (31.120) localizados em genes. Foi observado que as metilações de DNA detectadas ocorreram mais frequentemente em sítios localizados em genes ( $p$ -valor  $< 0,001$ ; Tabela 4), mesmo considerando-se que o total de sítios MSD em genes e regiões intergênicas é similar, 15.010 e 16.110, respectivamente.

Tabela 4 – Comparação das ocorrências de metilações nos sítios MSD em regiões gênicas e intergênicas.

Ocorrência de metilações do DNA nos sítios MSD em genes e regiões intergênicas			
	Sítios Metilados	Sítios não metilados	Total
Genes	3 532*	11.478	15.010
Regiões intergênicas	2.007	14.103	16.110

\* Metilações em TEs dentro de íntrons ou UTRs de um genes foram considerados como sítios metilados em genes.

A contextualização das demais 1.121 metilações de DNA verificadas demonstrou que, mesmo estando localizadas fora de genes e TEs, a grande maioria dessas encontram-se próximas a um desses elementos genômicos (509 sítios localizados a menos de 10 kb de um gene e 1.037, a menos de 10 kb de um TE) podendo, portanto, interferir na regulação dos mesmos (Figura 16). A distribuição desses sítios MSD foi comparada com a distribuição de todos os sítios *PstI/MspI* do genoma de *E. grandis* no mesmo intervalo de 10 kb de genes e TEs. O resultado do teste qui-quadrado demonstra que a concentração de sítios MSD é significativamente maior do que o esperado (p-valor  $\sim 0,006$ ), com a maior diferença sendo observada para intervalos de 1 kb (Figura 17).

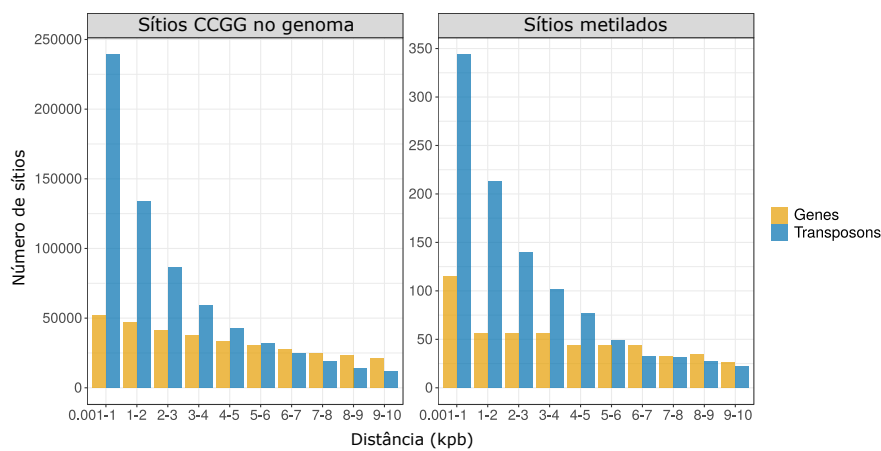


Figura 16 – **Distância das metilações de DNA para o gene ou TE mais próximo.** No gráfico da esquerda estão representadas as distâncias entre todos os sítios CCGG, reconhecidos pelas enzimas de restrição *MspI* e *HpaII*, do genoma para o gene ou TE mais próximo. À direita, as distâncias entre as metilações de DNA identificadas em sítios MSD para o gene ou TE mais próximo. Em ambos os gráficos são apresentados apenas sítios contidos em uma distância máxima de 10 mil pares de bases do elemento genômico. Cada barra representa o intervalo de mil pares de bases (1 kb).

#### 4.2.5 METILAÇÕES DO DNA EM ELEMENTOS TRANSPONÍVEIS

Das 992 metilações localizadas em TEs, aproximadamente 50% foram verificadas nas três amostras. Sítios metilados específicos de cada amostra também foram verificados (53 em folhas adultas, 93 em folhas juvenis e 76 no xilema). Aqui também, conforme o esperado, foi observada uma maior correspondência entre as metilações verificadas para as amostras de folha do que entre elas e o xilema (Figura 18-A).

Considerando-se apenas os elementos transponíveis que contêm, ao menos, uma metilação do DNA (TEs metilados), há uma pequena redução na correspondência entre as diferentes amostras. Aproximadamente, 48,2% (406) dos TEs metilados foram identificados em todas as amostras. Novamente, maior similaridade entre as amostras de folha do que destas com o xilema foi observada (Figura 18-B). Ao todo, 842 TEs do genoma de *E.*

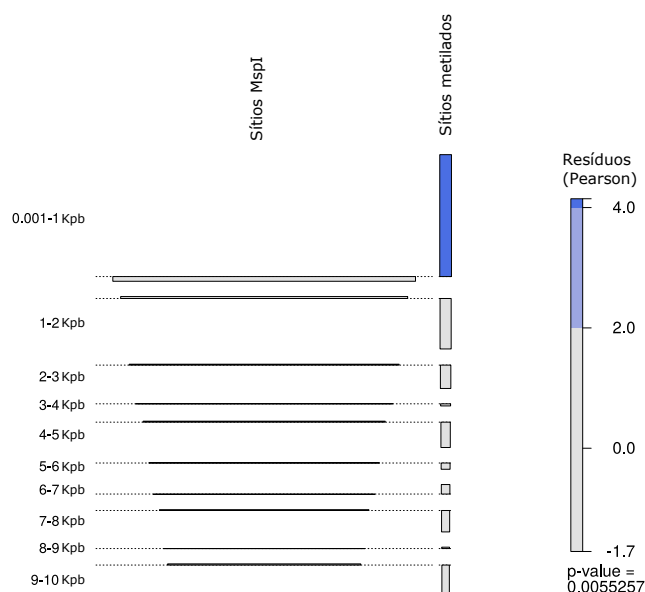


Figura 17 – Gráfico de associação demonstrando os resíduos (Pearson) do teste qui-quadrado aplicado para a distribuição dos sítios metilados na proximidade de genes (10 kb). O p-valor demonstrado na parte inferior do gráfico foi gerado pelo teste qui-quadrado aplicado em uma tabela de contingência de múltiplos fatores, rejeitando a hipótese nula de independência entre as amostras quando p-valor < 0,05. Em um gráfico de associação<sup>1</sup>, a área de cada quadrado equivale à diferença entre as frequências esperadas e observadas e as cores indicam as células individuais que são, provavelmente, significativas, ou seja, que apresentem grande desvio das frequências esperadas.

*grandis* contém ao menos uma citosina metilada detectada pela metodologia implementada neste estudo.

A) Sítios metilados localizados em TEs B) TEs contendo sítios metilados

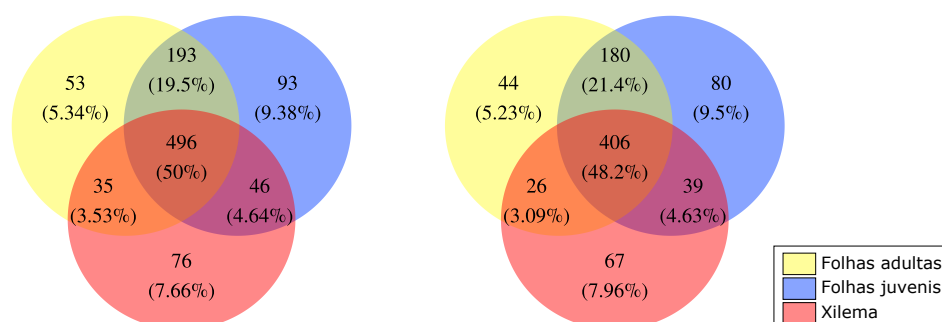


Figura 18 – Diagrama de Venn comparando os elementos transponíveis metilados em cada amostra.

Em relação à classificação dos elementos transponíveis que apresentam metilação, é possível observar que o perfil de distribuição desses se assemelha à distribuição de TEs no genoma, de maneira geral, sendo os TEs LARD e LTR os mais abundantes

<sup>1</sup> *Visualizing Independence Using Extended Association and Mosaic Plots.*

(Figura 19). O teste de qui-quadrado foi aplicado para verificar se a distribuição dos TEs metilados nas diferentes classes é dependente da amostra. Para esse teste, foram utilizadas as proporções de TEs metilados em cada classe. Os resultados demonstram que não há diferença significativa entre essas classes ( $p$ -valor = 0,99), indicando que a distribuição é independente da amostra avaliada (Figura 19).

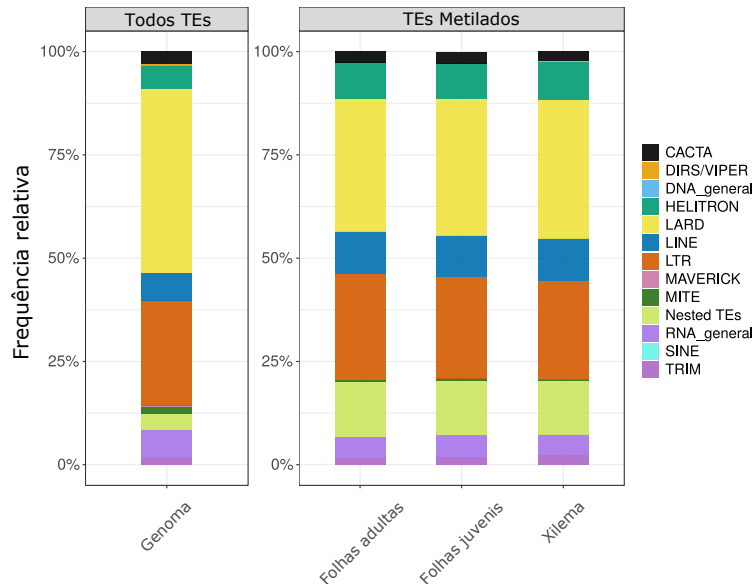


Figura 19 – **Classificação dos elementos transponíveis que contêm ao menos uma metilação do DNA.** Do lado esquerdo, está representada a distribuição dos TEs presentes no genoma de acordo com a sua classificação. Do lado direito, está representada a proporção dos TEs que contêm, ao menos, uma metilação do DNA para cada amostra, considerando-se as classes de TEs avaliadas.

Considerando todos os TEs do genoma de *E. grandis* e assumindo que estes possuem igual probabilidade de ser metilados, observou-se significativa diferença entre as proporções esperadas e observadas nas amostras avaliadas, indicando que algumas classes de TEs não são constitutivamente metiladas como o esperado ( $p$ -valor < 0,001 no teste de qui-quadrado). Pela visualização dos resíduos em um gráfico de associação (Figura 20), é possível demonstrar que a frequência de TEs metilados nas classes LINE e Helitron, bem como o número TEs classificados como aninhados, é maior do que o esperado para as três amostras avaliadas. Além disso, todas as amostras possuem menos TEs metilados nas classes LARD e MITE do que o esperado (Figura 20).

#### 4.2.6 METILAÇÕES DO DNA EM GENES CODIFICADORES DE PROTEÍNAS

Ao todo, 3.426 metilações foram identificadas em genes pela metodologia implementada neste estudo. Destas, a grande maioria é comum às três amostras avaliadas ( $\sim 73\%$ ; Figura 21-A). Considerando-se os genes que contêm ao menos uma metilação de DNA

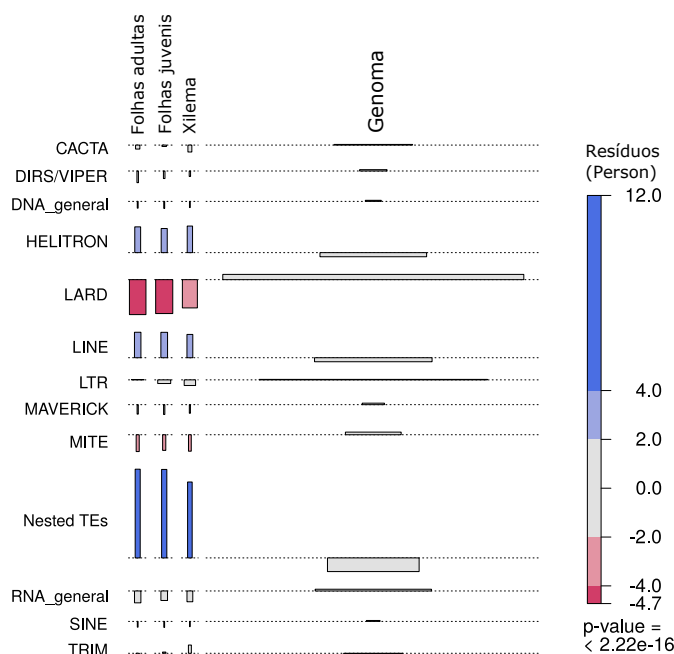


Figura 20 – Gráfico de associação demonstrando os resíduos (Pearson) do teste qui-quadrado aplicado para a distribuição de TEs metilados em classes. O p-valor demonstrado na parte inferior do gráfico foi gerado pelo teste qui-quadrado aplicado em uma tabela de contingência de múltiplos fatores, rejeitando a hipótese nula de independência entre as amostras quando  $p\text{-valor} < 0,05$ . Em um gráfico de associação<sup>2</sup>, a área de cada quadrado equivale à diferença entre as frequências esperadas e observadas e as cores indicam as células individuais que são provavelmente significativas, ou seja, que apresentem grande desvio das frequências esperadas. Em azul, são destacadas células cuja frequência observada é maior do que a frequência esperada e em vermelho, as células cuja frequência observada é menor do que a esperada.

detectada, em pelo menos uma das amostras, foram observados 2.783 genes codificadores de proteína metilados. Desse total, a maior parte ( $\sim 75,4\%$ ; 2.097 genes) possui metilações em todas as amostras avaliadas, embora não necessariamente na mesma citosina (Figure 21-B). Foram identificados 325 genes que possuem metilações detectáveis pela técnica exclusivamente em uma das amostras, sendo 74 genes exclusivamente metilados em folhas adultas, 119 em folhas juvenis e 132 no xilema (Figure 21-B).

#### 4.2.7 ANOTAÇÃO FUNCIONAL E ENRIQUECIMENTO DE TERMOS DO GENE ONTOLOGY PARA OS GENES METILADOS

Uma das informações relevantes para a interpretação biológica das metilações presentes em citosinas de genes é a classificação funcional desses. Especificamente, existe grande interesse em determinar se, no conjunto de genes metilados, existe um enriquecimento para alguma função biológica. Visando a obtenção dessa informação, foi realizada a

<sup>2</sup> *Visualizing Independence Using Extended Association and Mosaic Plots.*

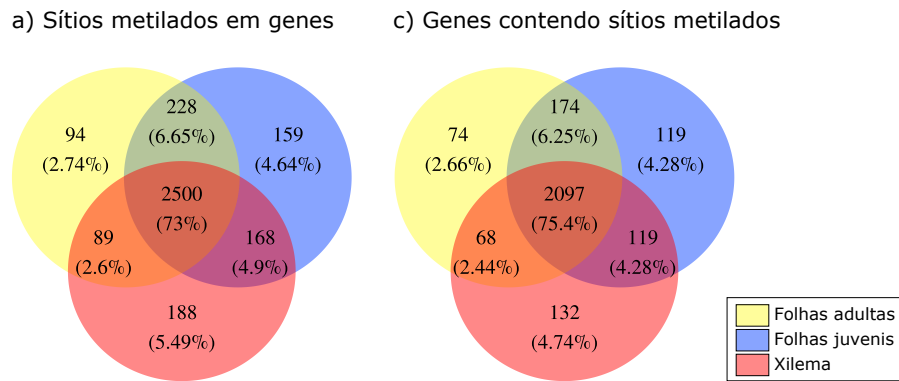


Figura 21 – Diagrama de Venn comparando os genes metilados em cada amostra.

anotação e a análise de enriquecimento de termos para os 2.783 genes metilados, cujos resultados estão descritos abaixo.

Foram verificados termos enriquecidos para os conjuntos de genes metilados referentes a cada uma das três amostras e nas três categorias do GO (BP - *Biological Process*, MF - *Molecular Function* e CC - *Cellular Component*). Os resultados demonstram que existem 149 GO termos enriquecidos no conjunto de genes metilados verificados neste estudo, quando comparados com o conjunto de genes que possuem ao menos um sítio MSD. Aproximadamente, 50% dos termos enriquecidos estão presentes em todas as amostras. Contudo, existem termos enriquecidos que são amostra-específicos.

Considerando-se os genes metilados, foram identificados 10 termos enriquecidos específicos para folhas adultas, 23 para folhas juvenis e 8 para o xilema (Tabela 5). A grande maioria dos termos enriquecidos representam processos biológicos essenciais à sobrevivência da planta. Já o conjunto de termos enriquecidos específicos para cada amostras possivelmente reflete processos específicos dos tipos celulares e fase do desenvolvimento de cada amostra. Por exemplo, termos relacionados com ciclo celular, desenvolvimento e biossíntese (ver Tabela 5), foram especificamente enriquecidos na amostra de folha em desenvolvimento.

Desta forma, as diferenças observadas entre os perfis de metilação são provavelmente relacionadas com a diferenciação celular. Contudo, vale ressaltar que os resultados aqui gerados demonstram a capacidade desta abordagem, baseada em MS-DArT-seq, para a identificação de metilações do DNA que ocorrem especificamente em uma amostra. Assim como em relacioná-las com processos biológicos que possam auxiliar na interpretação do efeito dessas metilações.

Tabela 5 – Lista de termos GO enriquecidos para o conjunto de genes metilados de cada amostra.

Termos enriquecidos exclusivamente na amostra de folhas adultas		
ID	Descrição	Classe
GO:0046488	phosphatidylinositol metabolic process	BP
GO:0048585	negative regulation of response to stimulus	BP
GO:0051188	cofactor biosynthetic process	BP
GO:0006457	protein folding	BP
GO:0072521	purine-containing compound metabolic process	BP
GO:0051049	regulation of transport	BP
GO:0007131	reciprocal meiotic recombination	BP
GO:0051241	negative regulation of multicellular organismal process	BP
GO:0031668	cellular response to extracellular stimulus	BP
GO:0019013	viral nucleocapsid	CC
Termos enriquecidos exclusivamente na amostra de folhas juvenis		
ID	Descrição	Classe
GO:0022402	cell cycle process	BP
GO:0007275	multicellular organism development	BP
GO:0010033	response to organic substance	BP
GO:0009648	photoperiodism	BP
GO:0048869	cellular developmental process	BP
GO:0071310	cellular response to organic substance	BP
GO:0051254	positive regulation of RNA metabolic process	BP
GO:0031325	positive regulation of cellular metabolic process	BP
GO:0044270	cellular nitrogen compound catabolic process	BP
GO:0042398	cellular modified amino acid biosynthetic process	BP
GO:0042559	pteridine-containing compound biosynthetic process	BP
GO:0071902	positive regulation of protein serine/threonine kinase activity	BP
GO:0001101	response to acid chemical	BP
GO:0017056	structural constituent of nuclear pore	MF
GO:0016881	acid-amino acid ligase activity	MF
GO:0031410	cytoplasmic vesicle	CC
GO:1990234	transferase complex	CC
GO:1990904	ribonucleoprotein complex	CC
GO:0000428	DNA-directed RNA polymerase complex	CC
GO:0030880	RNA polymerase complex	CC
GO:0099513	polymeric cytoskeletal fiber	CC
GO:0099080	supramolecular complex	CC
GO:0099081	supramolecular polymer	CC
Termos enriquecidos exclusivamente na amostra do xilema		
ID	Descrição	Classe
GO:0071705	nitrogen compound transport	BP
GO:0048573	photoperiodism, flowering	BP
GO:0009787	regulation of abscisic acid-activated signaling pathway	BP
GO:0009755	hormone-mediated signaling pathway	BP
GO:0006413	translational initiation	BP
GO:0000045	autophagosome assembly	BP
GO:0097306	cellular response to alcohol	BP
GO:0019200	carbohydrate kinase activity	MF

Termos e notações seguem a descrição do *Gene Ontology Consortium*. BP (*biological process*): Processo biológico, MF (*molecular function*): Função molecular, CC (*cellular component*): Componente celular.

# 5 DISCUSSÃO

Neste estudo, foi desenvolvida uma nova abordagem analítica para detecção de metilação de DNA ( $5^mC$ ) utilizando a técnica de sequenciamento MS-DArT-seq. Esta técnica se enquadra no conjunto de métodos baseados na redução genômica usando enzimas de restrição sensíveis à metilação, como por exemplo MSCC (BALL et al., 2009) e MRE-seq (MAUNAKEA et al., 2010). Contudo, a abordagem descrita diverge dessas técnicas, sendo baseada na dupla digestão genômica com a enzima *PstI* em adição à *MspI* (insensível à metilação) e *HpaII* (sensível à metilação), gerando duas bibliotecas de fragmentos sequenciados (*PstI-MspI* e *PstI-HpaII*). Os isosquisômeros *MspI* e *HpaII* reconhecem o sítio de restrição CCGG, sujeito a metilações nos contextos CG e CHG. Como ambas as enzimas de restrição são bloqueadas pela metilação no contexto CHG (FULNEČEK; KOVAŘÍK, 2014; POOTAKHAM et al., 2016), a metodologia investiga a ocorrência de metilações no contexto CG.

