



Centro de Investigación Científica de Yucatán, A.C.

Posgrado en Ciencias Biológicas

**ANÁLISIS DE LA EXPRESIÓN DIFERENCIAL DE  
METILTRANSFERASAS DE ADN EN LAS  
VARIANTES SOMACLONALES DE *AGAVE  
ANGUSTIFOLIA* HAW.**

Tesis que presenta

FERNANDO DE JESÚS COLORADO PAÚL

En opción al título de

MAESTRO EN CIENCIAS

(Ciencias Biológicas: Opción Biotecnología)

Mérida, Yucatán, México  
2022

---



*CENTRO DE INVESTIGACIÓN CIENTÍFICA DE YUCATÁN, A. C.*  
*POSGRADO EN CIENCIAS BIOLÓGICAS*



**RECONOCIMIENTO**

Por medio de la presente, hago constar que el trabajo de tesis de **Fernando de Jesús Colorado Paúl** titulado “**Análisis de la expresión diferencial de metiltransferasas de ADN en las variantes somaclonales de *Agave angustifolia* Haw.**” fue realizado en el laboratorio de Epigenética de la **Unidad de Biotecnología** del **Centro de Investigación Científica de Yucatán, A.C.** bajo la dirección de la **Dra. Clelia De la Peña Seaman**, dentro de la opción de Biotecnología, perteneciente al Programa de Posgrado en Ciencias Biológicas de este Centro.

Atentamente



---

Dra. Cecilia Hernández Zepeda  
Directora de Docencia

Mérida, Yucatán, México, a 2 de marzo de 2022



## DECLARACIÓN DE PROPIEDAD

Declaro que la información contenida en la sección de Materiales y Métodos, los Resultados y Discusión de este documento proviene de las actividades de investigación realizadas durante el período que se me asignó para desarrollar mi trabajo de tesis, en las Unidades y Laboratorios del Centro de Investigación Científica de Yucatán, A.C., y que a razón de lo anterior y en contraprestación de los servicios educativos o de apoyo que me fueron brindados, dicha información, en términos de la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, le pertenece patrimonialmente a dicho Centro de Investigación. Por otra parte, en virtud de lo ya manifestado, reconozco que de igual manera los productos intelectuales o desarrollos tecnológicos que deriven o pudieran derivar de lo correspondiente a dicha información, le pertenecen patrimonialmente al Centro de Investigación Científica de Yucatán, A.C., y en el mismo tenor, reconozco que si derivaren de este trabajo productos intelectuales o desarrollos tecnológicos, en lo especial, estos se registrarán en todo caso por lo dispuesto por la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, en el tenor de lo expuesto en la presente Declaración.



---

Fernando de Jesús Colorado Paúl

---

---

## **AGRADECIMIENTOS**

Al Centro de Investigación Científica de Yucatán y la Unidad de Biotecnología por las instalaciones prestadas para el desarrollo de la parte experimental de este trabajo.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca No. 747977.

A mi asesora, Dra. Clelia De la Peña por su paciencia, trato amable, los ánimos y por compartir conmigo sus conocimientos y su gran pasión por la ciencia. Crecí personal y académicamente con su asesoría.

A mi comité tutorial y revisores de tesis: Dra. Elsa Góngora Castillo, por su paciencia, su asesoramiento y múltiples consejos útiles para desarrollar mi trabajo. Dr. José Efraín Ramírez Benitez por su trato amable, asesoramiento, revisión detallada de la tesis y por sus comentarios que ayudaron a mejorar mi documento. Dr. Luis Sáenz Carbonell, por la revisión de mi trabajo escrito y por permitirme trabajar en el laboratorio del que está a cargo para completar la parte experimental de mi proyecto. Dra. Virginia Aurora Herrera Valencia, por formar parte de mi comité y por la revisión de mi tesis.

Agradezco al técnico Eduardo Castillo Castro por su amistad, su apoyo y asesoramiento en el laboratorio de Epigenética durante mi maestría.

A mis compañeros de laboratorio Dr. Alexis Salazar Iribe y MC. Mauricio Andrade Marcial, por su paciencia, amistad, asesoramiento, los ánimos y extensa ayuda para desarrollar mi proyecto.

A todos mis amigos del CICY por su amistad y las aventuras compartidas.

A mi familia por su cariño y apoyo incondicional en todas mis metas.

## **LISTA DE LOS PRODUCTOS GENERADOS**

**2021. Artículo de divulgación:** Fernando de Jesús Colorado Paúl, Dra. Clelia De la Peña. (Febrero 2021). Epigenética y el interruptor de las células. Cakotanú, 3, 20-24.