Destaca-se a escolha da enzima *PstI* por ela apresentar sensibilidade à metilação no contexto CHG. Desta forma, seu uso reduz a seleção de fragmentos em regiões hipermetiladas do genoma, como as regiões ricas em elementos repetitivos, promovendo, portanto, a amostragem preferencial de regiões hipometiladas como regiões ativas da cromatina.

Visando a validação do método, a abordagem desenvolvida foi aplicada para três amostras de um mesmo genótipo de *E. grandis*, propiciando a detecção de milhares de metilações do DNA.

## 5.1 IMPORTÂNCIA DA UTILIZAÇÃO DE *E. grandis* COMO MODELO EXPERIMENTAL

Amostras de clones da árvore BRASUZ1, cujo genoma foi sequenciado para a obtenção do genoma de referência da espécie *Eucalyptus grandis* (MYBURG et al., 2014), foram utilizadas como modelo para a implementação da análise descrita neste trabalho. A utilização de uma espécie cuja propagação é realizada por clonagem apresenta a intrínseca vantagem da disponibilidade de réplicas biológicas, ou seja, indivíduos livres de variação genética. Dessa forma, representando uma ótima oportunidade para o estabelecimento de um modelo para investigação dos processos epigenéticos em plantas, possibilitando a remoção do componente genético para avaliação exclusivamente da variação epigenética amostrada na forma de metilação em citosinas. Essa é uma característica importante, uma vez que diversos estudos têm demonstrado uma grande relação das variações epigenéticas com variações genéticas, particularmente, nos padrões de metilações de DNA, influenci-



ando diretamente a ocorrência de epialelos (GREALLY, 2017; SEYMOUR; BECKER, 2017). Este estudo foi realizado com um único clone de *E. grandis* como piloto para o estabelecimento da metodologia analítica descrita e abre possibilidades para aplicação em delineamentos experimentais mais complexos.

## 5.2 CARACTERÍSTICAS DOS DADOS GERADOS POR MS-DArT-seq

### 5.2.1 QUALIDADE DAS SEQUÊNCIAS GERADAS POR MS-DArT-seq

Diversas etapas de filtragem e controle de qualidade foram implementadas visando a obtenção de resultados robustos e reproduzíveis. Cerca de 30% (13.151.774) das sequências geradas não foram utilizadas no estudo porque não mapearam no genoma de referência (5.282.139) ou mapearam em múltiplas posições (7.869.635). Este fato pode ser explicado pelos filtros restritivos aplicados, como a exigência de mapeamento da leitura de sequenciamento completa, impedimento da abertura de *gaps* e da inclusão de leituras de sequência com mais de três *mismatches*. Adicionalmente, alta heteroziguidade apresentada nos clones de *E. grandis* e a grande fração do genoma representada por elementos repetitivos (MYBURG et al., 2014) também contribuíram para estes resultados.

Após a definição dos sítios MSD e durante a etapa de contagem das leituras mapeadas, verificou-se que uma pequena parcela das sequências mapeadas apresentaram sobreposição entre si. Essas leituras foram originadas por ausência de clivagem de sítios de restrição em um dos dois cromossomos homólogos. A ausência de clivagem de sítios de restrição pode ser resultado de três causas principais. Pode haver metilação na citosina externa em apenas um dos dois cromossomos levando ao bloqueio da enzima *MspI* nesse sítio enquanto, no cromossomo homólogo, o corte é realizado e um fragmento é gerado. Dessa forma, dois fragmentos de tamanhos distintos e oriundos da mesma região genômica são gerados, um para cada cromossomo homólogo. Outra possibilidade é que os sítios de ambos os cromossomos estejam acessíveis (livres de metilação) mas não sejam clivados devido à falta de eficiência enzimática nas condições experimentais implementadas para a construção das bibliotecas de sequenciamento. Ainda é possível que os cromossomos estejam em heterozigose na região de um sítio de restrição, levando à perda do sítio de restrição e, conseqüentemente, a não clivagem do DNA nesta região. Nos três casos, são formados fragmentos distintos correspondentes a uma mesma região genômica que são visualizados como uma série de leituras mapeadas e sobrepostas, conforme demonstrado na Figura 8.

A capacidade da metodologia em identificar corretamente metilações do DNA foi confirmada pela comparação com dados de BS-seq, que mostrou uma correspondência de

mais de 95% entre os sítios metilados identificados pelas duas metodologias (Tabela 1). Dessa forma, MS-DArT-seq pode ser considerada como uma alternativa acurada para a detecção de metilações do DNA no contexto CG em escala genômica.

Visando viabilizar a correta comparação entre amostras, optou-se pela remoção dos sítios MSD com dados faltantes. É relevante ressaltar que, em um estudo onde o objetivo principal não seja a comparação entre amostras mas a detecção e quantificação do maior número possível de metilações do DNA, os sítios que foram excluídos poderiam ser utilizados.

## 5.2.2 AMOSTRAGEM DOS SÍTIOS DE METILAÇÃO EM RELAÇÃO AO CONTEXTO GENÔMICO

A utilização da metodologia de sequenciamento MS-DArT-seq mostrou-se eficaz para a amostragem de dezenas de milhares de sítios de metilação, com mais de 30 mil sítios avaliados no genoma de *E. grandis*. Além disso, a abordagem mostrou-se bastante reprodutível, com uma forte correlação entre as réplicas biológicas tanto para a amostragem de sítios MSD como para a identificação das metilações do DNA (Figura 13).

Interessantemente, foi observado um grande direcionamento da detecção das metilações do DNA em regiões gênicas, a despeito da maior parte do genoma ser constituída de elementos repetitivos e regiões intergênicas (MYBURG et al., 2014; BARTHOLOMÉ et al., 2015). Esse fato deve-se, provavelmente, à escolha das enzimas de restrição utilizadas. Como mencionado anteriormente, ambas as enzimas de restrição *MspI* e *HpaII* são bloqueadas pela presença de metilações do DNA no contexto CHG (FULNEŽEK; KOVAŘÍK, 2014; POOTAKHAM et al., 2016), comuns em plantas, especialmente em regiões intergênicas (NIEDERHUTH et al., 2016). Adicionalmente, a enzima *PstI* também é sensível à metilação CHG (POOTAKHAM et al., 2016; MCCLELLAND; NELSON; RASCHKE, 1994), potencializando esse efeito. Uma potencial desvantagem na utilização da enzima *PstI* é a redução no número de sítios amostrados. Conforme demonstrado na Figura 9, apenas uma fração dos sítios preditos *in silico* foram, de fato, sequenciados, provavelmente, devido à baixa amostragem em regiões ricas em metilações no contexto CHG. Além disso, a exclusão de sequências com múltiplos mapeamentos pode também ter contribuído para a diminuição do número de fragmentos em regiões ricas em elementos repetitivos. Essas regiões afetam a capacidade dos *softwares* em promoverem o correto mapeamento das leituras curtas de sequenciamento, uma vez que estas são compatíveis com múltiplas regiões do genoma, e, assim, são mapeadas em múltiplas posições sendo removidas deste estudo.

Também foi demonstrado, neste estudo, que a utilização das enzimas *BamHI* e *EcoRI*, em substituição à enzima *PstI*, promove uma redução na co-localização entre os

fragmentos gerados e os genes anotados para o genoma, reforçando a hipótese da influência da enzima *PstI* no direcionamento da amostragem para regiões gênicas. Contudo, como também demonstrado pela digestão *in silico* e análise de co-localização com o programa Ologram (Figura 11), fragmentos gerados pela digestão usando apenas a enzima *MspI* são também co-localizados com genes, dessa forma, indicando um potencial viés na distribuição de sítios *MspI* no genoma de *E. grandis*, com um enriquecimento em regiões gênicas.

A amostragem preferencial de regiões gênicas está de acordo com o relatado por Pootakham et al. (2016) em estudo de GBS (*Genotyping By Sequencing*) com a espécie *Elaeis guineensis*, onde houve uma grande concentração de locos amostrados em regiões gênicas quando foi utilizada a combinação das enzimas de restrição *PstI* e *MspI* para o preparo das bibliotecas de sequenciamento. Também em um estudo visando a identificação de metilações de DNA em sítios de restrição CCGG, a partir da digestão com a enzima de restrição *HpaII*, Chwialkowska et al. (2017) verificaram o enriquecimento da amostragem para regiões gênicas. Os autores atribuíram esse fato à maior presença desses sítios de restrição em regiões gênicas, corroborando um possível enriquecimento de sítios CCGG em genes.

Durante a identificação do contexto genômico das metilações do DNA, foram verificados alguns possíveis problemas na anotação de genes e TEs no genoma de *E. grandis*. Por exemplo, foi identificado um grande número de predições de TEs em posições onde estão anotados éxons de genes. Foram verificadas também sobreposições entre genes anotados nas distintas fitas (*strands*) representadas no genoma. Esses achados chamam a atenção para o fato de que, embora o genoma de *E. grandis* tenha sido gerado com sequenciamento Sanger e, portanto, possua alta qualidade, é necessária uma curadoria para resolução desses problemas de anotação.

### 5.2.3 INFLUÊNCIA DE METILAÇÕES EM CORPOS GÊNICOS NOS PADRÕES DE METILAÇÕES VERIFICADOS POR MS-DART-SEQ

Ao todo, considerando-se as três amostras avaliadas, foram identificados 5.846 sítios metilados. A maior parte dessas metilações (~58,6%) foi verificada em sítios de metilação presentes em genes. A ocorrência de metilações do DNA, detectáveis pela metodologia MS-DArT-seq, é mais provável em genes do que em regiões intergênicas mesmo desconsiderando-se o viés na amostragem de sítios MSD (Tabela 4). A presença de metilações em corpos gênicos (*gene body methylation*; GBM) em uma significativa parcela dos genes de plantas pode ser a causa. Nesses genes, a grande maioria das citosinas no contexto de sequência “CG” encontra-se metilada. Uma vez que esse é também o contexto de metilação amostrado pela técnica MS-DArT-seq, é esperado que boa parte dos sítios de metilação amostrados em genes contendo GBM estejam, de fato, metilados. Consequentemente, espera-se que haja maior prevalência de metilações em genes do que

em regiões intergênicas.

No genoma de *E. grandis*, conforme relatado por [Niederhuth et al. \(2016\)](#), existem 7.731 genes que apresentam GBM, aproximadamente 21% dos genes do genoma (36.349). Usando a abordagem descrita neste estudo, 45,48% dos genes com GBM possuem ao menos um sítio MSD avaliado. Desta forma, MS-DArT-seq, em associação com nosso protocolo computacional, representa uma alternativa interessante e custo-efetiva para o estudo desse fenômeno biológico em plantas.

## 5.3 METILAÇÕES DO DNA IDENTIFICADAS E COMPARAÇÃO ENTRE AS AMOSTRAS AVALIADAS

### 5.3.1 COMPARAÇÃO DOS PERFIS DE METILAÇÃO ENTRE AS AMOSTRAS AVALIADAS

Não foi possível observar uma grande diferenciação entre os perfis de metilação de cada amostra. De fato, conforme o esperado, em se tratando de amostras do mesmo genótipo coletados simultaneamente das mesmas árvores e, portanto, submetidos às mesmas condições ambientais, a grande maioria das metilações detectadas foram comuns às três amostras avaliadas, embora existam pequenos conjuntos de citosinas metiladas específicas em cada uma das amostras. Essas diferenças, provavelmente, refletem modificações nos padrões de metilação relativas aos processos de desenvolvimento e diferenciação celular. Essa hipótese é reforçada pela observação do maior compartilhamento de metilações do DNA entre as amostras foliares do que entre estes e o xilema.

Em relação aos TEs metilados, não é possível observar diferenças relevantes entre as amostras com relação às classes de TEs (Figura 19). Além disso, é importante notar que a amostragem e definição dos sítios MSD não aparenta viés de seleção relacionado às classes de transposons (teste qui-quadrado) sendo, portanto, adequada para utilização em plantas, cujos genomas são enriquecidos nestes elementos. Contudo, comparando TEs metilados ao conjunto total de TEs do genoma, algumas classes de TEs são menos (LARD e MITE) ou mais (HELITRON, LINE e TEs aninhados, *Nested TEs*) representadas nas amostras do que o esperado (Figura 20). Com a capacidade de amostragem desta técnica, não é possível afirmar se estes resultados são um indício de que existem diferenças na regulação da metilação para cada classe de TEs em *E. grandis*. Igualmente, não é possível generalizar esta observação para outras espécies. Contudo, estas são questões interessantes que podem ser abordadas futuramente em estudos mais abrangentes.

Devido ao alto compartilhamento de metilações entre as amostras, também foi observada alta similaridade entre os conjuntos de genes que contêm ao menos uma metilação do DNA detectada (Figura 21). Nesse sentido, embora existam termos GO enriquecidos

para os diferentes conjuntos de genes, de maneira geral, estes são relacionados com processos biológicos básicos, não refletindo funções específicas que possam ser relacionadas com os papéis biológicos das diferentes amostras (Tabela 5).

## 5.4 COMPARAÇÃO COM OUTRAS METODOLOGIAS PARA DETECÇÃO DE METILAÇÃO DO DNA USANDO NGS

Outras abordagens associando a utilização de enzimas de restrição sensíveis à metilação e NGS têm sido relatadas na literatura. Ambos [Marsh, Cottrell e Goldman \(2016\)](#) e [Chwialkowska et al. \(2017\)](#) relataram a detecção de metilações do DNA a partir da digestão do DNA utilizando a enzima *HpaII*, seguida pelo sequenciamento dos fragmentos gerados. Em ambos os casos, foram desenvolvidas metodologias de análise que utilizam o pacote `edgeR` para a análise comparativa entre as contagens de fragmentos obtidas em amostras distintas, ou grupos de amostras, determinando assim a presença ou ausência de metilações. Adicionalmente, [Mayne et al. \(2018\)](#) desenvolveram o pacote de análise `msgbsr`, implementado em R, que se propõe a realizar a análise de dados de sequenciamento de bibliotecas de fragmentos gerados a partir da digestão do DNA com enzimas de restrição sensíveis à metilação, em uma abordagem também dependente da utilização do pacote `edgeR`.

Contudo, até o momento, as metodologias apresentadas diferem da abordagem utilizada neste estudo, principalmente, em relação à forma de determinação das metilações do DNA. Enquanto as demais metodologias apresentadas dependem da comparação entre diferentes grupos amostrais para a detecção das metilações do DNA, na metodologia de análise desenvolvida neste estudo e baseada no sequenciamento MS-DArT-seq, a detecção das metilações do DNA pode ser inferida diretamente pela comparação entre as bibliotecas *PstI-MspI* e *PstI-HpaII*, permitindo a detecção de milhares de metilações do DNA em cada amostra, de forma independente da comparação entre grupos, desde que sejam realizadas réplicas biológicas.

# CONCLUSÃO

Neste estudo, foi elaborado um protocolo computacional para detecção de metilações de DNA a partir de dados produzidos com a tecnologia MS-DArT-seq (*Methylation Sensitive DArT sequencing*). Além da detecção das metilações do DNA, caso exista um genoma de referência para a espécie avaliada, o protocolo analítico desenvolvido permite também a contextualização genômica dessas metilações a partir da utilização da anotação de genes e elementos transponíveis, caso estas estejam disponíveis para o genoma. Utilizando essas três amostras de *Eucalyptus grandis*, foi possível demonstrar que a técnica MS-DArT-seq possui grande reprodutibilidade entre as réplicas biológicas, além de permitir a amostragem de milhares de sítios de metilação do DNA a um baixo custo e com grande acurácia, como demonstrado pela comparação com BS-seq. Além disso, uma miríade de testes estatísticos foram aplicados visando a caracterização dos dados gerados por essa metodologia para que o protocolo seja confiavelmente aplicado em outros estudos.

Ao todo, foram avaliados mais de 30 mil sítios de metilação do DNA e, aproximadamente, 6 mil metilações foram identificadas. A caracterização genômica das metilações identificadas demonstrou a existência de um viés de amostragem da técnica MS-DArT-seq, que promove um forte direcionamento para amostragem de sítios de metilação em regiões gênicas, sendo esta uma característica interessante para estudos direcionados ao entendimento da regulação epigenética da expressão gênica.

Devido ao custo acessível da técnica, associado à capacidade de amostragem de milhares de sítios no genoma, essa abordagem é particularmente interessante para a aplicação em estudos com grande número de amostras, como estudos populacionais. Dessa forma, espera-se que o protocolo de análise apresentado neste estudo seja utilizado para a avaliação dos perfis de metilação do DNA em outras espécies, bem como em delineamentos experimentais mais complexos, contribuindo assim para o desenvolvimento do campo da epigenética.

## Parte III

# CAPÍTULO 2<sup>1</sup> - PADRÕES DE METILAÇÃO DE DNA EM CLONES ELITE DE *Eucalyptus* PLANTADOS EM AMBIENTES CONTRASTANTES.

---

<sup>1</sup> Os resultados descritos neste capítulo encontram-se também publicados na forma de artigo científico no periódico Forest Ecology and Management (<https://doi.org/10.1016/j.foreco.2020.118319>), como demonstrado na Figura 22.



## Forest Ecology and Management

Volume 474, 15 October 2020, 118319



# Patterns of DNA methylation changes in elite *Eucalyptus* clones across contrasting environments

Wendell Jacinto Pereira <sup>a, 1, 2</sup> ✉, Marília de Castro Rodrigues Pappas <sup>b, 1</sup> ✉, Otávio Camargo Campoe <sup>c, d</sup> ✉, José Luiz Stape <sup>e</sup> ✉, Dario Grattapaglia <sup>b, f</sup> ✉, Georgios Joannis Pappas Jr <sup>a</sup> ✉

Show more

<https://doi.org/10.1016/j.foreco.2020.118319>

[Get rights and content](#)

### Highlights

- Epigenetic changes play an important role in the adaptation response of forest trees.
- *Eucalyptus* is a robust system to track epigenetic changes in response to environment.
- Site-specific DNA methylation changes were identified for each of five eucalypt clones.
- No consistently shared epigenetic marks were found for all clones across environments.
- Results showed that the observed epigenetic variation is mainly genotype dependent.

Figura 22 – Página inicial do artigo publicado no periódico Forest Ecology and Management e baseado no estudo descrito neste capítulo.



## 6 INTRODUÇÃO

Há muito tempo geneticistas buscam o entendimento dos processos que permitem a adaptação de organismos aos desafios impostos pelo ambiente. Dentre estes processos, a plasticidade fenotípica é um dos mais intrigantes, cujas bases genéticas e moleculares não são completamente compreendidas. Esta pode ser definida como a capacidade de um mesmo genótipo de produzir diferentes fenótipos dependendo das condições ambientais na quais ele se encontra ou se desenvolveu (PIGLIUCCI; MURREN; SCHLICHTING, 2006). Atualmente, o estudo da variação epigenética é uma das muitas abordagens aplicadas para o estudo deste fenômeno (HUANG et al., 2017).

Em teoria, a modulação da expressão gênica em resposta à presença de sinais externos, ou como parte do desenvolvimento do organismo, poderia ser desencadeada por alterações reversíveis na estrutura da cromatina, como pela adição ou remoção de metilação do DNA e modificações de histonas, sem contudo haver mudança na sequência de bases (NIEDERHUTH; SCHMITZ, 2017; ECKER et al., 2018). Dessa forma, sugere-se que o “epigenoma”, definido como o conjunto dessas modificações não relacionadas com a alteração da sequência de DNA contida no genoma, funcione como uma camada intermediária de comunicação entre o genoma e o meio ambiente (ECKER et al., 2018).

Sendo as plantas organismos sésseis, a capacidade de adaptação rápida e de forma sustentável é fundamental para a sobrevivência frente a estresses bióticos e abióticos, especialmente, em um cenário de mudança climática. Diversos exemplos sugerem que a regulação epigenética, provavelmente, contribua com a capacidade de adaptação em curto prazo das plantas em face a estresses ambientais (PECINKA; ABDELSAMAD; VU, 2013; CORTIJO et al., 2014; SECCO et al., 2015; WIBOWO et al., 2016; SCHMID et al., 2018).

Já a longo prazo, o ganho de características adaptativas pelo acúmulo de mutações é particularmente vantajoso em espécies de vida longa, como as espécies florestais. O acúmulo de mutações com potencial adaptativo é ainda mais favorecido em espécies que produzem um grande número de sementes e com baixo nível de dispersão, aumentando a probabilidade de ocorrência de mutações que, selecionadas pelas condições ambientais locais, poderão ser herdadas pelas gerações futuras (HERMAN et al., 2014; PLATT et al., 2015). Nesse sentido, uma vez que citosinas metiladas sofrem mutação mais frequentemente do que não metiladas, estas traduzem-se em um aumento das taxas mutacionais que é diretamente relacionado a fatores epigenéticos e pode estar envolvido na adaptação local (PLATT et al., 2015).

Estudos epigenéticos em populações naturais ou grupos de indivíduos geneticamente distintos enfrentam um desafio de diferenciar efeitos causados pela epigenética e a própria

variação genética (LATZEL et al., 2013; DUBIN et al., 2015; SCHMID et al., 2018; SU et al., 2018). Parte da metilação do DNA está relacionada, ou é controlada, pela variação genética em genes que controlam as vias responsáveis pelo estabelecimento e manutenção da metilação do DNA (PECINKA; ABDELSAMAD; VU, 2013). No entanto, uma parcela das metilações do DNA são independentes da variação genética (SCHMITZ et al., 2013) e pode estar relacionada com processos adaptativos em plantas, representando, portanto, um relevante aspecto da variação biológica a ser estudada.

Os efeitos das marcas epigenéticas na adaptação às condições ambientais, resistência a estresses bióticos e abióticos vêm sendo estudados por vários grupos em diferentes espécies, incluindo organismos modelo como *Arabidopsis thaliana* (JIANG et al., 2014; KOOKE et al., 2015; KELLER; LASKY; YI, 2016; GANGULY et al., 2017; SCHMID et al., 2018), espécies cultivadas (RAZA et al., 2017; SPRINGER; SCHMITZ, 2017) e espécies florestais (CI et al., 2016; LAFON-PLACETTE et al., 2018; SU et al., 2018; BROECK et al., 2018). Evidências do papel adaptativo das metilações do DNA são também encontradas em iniciativas mais amplas, como o 1001 metilomas de *Arabidopsis* (*1001 Arabidopsis DNA methylomes*) que demonstrou que os níveis de metilação do DNA em elementos transponíveis de diferentes populações apresentam correlação positiva com a latitude e níveis de precipitação, e correlação negativa com temperaturas mais elevadas (KAWAKATSU et al., 2016). Mudanças na metilação do DNA foram também associadas com práticas agrônômicas e coordenadas geográficas no cultivo de uva (*Vitis vinifera*), permitindo inclusive o uso dessas marcas para inferência sobre a qualidade de uma região para produção de vinho (XIE et al., 2017b; XIE et al., 2017a). Adicionalmente, metilação do DNA apresenta o potencial para ser utilizada como biomarcador para a avaliação de estresses ambientais (FEIL; FRAGA, 2012).

Devido ao papel central das espécies florestais na manutenção de diversos biomas, diversos estudos têm sido realizados visando o entendimento da influência de fatores ambientais na capacidade adaptativa desses organismos (BRÄUTIGAM et al., 2013; CI et al., 2016; GUGGER et al., 2016). Espécies do gênero *Populus* estão entre as mais estudadas e diversos estudos têm reportado mudanças no epigenoma em resposta a fatores ambientais, como a disponibilidade de água (GOURCILLEAU et al., 2010; RAJ et al., 2011; LAFON-PLACETTE et al., 2018), biodisponibilidade de fósforo (SCHÖNBERGER et al., 2016) e estresse salino (SU et al., 2018). Em *Eucalyptus grandis*, recentemente, foi demonstrado o papel funcional de regiões de cromatina aberta e de modificações de histonas na regulação gênica (BROWN et al., 2019). Além disso, foi demonstrada a regulação epigenética pela indução e/ou remoção de metilação do DNA em plantas da espécie *Eucalyptus globulus* submetidas a condições de deficit hídrico seguido por reidratação (CORREIA et al., 2016)), usando a técnica MS-RAPD (*methyl sensitive random amplified polymorphic DNA*).

Mais recentemente, Champigny et al. (2020) demonstraram que, em populações de

---

*Populus balsamifera*, é possível o uso de epigenomas para a estimar caracteres fenotípicos. Embora seja tecnicamente factível a obtenção de múltiplos metilomas completos para qualquer organismo, o alto custo associado torna essa abordagem ainda impraticável para um grande número de espécies para as quais os recursos financeiros são escassos. Desta forma, muitos estudos dependem da aplicação de técnicas que sejam financeiramente acessíveis, a despeito do menor poder amostral. MS-DArT-seq, descrita detalhadamente no Capítulo II e em [Pereira et al. \(2020\)](#), é uma tecnologia que se enquadra neste perfil, sendo adotada na investigação descrita neste capítulo.

Neste estudo, a metodologia MS-DArT-seq, combinada com a metodologia de análise descrita no Capítulo II, foi aplicada para a avaliação do padrão de metilação de milhares de sítios ao longo do genoma em *Eucalyptus*. O material vegetal foi cedido pelo programa TECHS (*Tolerance of Eucalyptus Clones to the Hydrous and Thermal Stresses*), cujo desenho experimental permitiu a comparação dos mesmos genótipos plantados em diferentes localidades, com diferentes condições ambientais ([BINKLEY et al., 2017](#); [BINKLEY et al., 2020](#)), minimizando os efeitos confundidores relacionados ao *background* genético. Ao todo, foram avaliados os perfis de metilação de duas amostras, folhas e xilema, em duas réplicas de cinco clones elite por ambiente. Além da comparação geral dos perfis de metilação entre clones e ambientes, análises foram conduzidas visando identificar se existe associação entre as metilações do DNA identificadas e caracteres fenotípicos associados ao crescimento.

# OBJETIVOS

## OBJETIVO GERAL

Comparação dos perfis de metilação de DNA em clones de *Eucalyptus* plantados em diferentes ambientes, com relevante contraste de temperatura e regime hídrico.

## OBJETIVOS ESPECÍFICOS

- Identificação dos perfis de metilação de DNA de duas amostras vegetais, folhas e xilema, em cinco clones elite de *Eucalyptus*, plantados em duas localidades com ambientes contrastantes.
- Comparação dos perfis de metilação de DNA de cada clone entre ambientes e entre clones no mesmo ambiente.
- Verificar se existem metilações de DNA específicas de cada ambiente, ou seja, detectadas em todos os clones plantados em um mesmo ambiente.
- Caracterização do contexto genômico das metilações de DNA identificadas.

# FLUXOGRAMA DO CAPÍTULO

Abaixo encontra-se uma visão global das etapas do presente capítulo.

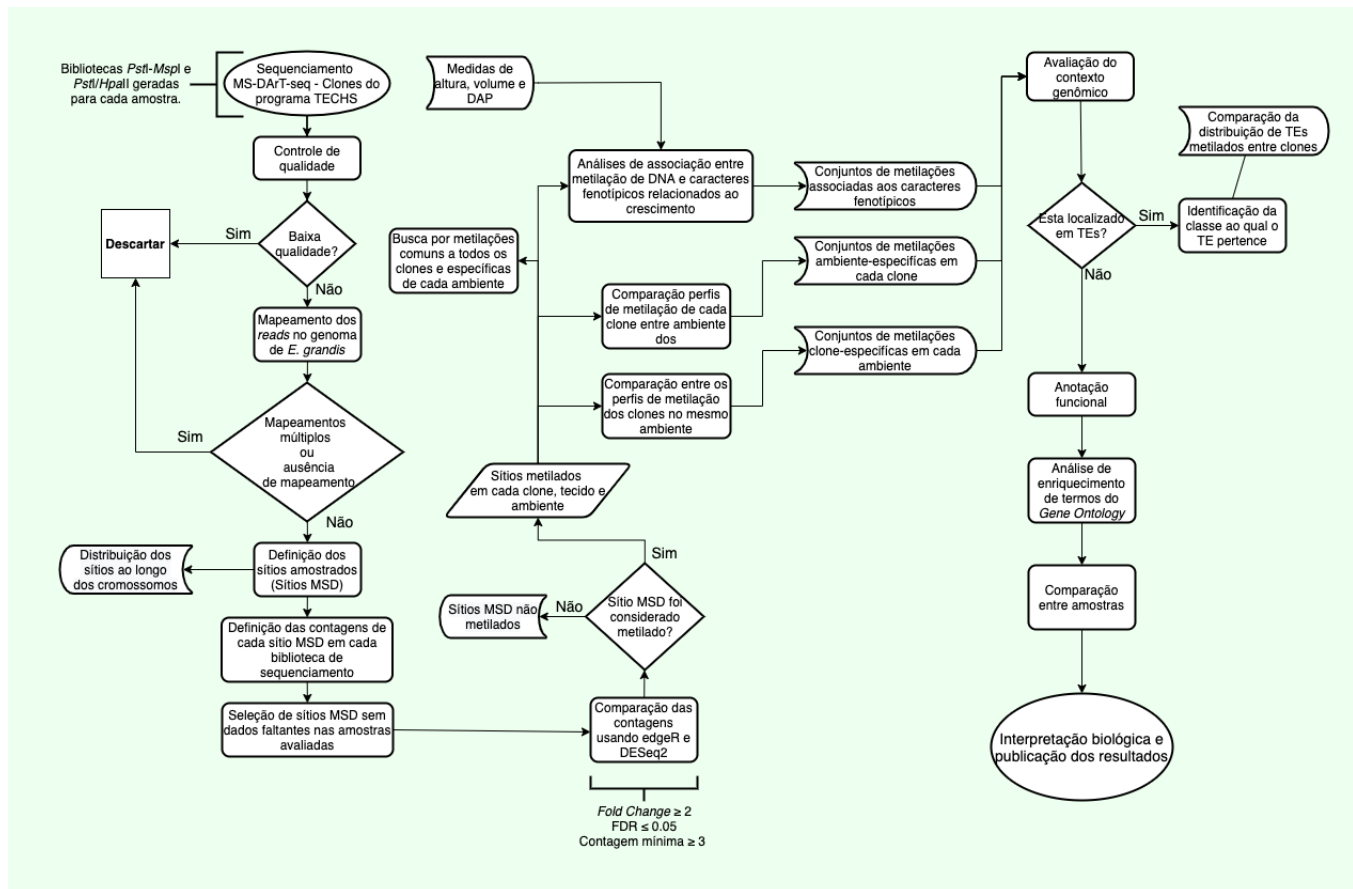


Figura 23 – Fluxograma representando as etapas do estudo conduzido para a geração dos resultados descritos neste capítulo.

# 7 MATERIAL E MÉTODOS

## 7.1 MATERIAL VEGETAL

### 7.1.1 CLONES SELECIONADOS PARA O ESTUDO

Neste estudo, foram avaliados os perfis de metilação de cinco genótipos (clones) de eucaliptos, quatro híbridos interespecíficos derivados do cruzamento entre *Eucalyptus grandis* e *Eucalyptus urophylla* e um de espécie pura (*Eucalyptus urophylla*). Esses genótipos foram propagados via clonagem e plantados em sítios experimentais que fazem parte do programa TECHS (*Tolerance of Eucalyptus Clones to the Hydrous and Thermal Stresses*) (BINKLEY et al., 2017; BINKLEY et al., 2020). Um aspecto relevante é que esses são clones comerciais amplamente utilizados por empresas florestais. Foram aqui codificados como A1, B2, D4, E5 e Q8, seguindo nomenclatura adotada pelo programa TECHS e apresentada em Binkley et al. (2020). Os cinco clones utilizados neste estudo foram selecionados dentre os clones TECHS por serem híbridos de *E. grandis*, espécie que possui genoma de referência, e *E. urophylla*, espécie mais próxima filogeneticamente de *E. grandis*, pertencente à mesma seção Latoangulatae no sugbgênero Symphyomyrtus.

Para este estudo, foram coletadas amostras de duas árvores de cada clone, duas réplicas biológicas, com idade aproximada de 3 anos e meio em cada ambiente. Os mesmos cinco genótipos tiveram amostras de folha e xilema coletadas em dois ambientes com características climáticas distintas, nomeados como seco e chuvoso.

### 7.1.2 AMBIENTES EXPERIMENTAIS

Dois ambientes contrastantes foram selecionados para este estudo, sítio TECHS '30' (seco) e TECHS '22' (chuvoso), seguindo a numeração descrita em Binkley et al. (2020). O ambiente chuvoso é localizado no estado de Minas Gerais (Bocaiuva, 17° 6' 55" sul, 43° 49' 16" oeste) e o seco no estado do Paraná (Telêmaco Borba, 24° 19' 28" sul, 50° 36' 59" oeste). A precipitação anual média no sítio chuvoso é de 1.435 mm e de 1.181 mm no ambiente seco. Contudo, a característica mais marcante é que o ambiente seco apresenta uma temporada de seca que pode prolongar-se por até 5 meses. Ainda, o ambiente chuvoso apresenta médias menores para a temperatura anual (18,9 °C versus 21,4 °C), índice de seca (0,615 versus 0,921) e potencial de evapotranspiração anual (883 versus 1.088).

## 7.2 MS-DArT-seq PARA DETECÇÃO DE METILAÇÕES DO DNA

A detecção das metilações do DNA nos genomas dos clones selecionados foi realizada usando a metodologia MS-DArT-seq. Após a extração do DNA, as amostras foram enviadas para a empresa DArT (Canberra, Austrália) que realizou o preparo das bibliotecas e o sequenciamento NGS. Após o recebimento das sequências, o método computacional descrito no Capítulo II foi aplicado para a detecção das metilações do DNA. Para uma descrição completa do métodos, veja as sessões 3.2, 3.3, 3.4 e 3.5 do Capítulo II. Uma ampla descrição das características dos dados gerados por MS-DArT-seq está também disponível em [Pereira et al. \(2020\)](#).

## 7.3 ANÁLISES ESTATÍSTICAS E DE ASSOCIAÇÃO ENTRE AS METILAÇÕES DO DNA IDENTIFICADAS E CARACTERES FENOTÍPICOS RELACIONADOS COM O CRESCIMENTO

Após a identificação das metilações do DNA, análises estatísticas foram aplicadas visando a descrição dos dados obtidos para os caracteres fenotípicos mensurados em cada clone e ambiente. Análises adicionais foram aplicadas para avaliar se existe associação entre as metilações do DNA identificadas e os caracteres fenotípicos nos clones avaliados.

Testes estatísticos foram implementados na linguagem R usando pacotes que fazem parte da instalação padrão do programa. O pacote FactoMineR<sup>1</sup> ([LÊ et al., 2008](#)) foi utilizado para a análise de MCA (*Multiple Correspondence Analysis*). Usando o pacote ggstatplot<sup>2</sup>, testes T de Student (*Student's t-test*) foram aplicados para realizar a comparação entre as medidas fenotípicas de cada clone nos sítios seco e chuvoso. Para comparar as medidas dentro de cada ambiente, testes ANOVA foram realizados seguidos por testes t de Student aplicados par-a-par. Todos os p-valores foram corrigidos para múltiplos testes aplicando-se o procedimento de Benjamini-Hochberg (FDR - *false discovery rate*).

Para identificar se sítios metilados poderiam ser significativamente associados com os caracteres fenotípicos relacionados com o crescimento, especificamente altura (m), diâmetro a altura do peito (cm) e volume (m<sup>3</sup>), foi aplicado um método baseado em modelos mistos de fatores latentes (*latent factor mixed models*) e implementado no pacote R LFMM2<sup>3</sup> ([CAYE et al., 2019](#)).

<sup>1</sup> [cran.r-project.org/web/packages/FactoMineR/](https://cran.r-project.org/web/packages/FactoMineR/)

<sup>2</sup> <https://indrajeetpatil.github.io/ggstatsplot/>

<sup>3</sup> <https://bcm-uga.github.io/lfmm/index.html>

## 7.4 DISPONIBILIDADE DE DADOS E REPRODUTIVIDADE DAS ANÁLISES

Dados gerados pelo sequenciamento MS-DArT-seq, e necessários para reproduzir este estudo, estão disponíveis no sítio Figshare: <<https://doi.org/10.6084/m9.figshare.11782203>>. As análises foram executadas usando programas com código aberto (*open-source*) e *scripts* elaborados em R ou Python para este estudo. Todos os passos da análise foram coordenados e executados usando o gerenciador de fluxos de trabalho Snakemake (KÖSTER; RAHMANN, 2012). O código utilizado e os arquivos de configuração do Snakemake estão disponíveis no sítio GitHub: <[https://github.com/wendelljpereira/techs\\_ms\\_dart\\_seq](https://github.com/wendelljpereira/techs_ms_dart_seq)>.

Além da disponibilização do código, um aplicativo gráfico usando o pacote Shiny<sup>4</sup> foi criado para facilitar a visualização e interpretação dos dados. Neste aplicativo, é possível comparar as metilações do DNA entre clones e entre localidades, visualizar os resultados das análises estatísticas, explorar a anotação gênica, bem como realizar a análise de enriquecimento de termos do *Gene Ontology*<sup>5</sup> utilizando diferentes conjuntos de genes que contêm metilação do DNA identificadas neste estudo. O aplicativo está também disponível em <[https://github.com/wendelljpereira/techs\\_ms\\_dart\\_seq](https://github.com/wendelljpereira/techs_ms_dart_seq)>.

---

<sup>4</sup> [shiny.rstudio.com/](https://shiny.rstudio.com/)

<sup>5</sup> [geneontology.org/](https://geneontology.org/)



## 8 RESULTADOS

### 8.1 CRESCIMENTO DOS CLONES NOS AMBIENTES CONTRASTANTES

No desenho experimental do programa TECHS, ao menos 80 árvores de cada clone foram plantadas em cada localidade, com oito parcelas contendo 10 árvores cada. Três diferentes medidas de crescimento relacionadas com a produtividade foram mensuradas no mesmo mês da coleta das amostras, volume (m<sup>3</sup>), altura (m) e DAP (diâmetro na altura do peito; cm). Estes dados estão consolidados na Figura 24, evidenciando que a localidade influencia de forma significativa o desempenho dos clones, como observado pelas medidas no ambiente chuvoso, em comparação com o ambiente seco, para todos os clones (Figura 24).

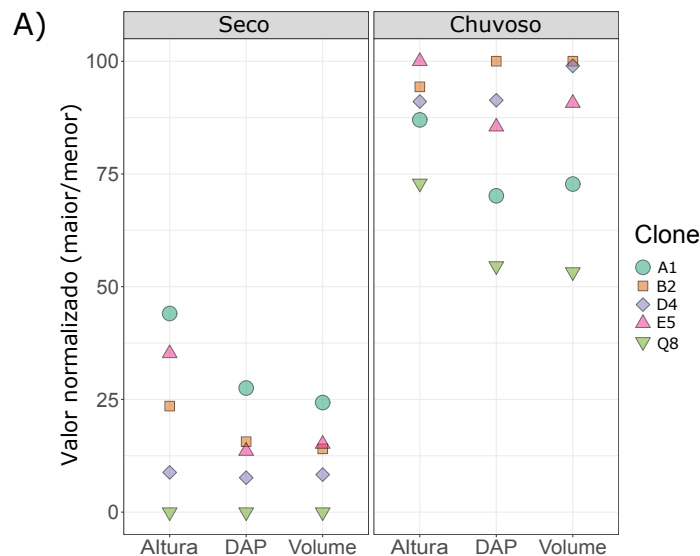


Figura 24 – **Comparação entre o volume dos clones nos ambientes seco e chuvoso.** Em cada medição, os menores valores foram definidos como 0 e os maiores valores como 100. “Seco” representa o sítio de plantio localizado em MG e “Chuvoso” o sítio de plantio localizado no PR.

Além do aumento no volume, é importante observar também a mudança no ranking de produtividade dos clones nos diferentes ambientes. Por exemplo, o clone mais produtivo no ambiente seco (clone A1) é apenas o quarto mais produtivo no ambiente chuvoso (Figura 24). Uma exceção à essa mudança é o clone Q8, que apresentou a menor produtividade nos dois ambientes avaliados.

As mudanças no ranking de produtividade indicam diferentes níveis de plasticidade

fenotípica nesses genótipos. As comparações par-a-par dos valores fenotípicos dentro de cada ambiente demonstram que a maioria das posições dos clones no ranking de produtividade são estatisticamente significativas ( $p$ -valor  $< 0,05$ ), sendo estas diferenças avaliadas pela análise de ANOVA seguida por testes  $t$  de Student (Figura 25). Por exemplo, no ambiente seco, o clone A1 é significativamente mais produtivo do que todos os demais clones, conforme inferido pelas três medidas de crescimento utilizadas (Figura 25). Também no ambiente seco, os clones B2 e E5 são igualmente ranqueados como segundo lugar, de acordo com o volume, não havendo diferença significativa entre eles (Figura 25).

## 8.2 GERAÇÃO DE DADOS VIA MS-DArT-seq

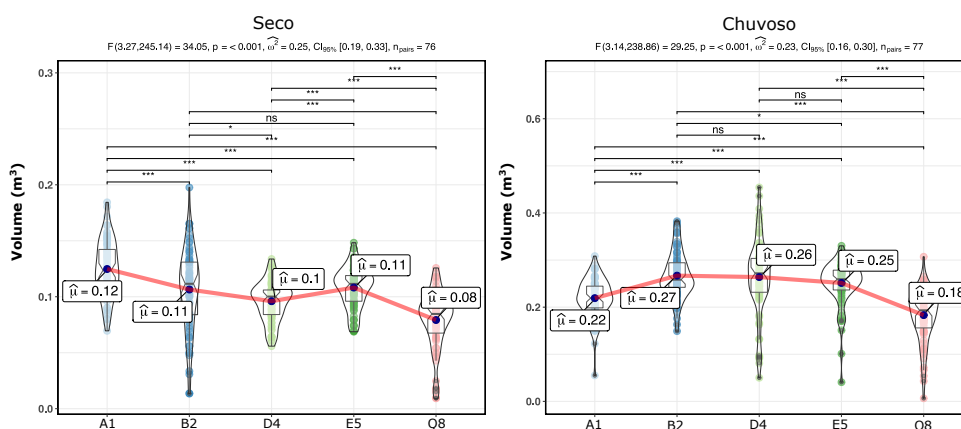
Neste estudo, duas amostras vegetais foram coletadas de cinco clones de *Eucalyptus*, plantados em dois ambientes contrastantes. Visando a obtenção de dados para detecção de metilação de DNA, as amostras foram sequenciadas utilizando a tecnologia MS-DArT-seq e análises com o protocolo computacional descrito no Capítulo II.

Ao todo, foram gerados mais de 246 milhões de sequências (77 pb) pelo sequenciamento MS-DArT-seq dos clones e amostras averiguados neste estudo. Após o controle de qualidade e mapeamento no genoma de referência (*E. grandis*), 60,32% das sequências mapearam em uma única posição genômica e foram utilizadas para a determinação dos sítios de metilação (nomeados como sítios MSD), bem como para a determinação das contagens de cada sítio MSD nas duas bibliotecas de sequenciamento (*PstI-HpaII/MspI*) de cada clone e tecido. Um total de 90.378 sítios MSD foram identificados e tiveram as contagens determinadas para cada biblioteca MS-DArT-seq. Ademais, existe uma alta reprodutibilidade entre as contagens obtidas para as réplicas biológicas, de acordo com coeficiente de correlação de Pearson, ( $\rho=0,973\pm 0,017$ ). A distribuição dos sítios MSD ao longo do genoma de referência (*E. grandis*) demonstra que eles estão distribuídos ao longo dos 11 cromossomos, conforme demonstrado na Figura 26. Uma vez que nem todos os sítios MSD foram amostrados em todos os clones e localidades, optou-se por excluir aqueles com dados faltantes, permitindo assim a comparação fidedigna dos perfis de metilação. Após essa filtragem, 18.569 sítios MSD foram selecionados para folha e 12.885 selecionados para o xilema. Interessantemente, a grande maioria dos sítios identificados nas amostras de xilema, 99,1%, foram também identificados nas amostras de folhas.

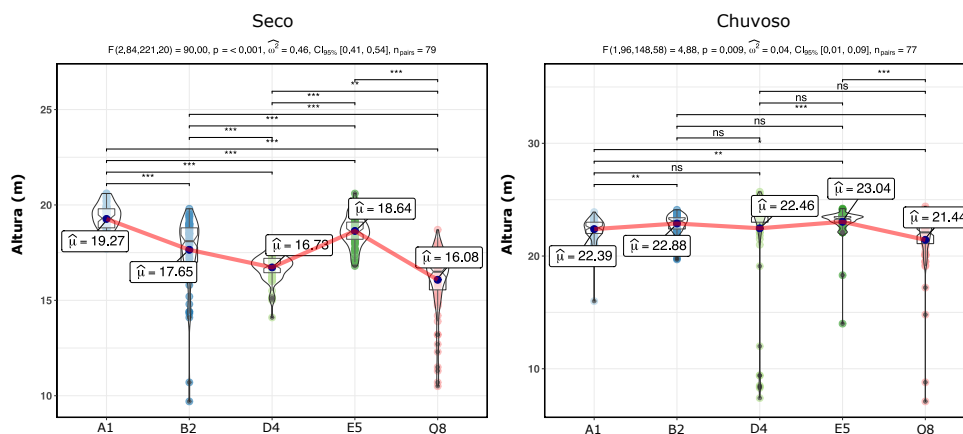
## 8.3 DISTRIBUIÇÃO DAS METILAÇÕES DO DNA ENTRE CLONES

Um total de 4.627 sítios metilados foram identificados nos cinco clones de *Eucalyptus* utilizados neste estudo, excluindo a redundância entre amostras. Na Figura 27, são

**A) Comparação da produtividade dos clones selecionados em cada ambiente - Volume**



**B) Comparação da produtividade dos clones selecionados em cada ambiente - Altura**



**C) Comparação da produtividade dos clones selecionados em cada ambiente - DAP**

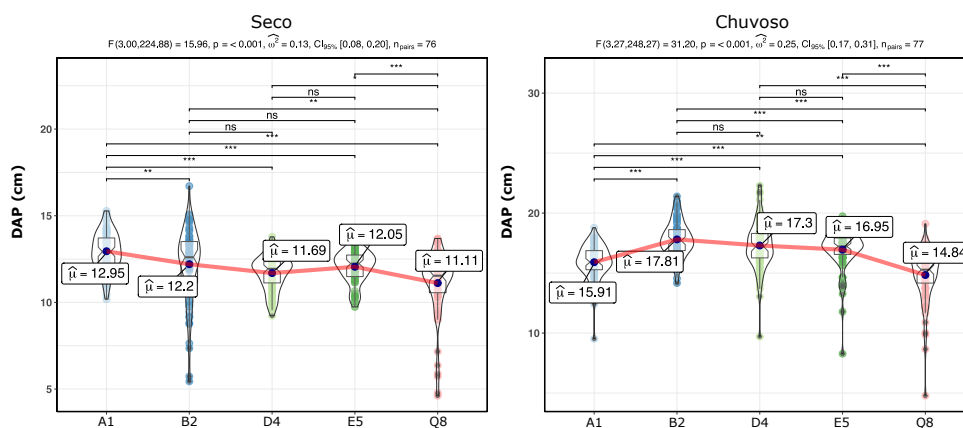


Figura 25 – Comparação entre a produtividade dos clones em um mesmo ambiente. Para cada caractere fenotípico, volume (A), altura (B) e DAP (C), foi realizada a análise de variância (ANOVA) seguida por testes t de Student aplicados par-a-par. À esquerda estão representadas as comparações no ambiente seco e à direita, as comparações no ambiente chuvoso. Em cada comparação, ns (*não significativo*) corresponde a p-valor  $\geq 0,05$ ; \*\*\*, a p-valor  $< 0,001$ ; \*\*, a p-valor  $< 0,01$  e \*, a p-valor  $< 0,05$ .

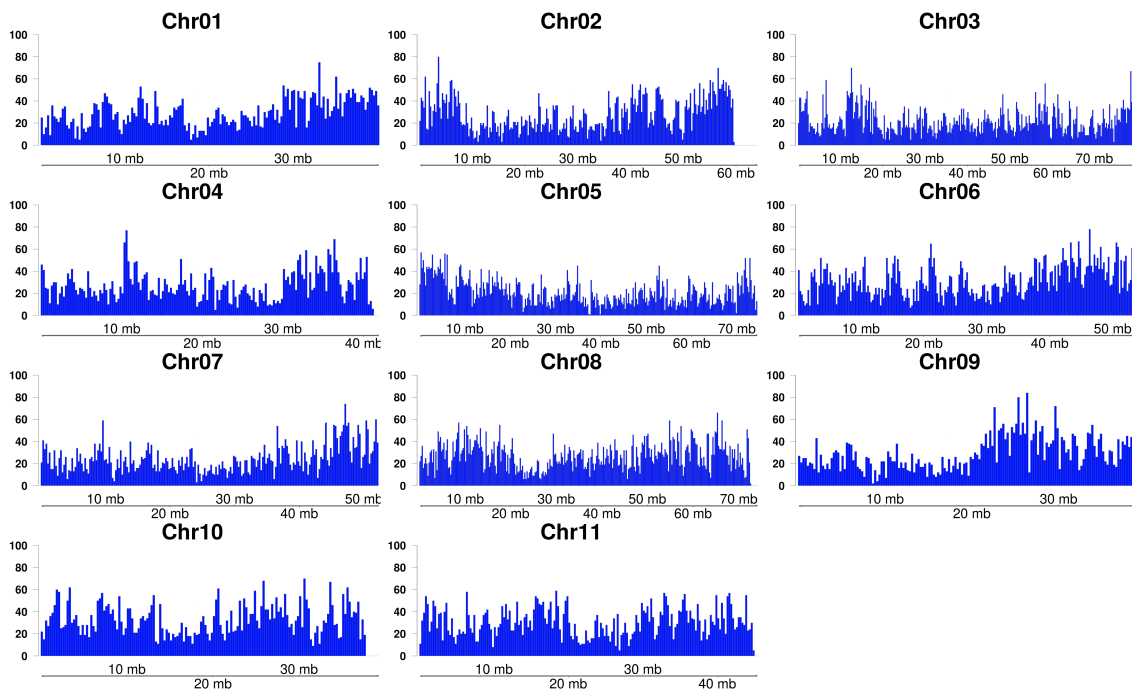


Figura 26 – Distribuição dos sítios MSD ao longo dos cromossomos do genoma de referência de (*E. grandis*).

demonstradas as comparações entre os sítios metilados verificados em cada localidade, para cada clone. Embora a maioria dos sítios metilados sejam comuns a ambos os ambientes, existem sítios metilados específicos para cada ambiente e clone (Figure 27).

Foi também realizada a comparação entre clones no mesmo ambiente. Essa comparação foi realizada para cada tecido separadamente e é demonstrada nas Figuras 28 e 29. Para ambos as amostras foi verificada uma importante parcela de sítios metilados compartilhados por todos os clones em cada ambiente. Além disso, a maioria dos sítios metilados foi compartilhada por dois ou mais clones. Contudo, foram também encontrados conjuntos de sítios metilados específicos para cada clone nas duas amostras investigadas (Figuras 28 e 29). Em particular, no xilema, um clone (Q8) destacou-se devido ao grande número de metilações específicas no ambiente seco, com 763 metilações identificadas.

## 8.4 ANOTAÇÃO FUNCIONAL DOS GENES CONTENDO SÍTIOS METILADOS

Como demonstrado no Capítulo II e em Pereira et al. (2020), uma característica intrínseca da detecção das metilações do DNA via MS-DArT-seq é um forte viés de seleção em favor de regiões gênicas devido ao uso de enzimas sensíveis à metilação, incluindo *Pst*I. O mesmo viés foi verificado para o grupo de metilações identificadas nos cinco clones avaliados neste estudo. A grande maioria dos sítios metilados estão localizados em genes

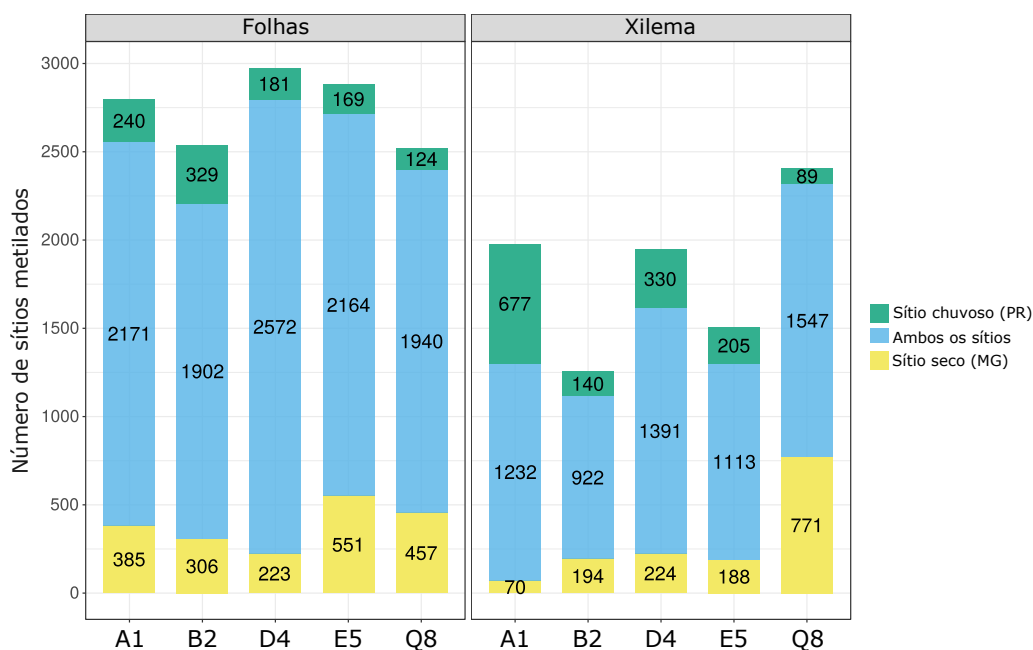


Figura 27 – **Frequência de sítios metilados por clone em cada local de plantio de acordo com o tecido avaliado (folhas e xilema).** Em amarelo, estão representados os sítios metilados que ocorrem especificamente no ambiente seco e, em verde, os sítios metilados exclusivos do ambiente chuvoso.

(69,83%), predominantemente em éxons. Interessantemente, elementos transponíveis, que ocupam uma região muito maior do genoma do que os genes, contribuíram com apenas 10,35% dos sítios metilados identificados. Distribuições similares foram verificadas para todos os clones e amostras, sendo apresentadas na Tabela 6.

Uma vez que conjuntos de metilações do DNA específicos para clone e ambiente foram identificados e verificou-se que grande parte dessas metilações estão localizadas em genes, optou-se por conduzir análises visando determinar se esses conjuntos de genes metilados apresentam funcionalidades específicas que poderiam ser associadas com os caracteres fenotípicos verificados para cada clone. Para cada conjunto de genes metilados, realizou-se a anotação funcional seguida pela análise de enriquecimento de termos do *Gene Ontology* (termos GO). Embora alguns conjuntos de termos GO tenham sido considerados enriquecidos para alguns clones e/ou amostras, a maioria deles representa processos biológicos basais e optou-se por não apresentá-los.

## 8.5 METILAÇÕES DO DNA ASSOCIADAS AO AMBIENTE

O fato de que a maioria de metilações de DNA detectadas foram compartilhadas entre os genótipos indica a possibilidade de que estas sejam derivadas de uma resposta ambiental que independe do *background* genético. Para avaliar essa hipótese, foi realizada a busca por sítios MSD que sejam metilados em todos os clones especificamente em um

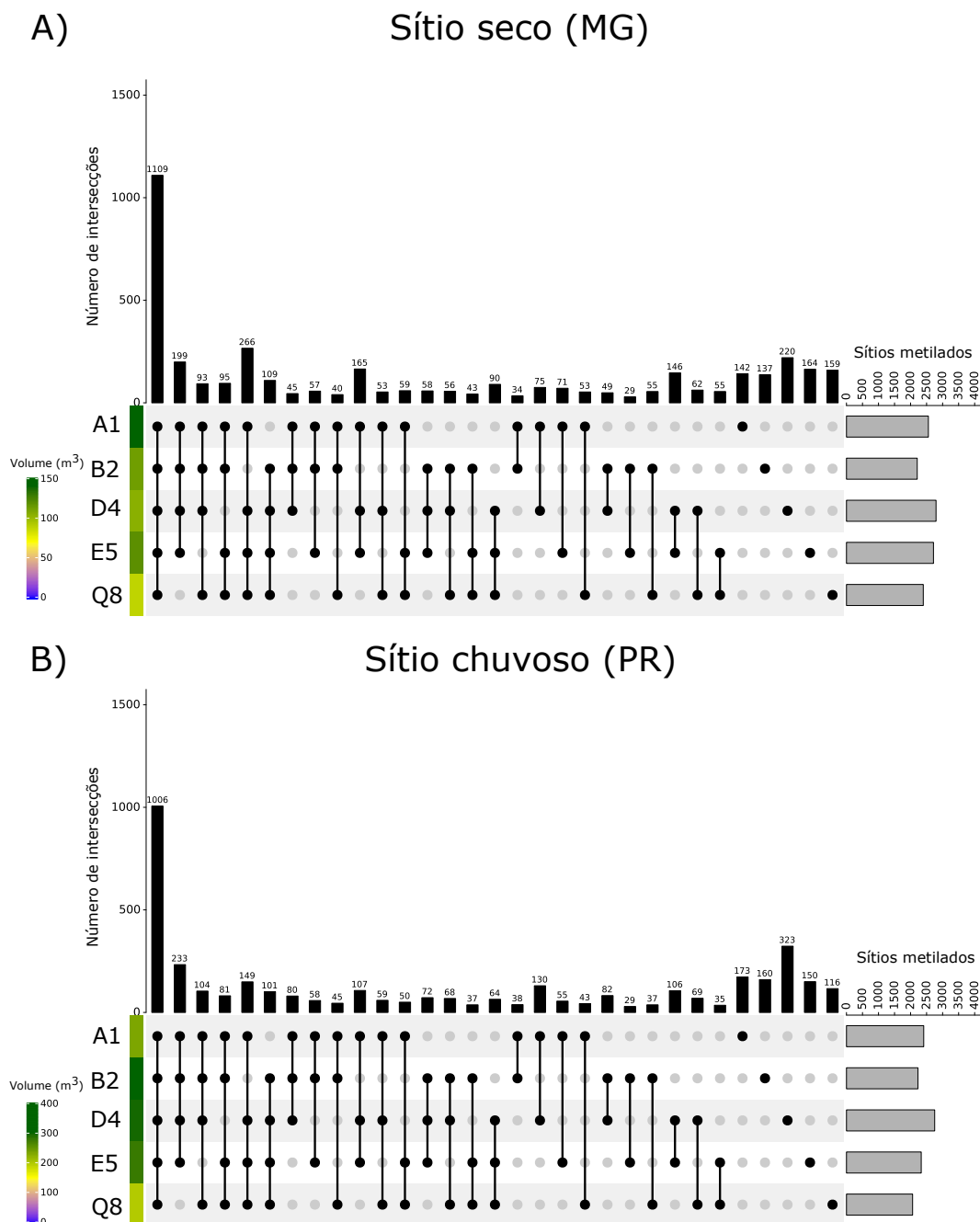


Figura 28 – Comparação entre todos os sítios metilados detectados nas amostras de folha de cada clone em cada sítio experimental. Nesta representação, as barras pretas demonstram o número de sítios metilados que são exclusivamente compartilhados pelos clones sendo comparados, representados pelos pontos pretos conectados na parte inferior de cada gráfico. As barras cinzas na lateral direita dos gráficos demonstram o número total de sítios metilados detectados em cada clone. Os *heatmaps* localizados à esquerda de cada gráfico demonstram a produtividade de cada clone no ambiente seco (gráfico superior) e chuvoso (gráfico inferior) inferida pelo volume ( $m^3$ ).

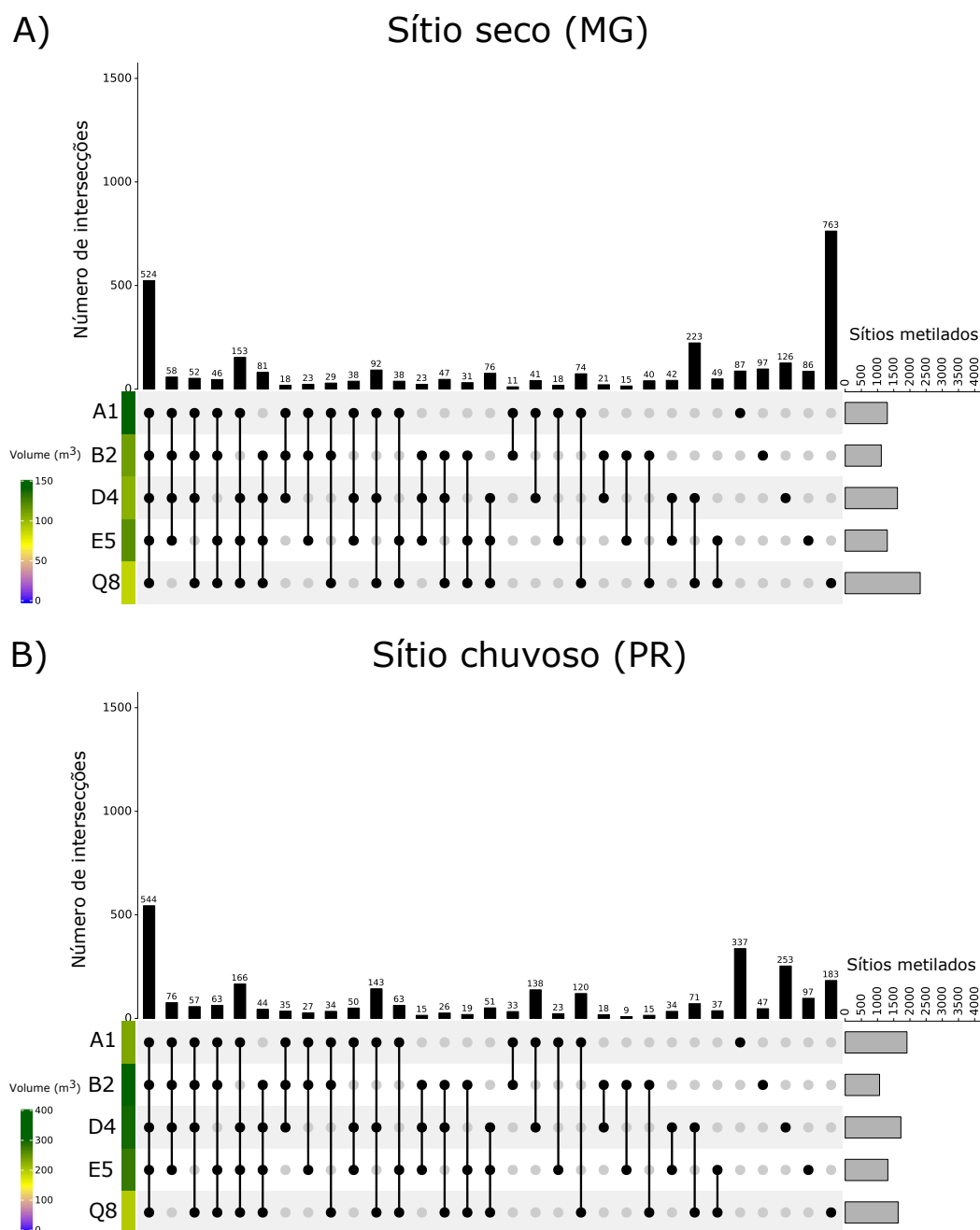


Figura 29 – Comparação entre todos os sítios metilados detectados nas amostras de xilema de cada clone em cada sítio experimental. Nesta representação, as barras pretas demonstram o número de sítios metilados que são exclusivamente compartilhados pelos clones sendo comparados, representados pelos pontos pretos conectados na parte inferior de cada gráfico. As barras cinzas na lateral direita dos gráficos demonstram o número total de sítios metilados detectados em cada clone. Os *heatmaps* localizados à esquerda de cada gráfico demonstram a produtividade de cada clone no ambiente seco (gráfico superior) e chuvoso (gráfico inferior) inferida pelo volume ( $m^3$ ).

Tabela 6 – Distribuição dos sítios metilados em cada clone de acordo com o contexto genômico.

Classe	Subclasse	Sítio TECHS	A1	B2	E5	Q8	D4
Genes	Íntron ou UTR	Seco - MG	283	184	247	393	293
		Chuvoso - PR	407	283	251	294	319
	Éxon	Seco - MG	965	714	907	1.347	1.036
		Chuvoso - PR	1.328	1.014	863	1.070	1.120
	Total	Seco - MG	1.248	898	1.154	1.740	1.329
		Chuvoso - PR	1.735	1.297	1.114	1.364	1.439
TEs	Íntron or UTR	Seco - MG	102	45	101	136	100
		Chuvoso - PR	193	72	81	107	116
	Intergenes	Seco - MG	30	19	22	40	35
		Chuvoso - PR	44	32	20	25	34
	Total	Seco - MG	132	64	123	176	135
		Chuvoso - PR	237	104	101	132	150
Anotações aninhadas	Sobreposição entre TEs e genes	Seco - MG	8	6	6	8	9
		Chuvoso - PR	11	8	6	7	9
	TEs aninhados	Seco - MG	179	51	63	100	73
		Chuvoso - PR	110	64	62	77	78
	Total	Seco - MG	78	57	69	108	82
		Chuvoso - PR	121	72	68	84	87
Intergenes	Intergenes e fora de TEs	Seco - MG	179	108	141	230	173
		Chuvoso - PR	320	141	134	185	197

ambiente. Por exemplo, sítios MSD que foram detectados como metilados no ambiente seco e não metilados no sítio chuvoso e vice-versa.

Tanto para as amostras de folha quanto de xilema, não há casos de metilações que ocorram em todos os clones e que discriminem os dois ambientes. Mesmo após relaxar a busca de metilações ocorrendo em quatro ou três clones, não foi possível identificar nenhum sítio metilado específico para um ambiente. Somente quando os clones são comparados par-a-par, é possível identificar conjuntos de sítios MSD com perfis de metilação contrastantes entre os ambientes, como apresentado na Figura 30. Por exemplo, 27 sítios são compartilhados pelos clones A1 e B2 e possuem estado de metilação diferente entre os dois ambientes (Figura 30).

Em seguida, a análise de correspondências múltiplas (MCA), que é similar à PCA, mas desenvolvida para dados categóricos ao invés de contínuos (HUSSON; LÊ; PAGÈS, 2017), foi aplicada para verificar como os genótipos e ambientes influenciam os perfis de metilação detectados. Como apresentado na Figura 31, os resultados demonstram que os genótipos se agrupam de forma independente do ambiente. Indicando assim que, possivelmente, o *background* genético influencia substancialmente a ocorrência de metilações de DNA identificadas por MS-DaT-seq nos clones e amostras avaliadas neste estudo (Figura 31).

Além disso, usando as metilações investigadas no xilema, a análise de MCA demonstra que o agrupamento também reflete a produtividade dos clones. Por exemplo, os dois clones menos produtivos no ambiente seco, Q8 e D4 estão localizados no quadrante



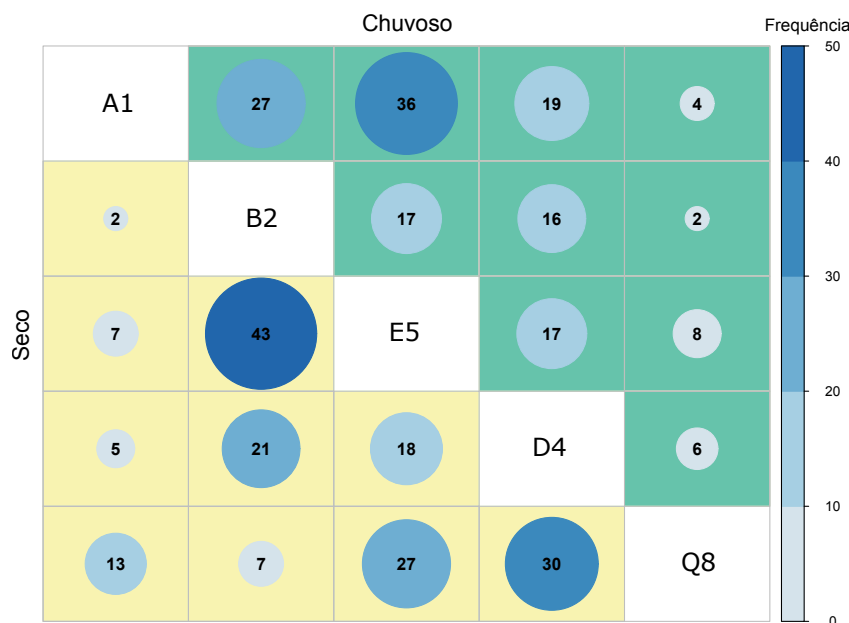


Figura 30 – **Comparações par-a-par dos sítios metilados ambiente-específicos.** Os números dentro de cada círculo representam a quantidade de sítios metilados compartilhados pelo par de clones em um ambiente e que possuem o estado de metilação oposto (não metilados) no outro ambiente. Na diagonal superior direita (em verde), estão representados os sítios que são metilados no ambiente chuvoso e, na diagonal inferior esquerda, em amarelo, estão os sítios metilados no ambiente seco.

inferior direito enquanto os clones mais produtivos, E5 e B2, estão localizados no quadrante superior direito (Figura 31). Também chama a atenção o maior distanciamento entre as amostras do clone A1 nos ambiente seco e chuvoso. Quando comparado ao observado para os outros clones, verifica-se que a distância entre as amostras deste clone é maior do que dos demais. Esse fato é interessante porque este é o clone que apresenta a mudança mais intensa no ranking de produtividade, sendo o clone mais produtivo no ambiente seco e apenas o quarto mais produtivo no ambiente chuvoso (Figura 24).

Para folhas, a relação entre a separação dos clones no gráfico MCA e as medidas de produtividade é menos clara. Contudo, é também possível observar o agrupamento dos clones mais produtivos no ambiente seco, clones A1 e E5, no quadrante inferior direito da Figura 32.

## 8.6 ASSOCIAÇÃO ENTRE SÍTIOS METILADOS E CARACTERES FENOTÍPICOS RELACIONADOS COM CRESCIMENTO

A partir da observação obtida pela análise de MCA, de que o status de metilação dos sítios MSD pode estar relacionado a caracteres fenotípicos, decidiu-se aplicar uma análise de associação entre gene e ambiente (GEA, *gene-environment association*). Os 4.627

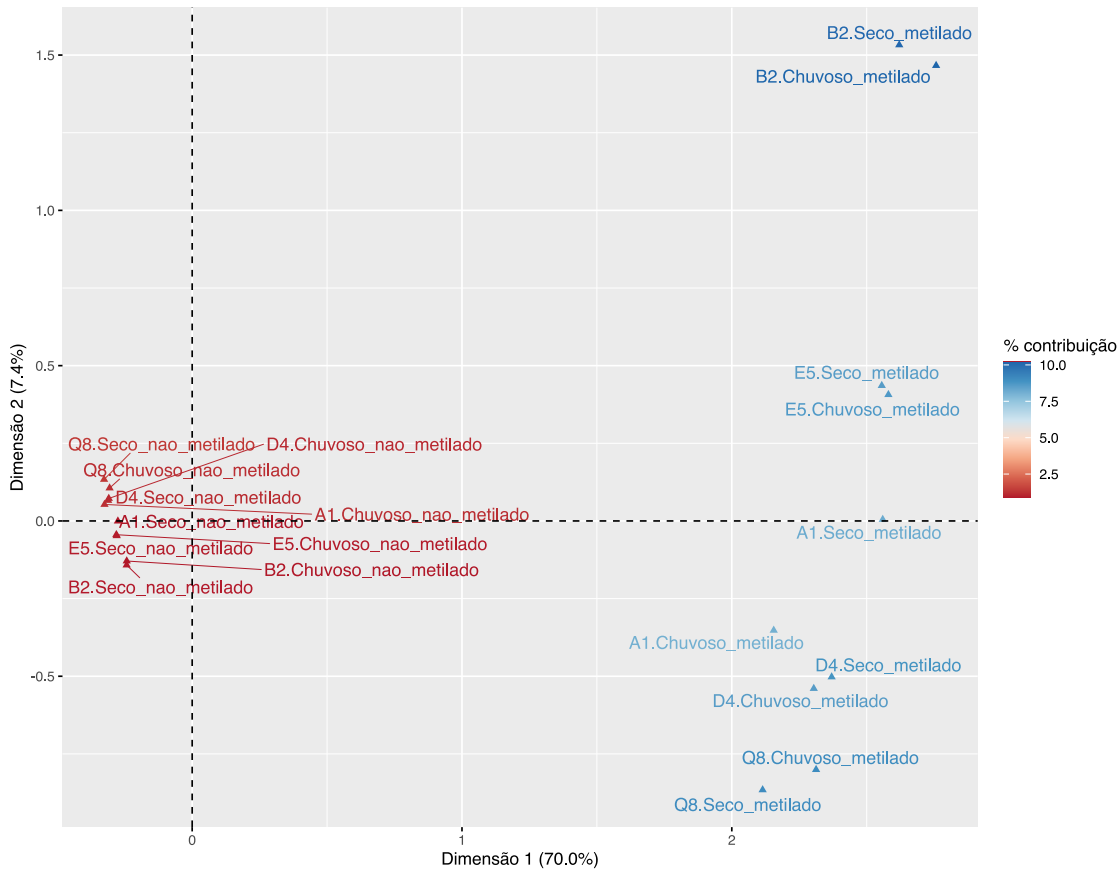


Figura 31 – **Análise de correspondências múltiplas (MCA) dos sítios MSD identificados nas amostras de xilema.** No eixos Y e X estão as porcentagens da variância explicada. Os rótulos representam o nome dos clones, local de plantio e o estado de metilação dos sítios MSD. A escala de cores representa a porcentagem da contribuição de cada amostra para a variância explicada.

sítios metilados não redundantes amostrados em todos os clones foram utilizados para detectar associações entre as marcas epigenéticas e as medidas dos caracteres fenotípicos relacionados com crescimento (volume, altura e DAP), usando o pacote LFMM2.

Um total de 445 sítios metilados foram significativamente associados ( $p$ -valor < 0,01) com um ou mais dos três caracteres fenotípicos avaliados (Figura 33-A, 33-C, 33-D e Figura 35-A), quando consideradas as amostras de folha. Deste total, 242 foram associados com o volume, 418 com a altura e 68 com DAP (Figura 35-A). Vários dos sítios metilados foram associados com mais de um caractere fenotípico, incluindo 188 (42,25%) que foram associados com volume e altura e 27 (6,1%) que foram associados com os três caracteres fenotípicos. Adicionalmente, parte dos sítios metilados foram associados exclusivamente com um caractere fenotípico, sendo 27 associados com o volume (6,1%) e 162 com a altura (36,4%).

Considerando o xilema, um número menor de sítios metilados foi associado com os

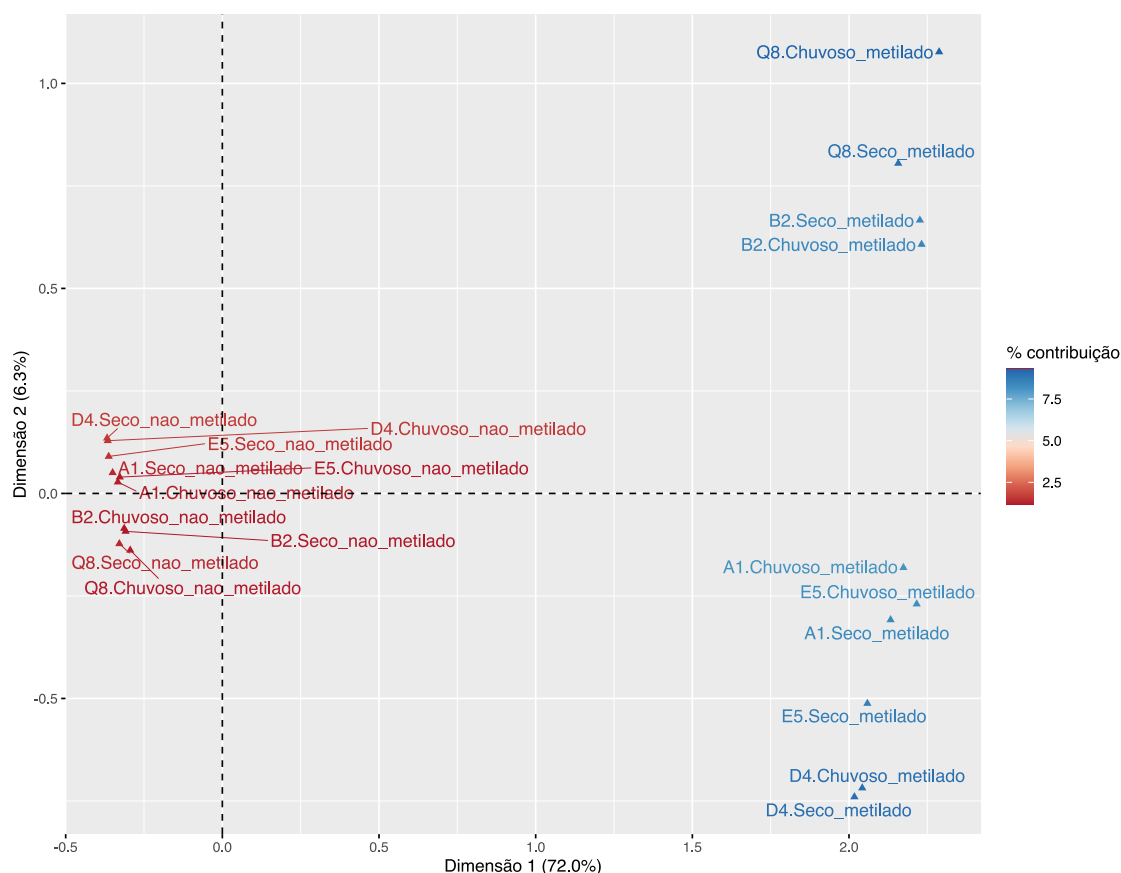


Figura 32 – **Análise de correspondências múltiplas (MCA) dos sítios MSD identificados nas amostras de folha.** No eixos Y e X, estão as porcentagens da variância explicada. Os rótulos representam o nome dos clones, local de plantio e o estado de metilação dos sítios MSD. A escala de cores representa a porcentagem da contribuição de cada amostra para a variância explicada.

caracteres fenotípicos avaliados, um total de 312 sítios metilados. Destes, 20,19% foram associados com o volume, 78,21% com a altura e 39,42% com DAP (Figura 35-B). Contudo, o número de sítios associados com os três caracteres fenotípicos foi maior neste tecido do que o verificado para folhas, 55 comparado aos 27 associados aos três caracteres fenotípicos para folhas. Novamente, conjuntos de sítios metilados especificamente associados com um dos caracteres fenotípicos foram identificados, sendo 60,6% exclusivamente associados com a altura e 19,2% com DAP.

### 8.6.1 ANÁLISE FUNCIONAL DOS SÍTIOS METILADOS ASSOCIADOS COM MEDIDAS DE CRESCIMENTO

A partir dos sítios metilados associados com as medidas de crescimento, foi possível identificar os genes onde esses sítios se localizam. Para verificar se existem grupos de genes envolvidos em processos biológicos correlatos sendo alvo de metilação, foi realizada uma

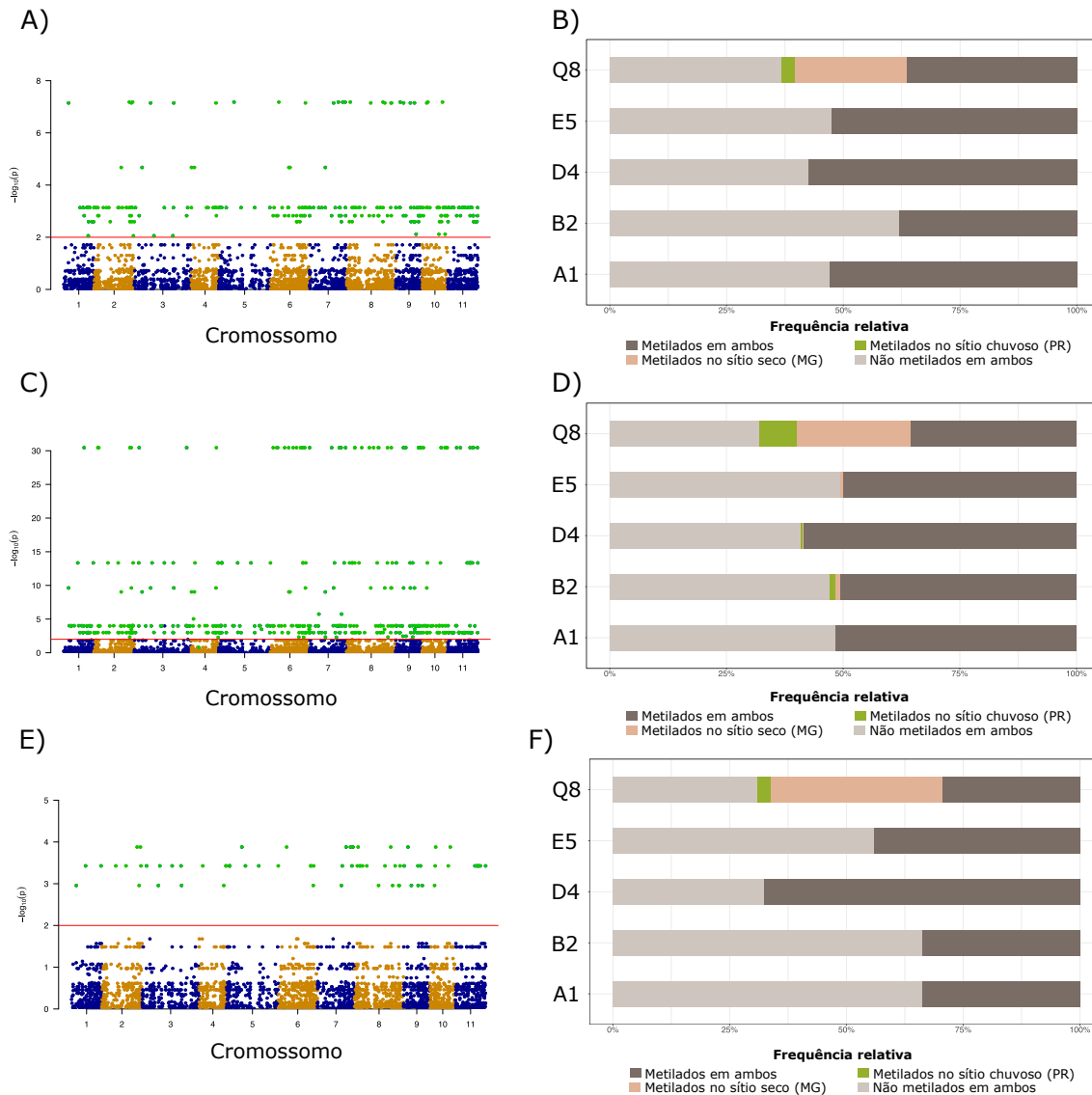


Figura 33 – Sítios metilados significativamente associados com volume (A e B), altura (C e D) e DAP (E e F) considerando o tecido foliar. A, C e E demonstram as distribuições dos sítios significativamente associados ( $p$ -valor  $< 0.01$ ) ao longo dos 11 cromossomos de *E. grandis*. B, D e F apresentam o *status* de metilação dos sítios significativamente associados com os caracteres fenotípicos nos dois ambientes avaliados.

análise de enriquecimento de termos GO usando este conjunto de genes considerando as amostras separadamente.

Alguns termos foram considerados enriquecidos para o conjunto de genes contendo as marcas significativas em ambos as amostras. Para amostras de folha, os termos atividade de hidrolases e GTPases (GO:0043547, GO:0051345 e GO:0043087), regulação positiva da função molecular (GO:0044093) e histogênese do floema ou xilema (GO:0010087) aparecem como enriquecidos no conjunto de genes contendo metilações associadas a volume. Já os termos *Golgi vesicle budding* e *Golgi vesicle transport* (GO:0048193 e GO:0048194) foram enriquecidos para altura. Nenhum termo foi enriquecido no conjunto de genes contendo os

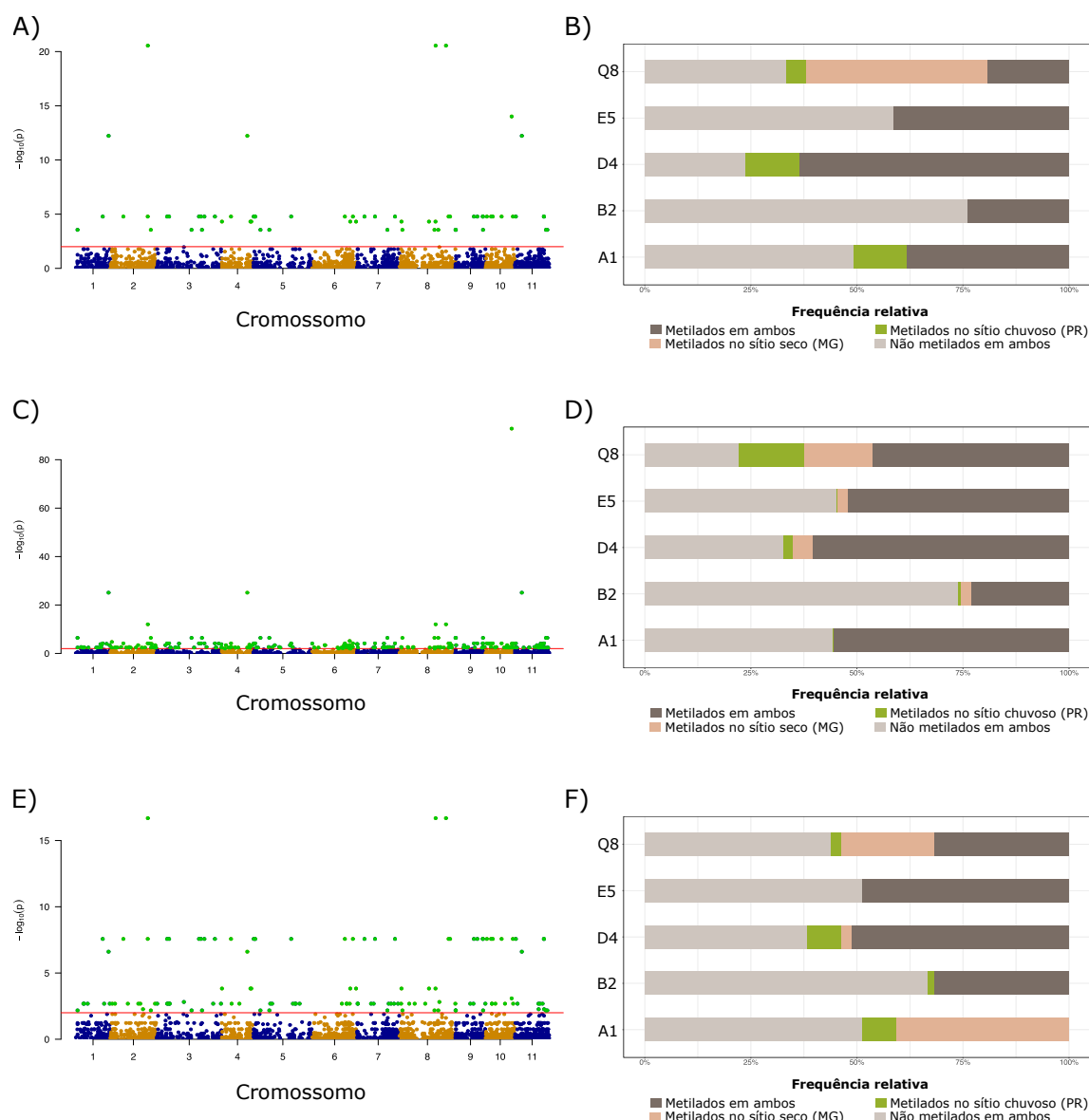


Figura 34 – Sítios metilados significativamente associados com volume (A e B), altura (C e D) e DAP (E e F) considerando o tecido foliar. A, C e E demonstram as distribuições dos sítios significativamente associados ( $p$ -valor  $< 0.01$ ) ao longo dos 11 cromossomos de *E. grandis*. B, D e F apresentam o *status* de metilação dos sítios significativamente associados com os caracteres fenotípicos nos dois ambientes avaliados.

sítios metilados associados com DAP.

Considerando os dados para xilema, um termo enriquecido, *cytoskeletal protein binding* (GO:0008092), foi verificado no conjunto de genes contendo os sítios metilados associados aos três caracteres fenotípicos. Um termo, *helicase activity* (GO:0004386), foi associado à altura. A análise foi também realizada para termos associados unicamente à altura, que representam mais de 60% do conjunto total, e três termos foram enriquecidos para esse conjunto de genes, *helicase activity* (GO:0004386), *phospholipid-translocating ATPase activity* (GO:0004012) e *phospholipid transporter activity* (GO:0005548).

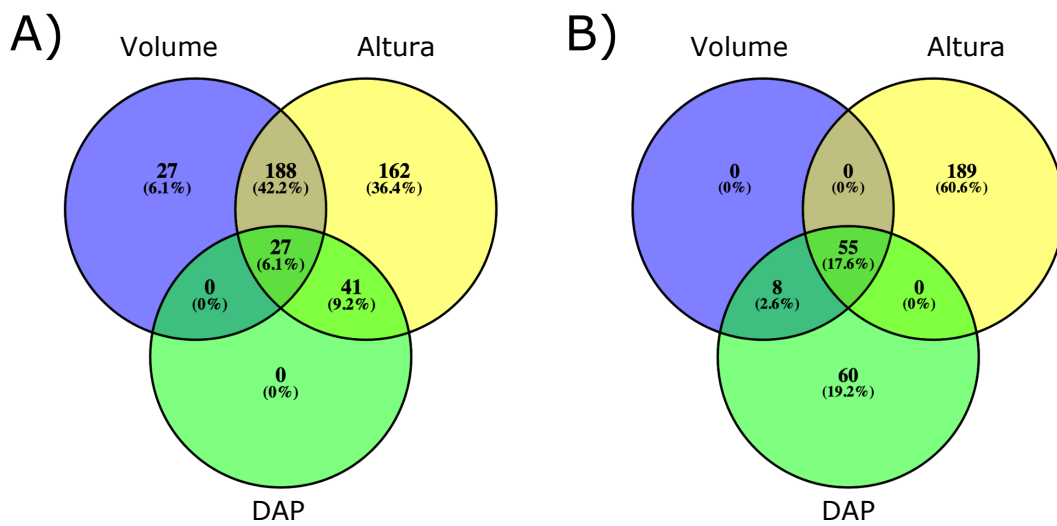


Figura 35 – Diagramas de Venn demonstrando a comparação entre os conjuntos de sítios metilados significativamente associados ( $p$ -valor < 0.01) com os três caracteres fenotípicos avaliados neste estudo (volume, altura e DAP). Em A) são mostrados os conjuntos do tecido foliar e em B, os do xilema.

## 9 DISCUSSÃO

Neste capítulo, buscou-se abordar os seguintes questionamentos:

1. Dada uma mesma planta crescendo em duas localidades com ambientes distintos, existem alterações epigenéticas consistentemente detectáveis?
2. Este conjunto de alterações é consistente entre diferentes indivíduos?
3. Estes padrões epigenéticos podem ser utilizados como indicadores para explicar diferenças fenotípicas?

Para tanto, utilizou-se um desenho experimental delineado no contexto do projeto TECHS (*Tolerance of Eucalyptus Clones to the Hydrous and Thermal Stresses*) (BINKLEY et al., 2017; BINKLEY et al., 2020), que dispunha de clones de árvores do gênero *Eucalyptus*, plantados em ambientes contrastantes quanto ao regime hídrico.

Neste contexto, dispondo de casos onde o mesmo *background* genético foi crescido em duas condições climáticas distintas, buscou-se investigar a emergência de diferentes padrões epigenéticos por clone, em uma amostra de indivíduos não relacionados.

Para tanto, adotou-se a metodologia MS-DArT-seq para investigar metilação de DNA em milhares de sítios genômicos em cinco clones de eucaliptos, plantados em dois ambientes contrastantes. Ao todo, 4.627 sítios metilados foram identificados, incluindo conjuntos de sítios metilados específicos para cada clone e ambiente. Contudo, não foi possível identificar um padrão discriminatório de metilação associado com a variação ambiental e que não esteja interligado com os genótipos dos clones avaliados. Dessa forma, as alterações na metilação observadas utilizando-se a técnica MS-DArT-seq aparentam ser fortemente genótipo dependentes.

Existe um crescente acúmulo de evidências demonstrando que o *background* genético é um dos principais fatores direcionando a variação epigenética (STUART et al., 2016; VIDALIS et al., 2016; UNDERWOOD; HENDERSON; MARTIENSSEN, 2017). Especialmente na espécie de crescimento anual *Arabidopsis thaliana*, a variação epigenética aparenta seguir padrões genéticos subjacentes (DUBIN et al., 2015; KAWAKATSU et al., 2016; MENG et al., 2016). Contudo, essas observações não podem ser generalizadas para outras espécies, uma vez que tanto o genoma quanto a estrutura epigenética de *Arabidopsis* são considerados atípicos dentre as espécies vegetais (MIROUZE; VITTE, 2014; ALONSO et al., 2015; VIDALIS et al., 2016; RICHARDS et al., 2017b). Por exemplo, *Arabidopsis thaliana* possui um genoma pequeno e com menor proporção de elementos transponíveis

do que demais organismos modelo como medicago, soja e arroz, dentre outras espécies investigadas.

Além disso, ocorrências de variação epigenética independente de variação genética têm sido reportadas em espécies cultiváveis como milho (XU et al., 2019) e soja (SHEN et al., 2018) e em espécies florestais como *Quercus lobata* (GUGGER et al., 2016) e *Populus balsamifera* (CHAMPIGNY et al., 2020). Dessa forma, a relevância dos fatores epigenéticos sobre a variação fenotípica deve ser investigada para cada espécie, buscando distinguir-se a participação dos componentes genéticos e epigenéticos.

Como esperado, todos os cinco clones investigados apresentaram produtividade consideravelmente menor no ambiente seco em comparação com o chuvoso. Adicionalmente, as medidas dos caracteres relacionados ao crescimento revelaram significativa diferença no desempenho dos clones investigados de acordo com o ambiente (Figura 24), evidenciando os diferentes níveis de plasticidade fenotípica desses clones e a forte interação com os diferentes ambientes. Portanto, os clones selecionados representam material biológico adequado para investigar os padrões gerais de alterações epigenéticas e, possivelmente, identificar sítios metilados associados com os caracteres fenotípicos relacionados ao crescimento.

Com a aplicação de análises multivariadas para testar a associação entre os sítios metilados e as medidas dos caracteres fenotípicos relacionados ao crescimento, foram identificados 445 sítios metilados associados com ao menos um dos caracteres e dispersos ao longo dos 11 cromossomos do genoma de referência. É importante ressaltar que os caracteres fenotípicos mensurados neste estudo são fortemente correlacionados ( $\rho > 0,8$  no ambiente seco e  $\rho > 0,65$  no ambiente chuvoso) podendo enviesar as análises estatísticas. Neste caso, esperaria-se que a maioria dos sítios significativos fossem associados aos três caracteres fenotípicos. Contudo, uma grande parte dos 445 sítios significativos foi associado apenas a um dos caracteres fenotípicos mensurados. Por exemplo, 36,4% dos sítios metilados foram associados unicamente à altura considerando-se as amostras de folhas e 60,6%, considerando o xilema.

Não foi possível identificar grupos de metilações ambiente específicos (Figura 30). Igualmente, não se estabeleceu padrões bem definidos entre metilação e os níveis de produtividade dos clones avaliados. Essa ausência de padrões claros é consistente com o observado em estudos aplicando GWAS (*genome-wide association study*) e predição genômica para a investigação de caracteres complexos relacionados ao crescimento. As evidências acumuladas sugerem que esses caracteres são controlados por um grande número de locos genômicos com pequeno efeito sobre o caractere (RESENDE et al., 2017; GRATAPAGLIA et al., 2018; MÜLLER et al., 2019). Se os fatores epigenéticos apresentam características similares ao observado para os genéticos, espera-se que um grande número de sítios de metilação com pequeno efeito também sejam associados aos diferentes caracteres, dificultando sua identificação.



Avaliando as metilações verificadas em relação à presença em genes, foi possível observar que os conjuntos de genes metilados nos clones mais produtivos no ambiente seco (A1 e E5) apresentaram enriquecimento de termos GO relacionados com processos de sinalização e resposta a estímulos. A taxa de crescimento superior desses clones nesse ambiente talvez indique um possível cenário onde a modulação epigenética influencie os mecanismos de sinalização e resposta para superar o ambiente hostil, ao invés de atuar diretamente nas vias metabólicas responsáveis pelo crescimento e produção de madeira. O papel fundamental dos mecanismos de percepção e sinalização na resposta das plantas à variação ambiental foi, recentemente, revisado por [Bigot et al. \(2018\)](#). Igualmente, [Kawakatsu et al. \(2016\)](#) revisaram o papel da epigenética na resposta ao estresse e nos processos de adaptação.

No melhoramento de espécies florestais, e, particularmente, no melhoramento de eucaliptos, as interações genótipo-ambiente são um importante componente das práticas de melhoramento. O entendimento da influência epigenética nas interações genótipo-ambiente, bem como a capacidade de seleção de epialelos envolvidos, pode beneficiar os programas de melhoramento de cultivares e espécies florestais, onde e quando é interessante reduzir esses efeitos, visando a geração de genótipos que sejam capazes de manter uma alta produtividade em diferentes ambientes contrastantes ([SPRINGER; SCHMITZ, 2017](#); [KAWAKATSU; ECKER, 2019](#); [LATUTRIE; GOURCILLEAU; PUJOL, 2019](#)). Contudo, os resultados obtidos aqui mostram que uma abordagem metodológica com maior amostragem, como o sequenciamento completo do epigenoma, associada à avaliação de um maior número de indivíduos, deve ser necessária para capturar tais epialelos de pequeno efeito. O que torna a estratégia pouco viável em virtude dos custos.

Neste estudo, os clones investigados são híbridos interespecíficos de *Eucalyptus grandis* e *Eucalyptus urophylla*. Em híbridos de primeira geração (F1), frequentemente, são observados o aumento da heterozigose e presença de vigor híbrido ([GLOMBIK et al., 2020](#)). Diversos fenômenos podem ocorrer como consequência da hibridização, incluindo rearranjos genômicos, reativação e/ou amplificação de elementos repetitivos, modificação nos padrões de expressão genômica, interações intra-locus entre alelos, dominância, sobredominância e interações epistáticas dos siRNAs e miRNAs ([FRIDMAN, 2015](#); [BASHIR et al., 2018](#); [GLOMBIK et al., 2020](#)). Alterações epigenéticas também decorrem da hibridização e podem afetar os fenótipos e a plasticidade fenotípica. Por exemplo, divergências no conteúdo de pequenos RNAs nos genomas parentais podem criar alterações epigenéticas herdáveis que não são associadas com modificações do DNA ([GLOMBIK et al., 2020](#)). Além disso, regiões genômicas diferencialmente metiladas nos genomas parentais são relacionadas com a presença de heterose nos híbridos, como demonstrado em estudos utilizando linhas puras recombinantes epigeneticamente divergentes (epiRILs; *epigenetic recombinant inbred lines*) ([LAUSS et al., 2018](#)). As epiRILs são linhagens (genótipos quase idênticos) que apresentam divergências epigenéticas causadas por mutações pontuais em regiões genômicas cruciais

para o estabelecimento e/ou manutenção de metilação do DNA.

É importante ressaltar que efeitos genéticos não foram considerados neste estudo. Optou-se por realizar um delineamento experimental que visou isolar potenciais efeitos ambientais dos efeitos genéticos. Para isso, utilizou-se réplicas dos mesmos genótipos nos dois ambientes contrastantes e realizada a comparação direta dos perfis de metilação do DNA para cada genótipo (Figura 27) no intuito de estabelecer a relação entre os padrões de metilações e o ambiente (Figura 30). Adicionalmente, buscou-se verificar a existência de associação entre os sítios metilados e caracteres fenotípicos relacionados ao crescimento.

Os resultados indicam que os perfis epigenéticos, mensurados pelas metilações do DNA detectáveis por MS-DArT-seq, são mais associados com o *background* genético dos clones avaliados do que com as condições ambientais. Dessa forma, é possível especular que a plasticidade fenotípica observada decorre de fatores genéticos, incluindo os que por ventura tenham emergido devido a hibridização interespecífica. Contudo, é importante frisar que este estudo foi baseado em uma técnica de redução genômica e com amostragem enviesada para regiões gênicas. Dessa forma, abordagens com maior poder analítico, como o estudo do metiloma de populações ao invés de indivíduos, são necessários para determinar com exatidão o nível de influência das variações epigenéticas sobre a plasticidade fenotípica.

Outro aspecto relevante relacionado ao uso de híbridos neste estudo é o fato de ter-se disponível apenas o genoma de *E. grandis* como referência durante a etapa de mapeamento das sequências. O uso deste genoma provavelmente cria um viés de seleção, favorecendo regiões herdadas do parental *E. grandis* em detrimento de regiões altamente divergentes herdadas do parental *E. urophylla*, que não mapeiam no genoma de referência. Uma vez que ainda não existe um genoma com alta qualidade de *E. urophylla* disponível, não é possível investigar esse viés em potencial ou a intensidade com que ele afetaria a amostragem de regiões genômicas oriundas de *E. urophylla*. Porém, sugere-se que a seleção preferencial de regiões com sequências altamente conservadas entre as espécies ocasiona uma subamostragem de metilações ocorrendo em alelos específicos do genoma de *E. urophylla* afetando as análises de associação entre sítios metilados e os caracteres fenotípicos relacionados ao crescimento, negligenciando uma fração das metilações do DNA que, potencialmente, respondem as alterações ambientais.

## CONCLUSÃO

Modulação epigenética é frequentemente sugerida como um mecanismo capaz de promover alterações fenotípicas em resposta a ocorrência de estresses bióticos e abióticos. Neste estudo, observou-se que, em clones de *Eucalyptus*, existe uma forte correlação entre a diversidade epigenética e *background* genético.

Uma questão relevante no estudo da contribuição epigenética na variação fenotípica é a necessidade de ampla amostragem dos sítios passíveis de metilação em um grande número de amostras, gerando assim poder estatístico suficiente para determinar mesmo efeitos pequenos. Dessa forma, os achados descritos neste estudo devem ser considerados não como uma resposta definitiva em relação ao efeito da metilação do DNA sobre a plasticidade fenotípica, mas sim como uma indicação de que estudos mais amplos, incluindo maior número amostral, bem como envolvendo uma ampla diversidade de locais e condições ambientais são necessários para o entendimento completo dos aspectos epigenéticos relacionados aos processos adaptativos.

# CONSIDERAÇÕES FINAIS

Os resultados verificados neste trabalho evidenciaram a capacidade de detecção de metilação de DNA em amostras de plantas a partir do sequenciamento MS-DArT-seq, combinado com protocolo computacional analítico desenvolvido no contexto deste estudo. Adicionalmente, a metodologia desenvolvida é capaz de identificar perfis de metilação especificamente associados a genótipos, bem como diferentes amostras vegetais.

Verificou-se ainda que metilações de DNA em clones de *Eucalyptus* plantados em ambientes com condições ambientais variadas possuem metilação de DNA especificamente detectadas em cada ambiente. Contudo, determinou-se que os perfis epigenéticos verificados possuem forte relação com o *background* genético.

O código que constitui o protocolo de análise desenvolvido foi depositado em repositórios de acesso público e está disponível para livre utilização. Dessa forma, espera-se que os resultados descritos auxiliem no entendimento da influência epigenética na resposta às variações ambientais em árvores, especialmente nas espécies do gênero *Eucalyptus*, bem como que os protocolos desenvolvidos possam ser implementados por outros grupos de pesquisa, auxiliando o avanço deste campo de estudo.

## REFERÊNCIAS

ALONSO, C. et al. Global dna cytosine methylation as an evolving trait: phylogenetic signal and correlated evolution with genome size in angiosperms. *Frontiers in Genetics*, Frontiers, v. 6, p. 4, 2015. Citado na página 109.

ALONSO, C.; RAMOS-CRUZ, D.; BECKER, C. The role of plant epigenetics in biotic interactions. *New Phytologist*, Wiley Online Library, v. 221, n. 2, p. 731–737, 2019. Citado na página 36.

ANNACONDA, M. L.; MAGERØY, M. H.; MARTINEZ, G. Stress response regulation by epigenetic mechanisms: changing of the guards. *Physiologia Plantarum*, v. 162, n. 2, p. 239–250, fev. 2018. ISSN 1399-3054. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/ppl.12662/abstract>>. Citado na página 37.

BAHIN, M. et al. Alfa: annotation landscape for aligned reads. *BMC genomics*, BioMed Central, v. 20, n. 1, p. 250, 2019. Citado na página 56.

BAI, F.; SETTLES, A. M. Imprinting in plants as a mechanism to generate seed phenotypic diversity. *Frontiers in Plant Science*, v. 5, 2015. ISSN 1664-462X. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fpls.2014.00780/full>>. Citado 2 vezes nas páginas 33 e 34.

BALL, M. P. et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature biotechnology*, Nature Publishing Group, v. 27, n. 4, p. 361, 2009. Citado 2 vezes nas páginas 49 e 78.

BARROS-SILVA, D. et al. Profiling dna methylation based on next-generation sequencing approaches: new insights and clinical applications. *Genes*, Multidisciplinary Digital Publishing Institute, v. 9, n. 9, p. 429, 2018. Citado na página 47.

BARTHOLOMÉ, J. et al. High-resolution genetic maps of Eucalyptus improve Eucalyptus grandis genome assembly. *The New Phytologist*, v. 206, n. 4, p. 1283–1296, jun. 2015. ISSN 1469-8137. Citado 2 vezes nas páginas 55 e 80.

BASHIR, T. et al. Effect of hybridization on somatic mutations and genomic rearrangements in plants. *International journal of molecular sciences*, Multidisciplinary Digital Publishing Institute, v. 19, n. 12, p. 3758, 2018. Citado na página 111.

BATISTA, R. A.; KÖHLER, C. Genomic imprinting in plants—revisiting existing models. *Genes & Development*, Cold Spring Harbor Lab, v. 34, n. 1-2, p. 24–36, 2020. Citado 2 vezes nas páginas 33 e 34.

BEDNAREK, P. T.; ORŁOWSKA, R.; NIEDZIELA, A. A relative quantitative Methylation-Sensitive Amplified Polymorphism (MSAP) method for the analysis of abiotic stress. *BMC Plant Biology*, v. 17, p. 79, abr. 2017. ISSN 1471-2229. Disponível em: <<https://doi.org/10.1186/s12870-017-1028-0>>. Citado na página 48.

BEWICK, A. J.; SCHMITZ, R. J. Gene body DNA methylation in plants. *Current Opinion in Plant Biology*, v. 36, p. 103–110, abr. 2017. ISSN 1369-5266. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1369526616301297>>. Citado 2 vezes nas páginas 32 e 33.

BEWICK, A. J. et al. Evolutionary and experimental loss of gene body methylation and its consequence to gene expression. *G3: Genes, Genomes, Genetics*, v. 9, n. 8, p. 2441–2445, 2019. Citado na página 33.

BIGOT, S. et al. Pivotal roles of environmental sensing and signaling mechanisms in plant responses to climate change. *Global change biology*, Wiley Online Library, v. 24, n. 12, p. 5573–5589, 2018. Citado na página 111.

BINKLEY, D. et al. The interactions of climate, spacing and genetics on clonal eucalyptus plantations across brazil and uruguay. *Forest Ecology and Management*, Elsevier, v. 405, p. 271–283, 2017. Citado 3 vezes nas páginas 89, 92 e 109.

BINKLEY, D. et al. Variation in whole-rotation yield among eucalyptus genotypes in response to water and heat stresses: The techs project. *Forest Ecology and Management*, Elsevier, v. 462, p. 117953, 2020. Citado 3 vezes nas páginas 89, 92 e 109.

BIRNBAUM, K. D.; ROUDIER, F. Epigenetic memory and cell fate reprogramming in plants. *Regeneration*, v. 4, n. 1, p. 15–20, fev. 2017. ISSN 2052-4412. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/reg2.73/abstract>>. Citado 2 vezes nas páginas 46 e 47.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, Oxford University Press, v. 30, n. 15, p. 2114–2120, 2014. Citado na página 55.

BOUYER, D. et al. DNA methylation dynamics during early plant life. *Genome Biology*, v. 18, p. 179, set. 2017. ISSN 1474-760X. Disponível em: <<https://doi.org/10.1186/s13059-017-1313-0>>. Citado na página 28.

BRÄUTIGAM, K. et al. Sexual epigenetics: gender-specific methylation of a gene in the sex determining region of populus balsamifera. *Scientific reports*, Nature Publishing Group, v. 7, n. 1, p. 1–8, 2017. Citado na página 39.

BRÄUTIGAM, K. et al. Epigenetic regulation of adaptive responses of forest tree species to the environment. *Ecology and evolution*, Wiley Online Library, v. 3, n. 2, p. 399–415, 2013. Citado na página 88.

BRINKMAN, A. B. et al. Whole-genome dna methylation profiling using methylcap-seq. *Methods*, Elsevier, v. 52, n. 3, p. 232–236, 2010. Citado na página 47.

BROECK, A. V. et al. Variability in dna methylation and generational plasticity in the lombardy poplar, a single genotype worldwide distributed since the eighteenth century. *Frontiers in plant science*, Frontiers, v. 9, p. 1635, 2018. Citado na página 88.

BROWN, K. et al. Identification and functional evaluation of accessible chromatin associated with wood formation in eucalyptus grandis. *New Phytologist*, Wiley Online Library, v. 223, n. 4, p. 1937–1951, 2019. Citado na página 88.

BRUNNER, A. L. et al. Distinct dna methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome research*, Cold Spring Harbor Lab, v. 19, n. 6, p. 1044–1056, 2009. Citado na página 49.

CARNEROS, E. et al. The epigenetic memory of temperature during embryogenesis modifies the expression of bud burst-related genes in norway spruce epitypes. *Planta*, Springer, v. 246, n. 3, p. 553–566, 2017. Citado na página 38.

CAYE, K. et al. Lfmm 2: fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular biology and evolution*, Oxford University Press, v. 36, n. 4, p. 852–860, 2019. Citado na página 93.

CHAMPIGNY, M. J. et al. Learning from methylomes: epigenomic correlates of populus balsamifera traits based on deep learning models of natural dna methylation. *Plant Biotechnology Journal*, Wiley-Blackwell, v. 18, n. 6, p. 1361, 2020. Citado 3 vezes nas páginas 39, 88 e 110.

CHANG, Y.-N. et al. Epigenetic regulation in plant abiotic stress responses. *Journal of Integrative Plant Biology*, Wiley Online Library, v. 62, n. 5, p. 563–580, 2020. Citado na página 36.

CHEN, L. et al. Direct identification of the active-site nucleophile in a dna (cytosine-5)-methyltransferase. *Biochemistry*, ACS Publications, v. 30, n. 46, p. 11018–11025, 1991. Citado na página 28.

CHOI, C.-S.; SANO, H. Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. *Molecular Genetics and Genomics*, Springer, v. 277, n. 5, p. 589–600, 2007. Citado na página 37.

CHOW, H. T.; CHAKRABORTY, T.; MOSHER, R. A. Rna-directed dna methylation and sexual reproduction: Expanding beyond the seed. *Current opinion in plant biology*, Elsevier, v. 54, p. 11–17, 2020. Citado na página 32.

CHWIALKOWSKA, K. et al. Methylation Sensitive Amplification Polymorphism Sequencing (MSAP-Seq)—A Method for High-Throughput Analysis of Differentially Methylated CCGG Sites in Plants with Large Genomes. *Frontiers in Plant Science*, v. 8, 2017. ISSN 1664-462X. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fpls.2017.02056/full>>. Citado 2 vezes nas páginas 81 e 83.

CI, D. et al. Variation in genomic methylation in natural populations of populus simonii is associated with leaf shape and photosynthetic traits. *Journal of experimental botany*, Oxford University Press, v. 67, n. 3, p. 723–737, 2016. Citado 2 vezes nas páginas 39 e 88.

CORREIA, B. et al. Depicting how eucalyptus globulus survives drought: involvement of redox and dna methylation events. *Functional Plant Biology*, CSIRO, v. 43, n. 9, p. 838–850, 2016. Citado na página 88.

CORTIJO, S. et al. Mapping the epigenetic basis of complex traits. *Science*, American Association for the Advancement of Science, v. 343, n. 6175, p. 1145–1148, 2014. Citado na página 87.

COURTOIS, B. et al. Genome-wide association mapping of root traits in a japonica rice panel. *PLoS one*, Public Library of Science, v. 8, n. 11, 2013. Citado na página 50.

DEICHMANN, U. Epigenetics: The origins and evolution of a fashionable topic. *Developmental Biology*, v. 416, n. 1, p. 249–254, ago. 2016. ISSN 0012-1606. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0012160616302974>>. Citado 3 vezes nas páginas 26, 28 e 30.

DELERIS, A.; HALTER, T.; NAVARRO, L. DNA Methylation and Demethylation in Plant Immunity. *Annual Review of Phytopathology*, v. 54, n. 1, p. 579–603, 2016. Disponível em: <<https://doi.org/10.1146/annurev-phyto-080615-100308>>. Citado na página 29.

DIEZ, C. M.; ROESSLER, K.; GAUT, B. S. Epigenetics and plant genome evolution. *Current Opinion in Plant Biology*, v. 18, p. 1–8, abr. 2014. ISSN 1369-5266. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1369526613001969>>. Citado 2 vezes nas páginas 26 e 27.

DOWEN, R. H. et al. Widespread dynamic dna methylation in response to biotic stress. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 109, n. 32, p. E2183–E2191, 2012. Citado 2 vezes nas páginas 36 e 37.

DOWN, T. A. et al. A bayesian deconvolution strategy for immunoprecipitation-based dna methylome analysis. *Nature biotechnology*, Nature Publishing Group, v. 26, n. 7, p. 779–785, 2008. Citado na página 47.

DOYLE, J. J.; DOYLE, J. L. *A rapid DNA isolation procedure for small quantities of fresh leaf tissue*. [S.l.], 1987. Citado na página 54.

DUBIN, M. J. et al. Dna methylation in arabidopsis has a genetic basis and shows evidence of local adaptation. *elife*, eLife Sciences Publications Limited, v. 4, p. e05255, 2015. Citado 2 vezes nas páginas 88 e 109.

ECKER, S. et al. Epigenetic and transcriptional variability shape phenotypic plasticity. *Bioessays*, Wiley Online Library, v. 40, n. 2, p. 1700148, 2018. Citado na página 87.

FEIL, R.; FRAGA, M. F. Epigenetics and the environment: emerging patterns and implications. *Nature reviews genetics*, Nature Publishing Group, v. 13, n. 2, p. 97–109, 2012. Citado na página 88.

FELSENFELD, G. A Brief History of Epigenetics. *Cold Spring Harbor Perspectives in Biology*, v. 6, n. 1, p. a018200, jan. 2014. ISSN , 1943-0264. Disponível em: <<http://cshperspectives.cshlp.org/content/6/1/a018200>>. Citado 2 vezes nas páginas 26 e 47.

FERRÉ, Q. et al. *OLOGRAM: Determining significance of total overlap length between genomic regions sets*. [S.l.]: Oxford University Press, 2020. Citado na página 56.

FERREIRA, L. J. et al. Salt tolerant and sensitive rice varieties display differential methylome flexibility under salt stress. *PLoS One*, Public Library of Science, v. 10, n. 5, p. e0124060, 2015. Citado na página 37.



- FRIDMAN, E. Consequences of hybridization and heterozygosity on plant vigor and phenotypic stability. *Plant Science*, Elsevier, v. 232, p. 35–40, 2015. Citado na página 111.
- FULNEČEK, J.; KOVAŘÍK, A. How to interpret Methylation Sensitive Amplified Polymorphism (MSAP) profiles? *BMC Genetics*, v. 15, p. 2, jan. 2014. ISSN 1471-2156. Disponível em: <<https://doi.org/10.1186/1471-2156-15-2>>. Citado 3 vezes nas páginas 48, 78 e 80.
- GALLUSCI, P. et al. Epigenetics for Plant Improvement: Current Knowledge and Modeling Avenues. *Trends in Plant Science*, v. 22, n. 7, p. 610–623, jul. 2017. ISSN 1360-1385. Disponível em: <[http://www.cell.com/trends/plant-science/abstract/S1360-1385\(17\)30089-4](http://www.cell.com/trends/plant-science/abstract/S1360-1385(17)30089-4)>. Citado na página 30.
- GANGULY, D. R. et al. The arabidopsis dna methylome is stable under transgenerational drought stress. *Plant physiology*, Am Soc Plant Biol, v. 175, n. 4, p. 1893–1912, 2017. Citado na página 88.
- GLOMBIK, M. et al. Competition of parental genomes in plant hybrids. *Frontiers in Plant Science*, Frontiers Media SA, v. 11, 2020. Citado na página 111.
- GOODSTEIN, D. M. et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, v. 40, n. D1, p. D1178–D1186, jan. 2012. ISSN 0305-1048. Disponível em: <<https://academic.oup.com/nar/article/40/D1/D1178/2903577/Phytozome-a-comparative-platform-for-green-plant>>. Citado na página 55.
- GOUIL, Q.; BAULCOMBE, D. C. DNA Methylation Signatures of the Plant Chromomethyltransferases. *PLOS Genetics*, v. 12, n. 12, p. e1006526, dez. 2016. ISSN 1553-7404. Disponível em: <<http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006526>>. Citado na página 29.
- GOURCILLEAU, D. et al. Dna methylation and histone acetylation: genotypic variations in hybrid poplars, impact of water deficit and relationships with productivity. *Annals of Forest Science*, EDP Sciences, v. 67, n. 2, p. 208, 2010. Citado 2 vezes nas páginas 38 e 88.
- GRATTAPAGLIA, D. et al. Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science*, Frontiers, v. 9, p. 1693, 2018. Citado na página 110.
- GREALLY, J. M. Population epigenetics. *Current Opinion in Systems Biology*, v. 1, p. 84–89, fev. 2017. ISSN 2452-3100. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2452310017300239>>. Citado na página 79.
- GREALLY, J. M. Comments and Opinion, *A user's guide to the ambiguous word 'epigenetics'*. 2018. Disponível em: <<https://www-nature.ez54.periodicos.capes.gov.br/articles/nrm.2017.135>>. Citado na página 26.
- GREAVES, I. K. et al. Epigenetic Changes in Hybrids. *Plant Physiology*, v. 168, n. 4, p. 1197–1205, ago. 2015. ISSN 0032-0889, 1532-2548. Disponível em: <<http://www.plantphysiol.org/content/168/4/1197>>. Citado na página 35.

- GROSZMANN, M. et al. The role of epigenetics in hybrid vigour. *Trends in Genetics*, v. 29, n. 12, p. 684–690, dez. 2013. ISSN 0168-9525. Disponível em: <[http://www.cell.com/trends/genetics/abstract/S0168-9525\(13\)00125-X](http://www.cell.com/trends/genetics/abstract/S0168-9525(13)00125-X)>. Citado na página 35.
- GU, H. et al. Preparation of reduced representation bisulfite sequencing libraries for genome-scale dna methylation profiling. *Nature protocols*, Nature Publishing Group, v. 6, n. 4, p. 468, 2011. Citado na página 46.
- GUGGER, P. F. et al. Species-wide patterns of dna methylation variation in quercus lobata and their association with climate gradients. *Molecular Ecology*, Wiley Online Library, v. 25, n. 8, p. 1665–1680, 2016. Citado 2 vezes nas páginas 88 e 110.
- GUO, H. et al. The methylation patterns and transcriptional responses to chilling stress at the seedling stage in rice. *International journal of molecular sciences*, Multidisciplinary Digital Publishing Institute, v. 20, n. 20, p. 5089, 2019. Citado na página 37.
- GÖMÖRY, D. et al. Epigenetic memory effects in forest trees: a victory of “Michurinian biology”? *Central European Forestry Journal*, v. 63, n. 4, p. 173–179, 2017. Disponível em: <<https://www.degruyter.com/view/j/forj.2017.63.issue-4/forj-2017-0024/forj-2017-0024.xml>>. Citado na página 38.
- HERMAN, J. J. et al. How stable ‘should’ epigenetic modifications be? insights from adaptive plasticity and bet hedging. *Evolution*, Wiley Online Library, v. 68, n. 3, p. 632–643, 2014. Citado na página 87.
- HOCHHOLDINGER, F.; BALDAUF, J. A. Heterosis in plants. *Current Biology*, Elsevier, v. 28, n. 18, p. R1089–R1092, 2018. Citado na página 35.
- HU, Y. et al. Cold stress selectively unsilences tandem repeats in heterochromatin associated with accumulation of h3k9ac. *Plant, cell & environment*, Wiley Online Library, v. 35, n. 12, p. 2130–2142, 2012. Citado na página 37.
- HUANG, X. et al. Rapid response to changing environments during biological invasions: Dna methylation perspectives. *Molecular ecology*, Wiley Online Library, v. 26, n. 23, p. 6621–6633, 2017. Citado na página 87.
- HUSSON, F.; LÊ, S.; PAGÈS, J. *Exploratory multivariate analysis by example using R*. [S.l.]: CRC press, 2017. Citado na página 102.
- IWASAKI, M.; PASZKOWSKI, J. Epigenetic memory in plants. *The EMBO Journal*, v. 33, n. 18, p. 1987–1998, set. 2014. ISSN 0261-4189, 1460-2075. Disponível em: <<http://emboj.embopress.org/content/33/18/1987>>. Citado na página 26.
- JIANG, C. et al. Environmentally responsive genome-wide accumulation of de novo arabidopsis thaliana mutations and epimutations. *Genome research*, Cold Spring Harbor Lab, v. 24, n. 11, p. 1821–1829, 2014. Citado na página 88.
- JOHANNES, F.; SCHMITZ, R. J. Spontaneous epimutations in plants. *New Phytologist*, Wiley Online Library, v. 221, n. 3, p. 1253–1259, 2019. Citado 3 vezes nas páginas 28, 29 e 32.

JOHNSEN, Ø.; DIETRICHSON, J.; SKARET, G. Phenotypic changes in progenies of northern clones of *Picea abies* (L.) Karst. grown in a southern seed orchard: III. climatic damage and growth in a progeny trial. *Scandinavian Journal of Forest Research*, Taylor & Francis, v. 4, n. 1-4, p. 343–350, 1989. Citado na página 38.

KAWAKATSU, T.; ECKER, J. R. Diversity and dynamics of DNA methylation: epigenomic resources and tools for crop breeding. *Breeding science*, v. 69, n. 2, p. 191–204, 2019. Citado na página 111.

KAWAKATSU, T. et al. Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. *Cell*, v. 166, n. 2, p. 492–505, jul. 2016. ISSN 0092-8674, 1097-4172. Disponível em: <[http://www.cell.com/cell/abstract/S0092-8674\(16\)30852-2](http://www.cell.com/cell/abstract/S0092-8674(16)30852-2)>. Citado 4 vezes nas páginas 33, 88, 109 e 111.

KELLER, T. E.; LASKY, J. R.; YI, S. V. The multivariate association between genomewide DNA methylation and climate across the range of *Arabidopsis thaliana*. *Molecular Ecology*, Wiley Online Library, v. 25, n. 8, p. 1823–1837, 2016. Citado na página 88.

KERMICLE, J. Dependence of the r-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics*, Genetics Society of America, v. 66, n. 1, p. 69, 1970. Citado na página 33.

KILIAN, A. et al. Diversity arrays technology: a generic genome profiling technology on open platforms. In: *Data production and analysis in population genomics*. [S.l.]: Springer, 2012. p. 67–89. Citado na página 50.

KIM, K. D.; BAIDOURI, M. E.; JACKSON, S. A. Accessing epigenetic variation in the plant methylome. *Briefings in Functional Genomics*, Oxford University Press, v. 13, n. 4, p. 318–327, 2014. Citado 2 vezes nas páginas 45 e 47.

KOOKE, R. et al. Epigenetic basis of morphological variation and phenotypic plasticity in *Arabidopsis thaliana*. *The Plant Cell*, Am Soc Plant Biol, v. 27, n. 2, p. 337–348, 2015. Citado na página 88.

KRUEGER, F.; ANDREWS, S. R. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics*, Oxford University Press, v. 27, n. 11, p. 1571–1572, 2011. Citado na página 60.

KURDYUKOV, S.; BULLOCK, M. DNA methylation analysis: choosing the right method. *Biology*, Multidisciplinary Digital Publishing Institute, v. 5, n. 1, p. 3, 2016. Citado na página 45.

KÖSTER, J.; RAHMANN, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, v. 28, n. 19, p. 2520–2522, out. 2012. ISSN 1367-4803. Disponível em: <<https://academic.oup.com/bioinformatics/article/28/19/2520/290322/Snakemake-a-scalable-bioinformatics-workflow>>. Citado 2 vezes nas páginas 61 e 94.

LAFON-PLACETTE, C. et al. Changes in the epigenome and transcriptome of the poplar shoot apical meristem in response to water availability affect preferentially hormone pathways. *Journal of experimental botany*, Oxford University Press UK, v. 69, n. 3, p. 537–551, 2018. Citado na página 88.

- LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, v. 9, n. 4, p. 357–359, abr. 2012. ISSN 1548-7091. Disponível em: <<http://www.nature.com/nmeth/journal/v9/n4/full/nmeth.1923.html?foxtrotcallback=true>>. Citado na página 55.
- LATUTRIE, M.; GOURCILLEAU, D.; PUJOL, B. Epigenetic variation for agronomic improvement: an opportunity for vegetatively propagated crops. *American journal of botany*, Wiley-Blackwell, v. 106, n. 10, p. 1281, 2019. Citado na página 111.
- LATZEL, V. et al. Epigenetic diversity increases the productivity and stability of plant populations. *Nature communications*, Nature Publishing Group, v. 4, n. 1, p. 1–7, 2013. Citado na página 88.
- LAUSS, K. et al. Parental dna methylation states are associated with heterosis in epigenetic hybrids. *Plant Physiology*, Am Soc Plant Biol, v. 176, n. 2, p. 1627–1645, 2018. Citado 2 vezes nas páginas 36 e 111.
- LÊ, S. et al. Factominer: an r package for multivariate analysis. *Journal of statistical software*, Los Angeles, v. 25, n. 1, p. 1–18, 2008. Citado na página 93.
- LE, T.-N. et al. Dna demethylases target promoter transposable elements to positively regulate stress responsive genes in arabidopsis. *Genome biology*, Springer, v. 15, n. 9, p. 458, 2014. Citado na página 37.
- LEE, K.; SEO, P. J. Dynamic Epigenetic Changes during Plant Regeneration. *Trends in Plant Science*, v. 0, n. 0, jan. 2018. ISSN 1360-1385. Disponível em: <[http://www.cell.com/trends/plant-science/abstract/S1360-1385\(17\)30274-1](http://www.cell.com/trends/plant-science/abstract/S1360-1385(17)30274-1)>. Citado na página 34.
- LI, N. et al. Whole genome dna methylation analysis based on high throughput sequencing technology. *Methods*, Elsevier, v. 52, n. 3, p. 203–212, 2010. Citado na página 47.
- LI, Y.; KUMAR, S.; QIAN, W. Active DNA demethylation: mechanism and role in plant development. *Plant Cell Reports*, v. 37, n. 1, p. 77–85, jan. 2018. ISSN 0721-7714, 1432-203X. Disponível em: <<https://link.springer.com/article/10.1007/s00299-017-2215-z>>. Citado na página 32.
- LIAO, Y.; SMYTH, G. K.; SHI, W. featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, Oxford University Press, v. 30, n. 7, p. 923–930, 2014. Citado na página 56.
- LINDSAY, S.; BIRD, A. P. Use of restriction enzymes to detect potential gene sequences in mammalian dna. *Nature*, Nature Publishing Group, v. 327, n. 6120, p. 336–338, 1987. Citado na página 47.
- LISTER, R.; ECKER, J. R. Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome research*, Cold Spring Harbor Lab, v. 19, n. 6, p. 959–966, 2009. Citado na página 46.
- LIU, R.; LANG, Z. The mechanism and function of active dna demethylation in plants. *Journal of Integrative Plant Biology*, Wiley Online Library, v. 62, n. 1, p. 148–159, 2020. Citado na página 32.

- LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, v. 15, p. 550, dez. 2014. ISSN 1474-760X. Disponível em: <<https://doi.org/10.1186/s13059-014-0550-8>>. Citado 2 vezes nas páginas 57 e 68.
- LYKO, F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*, v. 19, n. 2, p. 81, fev. 2018. ISSN 1471-0064. Disponível em: <<https://www.nature.com/articles/nrg.2017.80>>. Citado na página 30.
- MARK, V. C. V.; KILIAN, A.; DIERIG, D. A. Development of dart marker platforms and genetic diversity assessment of the us collection of the new oilseed crop lesquerella and related species. *PLoS One*, Public Library of Science, v. 8, n. 5, 2013. Citado na página 50.
- MARSH, A. G.; COTTRELL, M. T.; GOLDMAN, M. F. Epigenetic DNA Methylation Profiling with MSRE: A Quantitative NGS Approach Using a Parkinson's Disease Test Case. *Frontiers in Genetics*, v. 7, 2016. ISSN 1664-8021. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fgene.2016.00191/full>>. Citado na página 83.
- MATZKE, M. A.; MOSHER, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics*, v. 15, n. 6, p. 394, jun. 2014. ISSN 1471-0064. Disponível em: <<https://www.nature.com/articles/nrg3683>>. Citado na página 31.
- MAUNAKEA, A. K. et al. Conserved role of intragenic dna methylation in regulating alternative promoters. *Nature*, Nature Publishing Group, v. 466, n. 7303, p. 253–257, 2010. Citado 2 vezes nas páginas 49 e 78.
- MAURER-ALCALÁ, X. X.; KATZ, L. A. An epigenetic toolkit allows for diverse genome architectures in eukaryotes. *Current Opinion in Genetics & Development*, v. 35, n. Supplement C, p. 93–99, dez. 2015. ISSN 0959-437X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0959437X15001045>>. Citado na página 27.
- MAYNE, B. T. et al. msgbsR : An R package for analysing methylation-sensitive restriction enzyme sequencing data. *Scientific Reports*, v. 8, n. 1, p. 2190, fev. 2018. ISSN 2045-2322. Disponível em: <<https://www.nature.com/articles/s41598-018-19655-w>>. Citado na página 83.
- MCCLELLAND, M.; NELSON, M.; RASCHKE, E. Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Research*, v. 22, n. 17, p. 3640–3659, set. 1994. ISSN 0305-1048. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308336/>>. Citado na página 80.
- MENG, D. et al. Limited contribution of dna methylation variation to expression regulation in arabidopsis thaliana. *PLoS genetics*, Public Library of Science, v. 12, n. 7, 2016. Citado na página 109.
- MIROUZE, M.; VITTE, C. Transposable elements, a treasure trove to decipher epigenetic variation: insights from arabidopsis and crop epigenomes. *Journal of Experimental Botany*, Oxford University Press UK, v. 65, n. 10, p. 2801–2812, 2014. Citado na página 109.

- MÜLLER, B. S. et al. Independent and joint-gwas for growth traits in eucalyptus by assembling genome-wide data for 3373 individuals across four breeding populations. *New Phytologist*, Wiley Online Library, v. 221, n. 2, p. 818–833, 2019. Citado na página 110.
- MYBURG, A. A. et al. The genome of *Eucalyptus grandis*. *Nature*, v. 510, n. 7505, p. 356–362, jun. 2014. ISSN 0028-0836. Disponível em: <<http://www.nature.com/nature/journal/v510/n7505/full/nature13308.html?foxtrotcallback=true>>. Citado 4 vezes nas páginas 54, 78, 79 e 80.
- NIEDERHUTH, C. E. et al. Widespread natural variation of DNA methylation within angiosperms. *Genome Biology*, v. 17, p. 194, set. 2016. ISSN 1474-760X. Disponível em: <<https://doi.org/10.1186/s13059-016-1059-0>>. Citado 2 vezes nas páginas 80 e 82.
- NIEDERHUTH, C. E.; SCHMITZ, R. J. Covering Your Bases: Inheritance of DNA Methylation in Plant Genomes. *Molecular Plant*, v. 7, n. 3, p. 472–480, mar. 2014. ISSN 1674-2052. Disponível em: <[http://www.cell.com/molecular-plant/abstract/S1674-2052\(14\)60248-3](http://www.cell.com/molecular-plant/abstract/S1674-2052(14)60248-3)>. Citado na página 29.
- NIEDERHUTH, C. E.; SCHMITZ, R. J. Putting dna methylation in context: from genomes to gene expression in plants. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, Elsevier, v. 1860, n. 1, p. 149–156, 2017. Citado na página 87.
- ODA, M. et al. High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. *Nucleic acids research*, Oxford University Press, v. 37, n. 12, p. 3829–3839, 2009. Citado na página 49.
- ONG-ABDULLAH, M. et al. Loss of *Karma* transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature*, v. 525, n. 7570, p. 533–537, set. 2015. ISSN 1476-4687. Disponível em: <<https://www.nature.com/articles/nature15365>>. Citado 2 vezes nas páginas 30 e 35.
- O'DEA, R. E. et al. The role of non-genetic inheritance in evolutionary rescue: epigenetic buffering, heritable bet hedging and epigenetic traps. *Environmental Epigenetics*, v. 2, n. 1, jan. 2016. Disponível em: <<https://academic.oup.com/eep/article/2/1/dvv014/2464778>>. Citado na página 28.
- PARRILLA-DOBLAS, J. T. et al. Active dna demethylation in plants. *International journal of molecular sciences*, Multidisciplinary Digital Publishing Institute, v. 20, n. 19, p. 4683, 2019. Citado na página 32.
- PECINKA, A.; ABDELSAMAD, A.; VU, G. T. Hidden genetic nature of epigenetic natural variation in plants. *Trends in plant science*, Elsevier, v. 18, n. 11, p. 625–632, 2013. Citado 2 vezes nas páginas 87 e 88.
- PEREIRA, W. J. et al. A cost-effective approach to dna methylation detection by methyl sensitive dart sequencing. *Plos one*, Public Library of Science San Francisco, CA USA, v. 15, n. 6, p. e0233800, 2020. Citado 3 vezes nas páginas 89, 93 e 98.
- PIGLIUCCI, M.; MURREN, C. J.; SCHLICHTING, C. D. Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology*, The Company of Biologists Ltd, v. 209, n. 12, p. 2362–2367, 2006. Citado na página 87.

PIKAARD, C. S.; SCHEID, O. M. Epigenetic Regulation in Plants. *Cold Spring Harbor Perspectives in Biology*, v. 6, n. 12, p. a019315, dez. 2014. ISSN , 1943-0264. Disponível em: <<http://cshperspectives.cshlp.org/content/6/12/a019315>>. Citado 6 vezes nas páginas 27, 28, 29, 30, 31 e 32.

PLATT, A. et al. Genome-wide signature of local adaptation linked to variable c p g methylation in oak populations. *Molecular Ecology*, Wiley Online Library, v. 24, n. 15, p. 3823–3830, 2015. Citado na página 87.

POOTAKHAM, W. et al. Effects of methylation-sensitive enzymes on the enrichment of genic SNPs and the degree of genome complexity reduction in a two-enzyme genotyping-by-sequencing (GBS) approach: a case study in oil palm (<Emphasis Type="Italic">Elaeis guineensis</Emphasis>). *Molecular Breeding*, v. 36, n. 11, p. 154, nov. 2016. ISSN 1380-3743, 1572-9788. Disponível em: <<https://link.springer.com/article/10.1007/s11032-016-0572-x>>. Citado 3 vezes nas páginas 78, 80 e 81.

PROKOPUK, L.; WESTERN, P. S.; STRINGER, J. M. Transgenerational epigenetic inheritance: adaptation through the germline epigenome? *Epigenomics*, v. 7, n. 5, p. 829–846, ago. 2015. ISSN 1750-1911. Disponível em: <<https://www.futuremedicine.com/doi/10.2217/epi.15.36>>. Citado na página 30.

QUINLAN, A. R.; HALL, I. M. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, Oxford University Press, v. 26, n. 6, p. 841–842, 2010. Citado na página 55.

RAJ, S. et al. Clone history shapes populus drought responses. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 108, n. 30, p. 12521–12526, 2011. Citado na página 88.

RAZA, M. A. et al. Differential dna methylation and gene expression in reciprocal hybrids between solanum lycopersicum and s. pimpinellifolium. *DNA Research*, Oxford University Press, v. 24, n. 6, p. 597–607, 2017. Citado na página 88.

RESENDE, R. T. et al. Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in eucalyptus. *New Phytologist*, Wiley Online Library, v. 213, n. 3, p. 1287–1300, 2017. Citado na página 110.

RICHARDS, C. L. et al. Ecological plant epigenetics: Evidence from model and non-model species, and the way forward. *Ecology Letters*, v. 20, n. 12, p. 1576–1590, dez. 2017. ISSN 1461-0248. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/ele.12858/abstract>>. Citado na página 30.

RICHARDS, C. L. et al. Ecological plant epigenetics: Evidence from model and non-model species, and the way forward. *Ecology letters*, Wiley Online Library, v. 20, n. 12, p. 1576–1590, 2017. Citado na página 109.

RIVERA, C. M.; REN, B. Mapping human epigenomes. *Cell*, Elsevier, v. 155, n. 1, p. 39–55, 2013. Citado na página 46.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression

- data. *Bioinformatics*, v. 26, n. 1, p. 139–140, jan. 2010. ISSN 1367-4803. Disponível em: <<https://academic.oup.com/bioinformatics/article/26/1/139/182458/edgeR-a-Bioconductor-package-for-differential>>. Citado 2 vezes nas páginas 57 e 68.
- RODRIGUES, J. A.; ZILBERMAN, D. Evolution and function of genomic imprinting in plants. *Genes & Development*, v. 29, n. 24, p. 2517–2531, dez. 2015. ISSN 0890-9369, 1549-5477. Disponível em: <<http://genesdev.cshlp.org/content/29/24/2517>>. Citado 2 vezes nas páginas 33 e 34.
- SÁNCHEZ, A. L. et al. The role of dna (de) methylation in immune responsiveness of arabidopsis. *The Plant Journal*, Wiley Online Library, v. 88, n. 3, p. 361–374, 2016. Citado na página 37.
- SANTOS, A. P.; FERREIRA, L. J.; OLIVEIRA, M. M. Concerted Flexibility of Chromatin Structure, Methyome, and Histone Modifications along with Plant Stress Responses. *Biology*, v. 6, n. 1, p. 3, jan. 2017. Disponível em: <<http://www.mdpi.com/2079-7737/6/1/3>>. Citado na página 37.
- SCHMID, M. W. et al. Contribution of epigenetic variation to adaptation in arabidopsis. *Nature communications*, Nature Publishing Group, v. 9, n. 1, p. 1–12, 2018. Citado 2 vezes nas páginas 87 e 88.
- SCHMITZ, R. J.; LEWIS, Z. A.; GOLL, M. G. Dna methylation: shared and divergent features across eukaryotes. *Trends in Genetics*, Elsevier, v. 35, n. 11, p. 818–827, 2019. Citado na página 28.
- SCHMITZ, R. J. et al. Patterns of population epigenomic diversity. *Nature*, Nature Publishing Group, v. 495, n. 7440, p. 193–198, 2013. Citado na página 88.
- SCHÖNBERGER, B. et al. Site-dependent differences in dna methylation and their impact on plant establishment and phosphorus nutrition in populus trichocarpa. *PloS one*, Public Library of Science, v. 11, n. 12, 2016. Citado na página 88.
- SCHULTZ, M. D. et al. Human body epigenome maps reveal noncanonical dna methylation variation. *Nature*, Nature Publishing Group, v. 523, n. 7559, p. 212–216, 2015. Citado na página 60.
- SECCO, D. et al. Stress induced gene expression drives transient dna methylation changes at adjacent repetitive elements. *elife*, eLife Sciences Publications Limited, v. 4, p. e09343, 2015. Citado 2 vezes nas páginas 37 e 87.
- SEKHON, J. S. Multivariate and propensity score matching software with automated balance optimization: the matching package for r. *Journal of Statistical Software*, Forthcoming, 2008. Citado na página 56.
- SEYMOUR, D. K.; BECKER, C. The causes and consequences of DNA methyome variation in plants. *Current Opinion in Plant Biology*, v. 36, p. 56–63, abr. 2017. ISSN 1369-5266. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1369526616301546>>. Citado 3 vezes nas páginas 27, 28 e 79.
- SHEN, Y. et al. Dna methylation footprints during soybean domestication and improvement. *Genome biology*, BioMed Central, v. 19, n. 1, p. 1–14, 2018. Citado na página 110.



- SONG, Y. et al. Stable methylation of a non-coding rna gene regulates gene expression in response to abiotic stress in populus simonii. *Journal of Experimental Botany*, Oxford University Press UK, v. 67, n. 5, p. 1477–1492, 2016. Citado na página 38.
- SOW, M. D. et al. Narrow-sense heritability and p st estimates of dna methylation in three populus nigra l. populations under contrasting water availability. *Tree Genetics & Genomes*, Springer, v. 14, n. 5, p. 78, 2018. Citado na página 38.
- SPRINGER, N. M.; SCHMITZ, R. J. Exploiting induced and natural epigenetic variation for crop improvement. *Nature Reviews Genetics*, v. 18, n. 9, p. 563, set. 2017. ISSN 1471-0064. Disponível em: <<https://www.nature.com/articles/nrg.2017.45>>. Citado 5 vezes nas páginas 30, 31, 33, 88 e 111.
- STUART, T. et al. Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. *elife*, eLife Sciences Publications Limited, v. 5, p. e20777, 2016. Citado na página 109.
- SU, Y. et al. Single-base-resolution methylomes of populus euphratica reveal the association between dna methylation and salt stress. *Tree Genetics & Genomes*, Springer, v. 14, n. 6, p. 86, 2018. Citado na página 88.
- TAKUNO, S.; GAUT, B. S. Body-methylated genes in arabidopsis thaliana are functionally important and evolve slowly. *Molecular biology and evolution*, Oxford University Press, v. 29, n. 1, p. 219–227, 2012. Citado na página 33.
- UMER, M.; HERCEG, Z. Deciphering the epigenetic code: an overview of dna methylation analysis methods. *Antioxidants & redox signaling*, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 18, n. 15, p. 1972–1986, 2013. Citado na página 45.
- UNDERWOOD, C. J.; HENDERSON, I. R.; MARTIENSSEN, R. A. Genetic and epigenetic variation of transposable elements in arabidopsis. *Current opinion in plant biology*, Elsevier, v. 36, p. 135–141, 2017. Citado na página 109.
- US-CAMAS, R. et al. In vitro culture: an epigenetic challenge for plants. *Plant Cell, Tissue and Organ Culture (PCTOC)*, v. 118, n. 2, p. 187–201, ago. 2014. ISSN 0167-6857, 1573-5044. Disponível em: <<https://link.springer.com/article/10.1007/s11240-014-0482-8>>. Citado na página 34.
- VIDALIS, A. et al. Methylome evolution in plants. *Genome Biology*, v. 17, p. 264, dez. 2016. ISSN 1474-760X. Disponível em: <<https://doi.org/10.1186/s13059-016-1127-5>>. Citado 2 vezes nas páginas 32 e 109.
- WADDINGTON, C. H. The Epigenotype. *International Journal of Epidemiology*, v. 41, n. 1, p. 10–13, fev. 2012. ISSN 0300-5771. Disponível em: <<https://academic.oup.com/ije/article/41/1/10/648029>>. Citado na página 26.
- WALKER, J. et al. Sexual-lineage-specific dna methylation regulates meiosis in arabidopsis. *Nature Genetics*, Nature Publishing Group, v. 50, n. 1, p. 130–137, 2018. Citado na página 28.
- WENDTE, J. M.; PIKAARD, C. S. The RNAs of RNA-directed DNA methylation. *Biochimica Et Biophysica Acta*, v. 1860, n. 1, p. 140–148, 2017. ISSN 0006-3002. Citado na página 31.

- WIBOWO, A. et al. Hyperosmotic stress memory in arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by dna glycosylase activity. *Elife*, eLife Sciences Publications Limited, v. 5, p. e13546, 2016. Citado na página 87.
- XIE, H. et al. Environmental conditions and agronomic practices induce consistent global changes in dna methylation patterns in grapevine (*vitis vinifera* cv shiraz). *bioRxiv*, Cold Spring Harbor Laboratory, p. 127977, 2017. Citado na página 88.
- XIE, H. et al. Global dna methylation patterns can play a role in defining terroir in grapevine (*vitis vinifera* cv. shiraz). *Frontiers in plant science*, Frontiers, v. 8, p. 1860, 2017. Citado na página 88.
- XU, J. et al. Population-level analysis reveals the widespread occurrence and phenotypic consequence of dna methylation variation not tagged by genetic variation in maize. *Genome biology*, Springer, v. 20, n. 1, p. 243, 2019. Citado na página 110.
- YONG, W.-S.; HSU, F.-M.; CHEN, P.-Y. Profiling genome-wide DNA methylation. *Epigenetics & Chromatin*, v. 9, p. 26, jun. 2016. ISSN 1756-8935. Disponível em: <<https://doi.org/10.1186/s13072-016-0075-3>>. Citado 6 vezes nas páginas 28, 29, 30, 45, 46 e 47.
- YU, A. et al. Dynamics and biological relevance of dna demethylation in arabidopsis antibacterial defense. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 110, n. 6, p. 2389–2394, 2013. Citado na página 37.
- ZENG, H.; HE, B.; YI, C. Compilation of modern technologies to map genome-wide cytosine modifications in dna. *ChemBioChem*, Wiley Online Library, v. 20, n. 15, p. 1898–1905, 2019. Citado 2 vezes nas páginas 45 e 47.
- ZHANG, H.; LANG, Z.; ZHU, J.-K. Dynamics and function of dna methylation in plants. *Nature reviews Molecular cell biology*, Nature Publishing Group, v. 19, n. 8, p. 489–506, 2018. Citado 4 vezes nas páginas 30, 31, 32 e 37.
- ZHONG, X. Comparative epigenomics: a powerful tool to understand the evolution of DNA methylation. *New Phytologist*, v. 210, n. 1, p. 76–80, abr. 2016. ISSN 1469-8137. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/nph.13540/abstract>>. Citado 2 vezes nas páginas 27 e 28.
- ZILBERMAN, D. et al. Histone h2a. z and dna methylation are mutually antagonistic chromatin marks. *Nature*, Nature Publishing Group, v. 456, n. 7218, p. 125–129, 2008. Citado na página 33.