---

---

## **DEDICATORIA**

**A MI FAMILIA, SON MI APOYO Y MOTIVACIÓN MÁS GRANDE.**

---

---

## ÍNDICE

INTRODUCCIÓN .....	1
CAPÍTULO I .....	3
ANTECEDENTES.....	3
1.1. VARIACIÓN SOMACLONAL.....	3
1.2. EL ALBINISMO EN PLANTAS .....	4
1.3. EPIGENÉTICA.....	5
1.4. METILACIÓN DE ADN EN PLANTAS.....	6
1.4.1 Metiltransferasas en la metilación <i>de novo</i> .....	10
1.4.2 Metiltransferasas en la metilación de mantenimiento .....	11
1.6. GÉNERO AGAVE.....	13
1.7 AGAVE <i>ANGUSTIFOLIA</i> HAW. ....	14
1.8 JUSTIFICACIÓN.....	17
1.9 PREGUNTA DE INVESTIGACIÓN .....	18
1.10 HIPÓTESIS.....	19
1.11 OBJETIVO GENERAL.....	20
1.12 OBJETIVOS ESPECÍFICOS.....	20
1.13 DISEÑO EXPERIMENTAL.....	21
CAPÍTULO II .....	23

---

---

<b>IDENTIFICACIÓN DE SECUENCIAS DE GENES DE METILTRANSFERASAS DE ADN EN <i>A. ANGUSTIFOLIA</i> HAW.....</b>	<b>23</b>
<b>2.1. INTRODUCCIÓN .....</b>	<b>23</b>
<b>2.2. MATERIALES Y MÉTODOS.....</b>	<b>24</b>
2.2.1. Ensamble de transcriptoma.....	25
2.2.2. Análisis de expresión de genes diferencialmente expresados.....	27
2.2.3. Identificación de genes de metiltransferasas de ADN en el transcriptoma .....	27
2.2.4. Identificación de dominios conservados de metiltransferasas de ADN en el transcriptoma ensamblado.....	28
2.2.5. Árbol filogenético de proteínas de metiltransferasas de ADN .....	28
<b>2.3. RESULTADOS .....</b>	<b>29</b>
2.3.1. Análisis de calidad de lecturas de secuenciación del transcriptoma de <i>A. angustifolia</i> Haw.....	29
2.3.2. Distribución de longitud de transcritos y obtención de unigenes del transcriptoma de <i>A. angustifolia</i> Haw.....	30
2.3.3. Anotación funcional de genes identificados en el transcriptoma de <i>A. angustifolia</i> Haw.....	31
2.3.4. Mapeo de lecturas de secuenciación del transcriptoma de <i>A. angustifolia</i> Haw.....	32
2.3.5. Búsqueda de identidad de secuencias del transcriptoma de <i>A. angustifolia</i> Haw. con el algoritmo BLAST del NCBI .....	32
2.3.6. Identificación de dominios conservados en secuencias de genes de metiltransferasas de ADN.....	33
2.3.7. Análisis filogenético de las secuencias identificadas en el transcriptoma.....	35

---

---

<b>CAPÍTULO III.....</b>	<b>37</b>
<b>ANÁLISIS DE EXPRESIÓN DE GENES DE METILTRANSFERASAS DE ADN .....</b>	<b>37</b>
<b>3.1. INTRODUCCIÓN .....</b>	<b>37</b>
<b>3.2. MATERIALES Y MÉTODOS.....</b>	<b>38</b>
3.2.1. Diseño de cebadores .....	38
3.2.2. Condiciones de cultivo y características de las plántulas .....	39
3.2.3. Extracción de ARN .....	39
3.2.4. Síntesis de ADNc.....	40
3.2.5. PCR punto final.....	40
3.2.6. Determinación del nivel de expresión por densitometría de genes de metiltransferasas de ADN de <i>A. angustifolia</i> Haw. ....	40
3.2.7. Elección de genes de metiltransferasas de ADN para analizar por RT-qPCR .....	41
3.2.8. Análisis de expresión de metiltransferasas a partir del reporte del transcriptoma .....	41
3.2.9. Medición de niveles de expresión de metiltransferasas de ADN por RT-qPCR .....	41
<b>3.3. RESULTADOS .....</b>	<b>42</b>
3.3.1. Determinación del nivel de expresión por densitometría de genes de metiltransferasas de ADN de <i>A. angustifolia</i> Haw. ....	42
3.3.2. Niveles de expresión de genes de metiltransferasas de ADN reportados en el transcriptoma <i>A. angustifolia</i> Haw.....	45
3.3.3. Análisis de expresión de genes que codifican para metiltransferasas de ADN por RT-qPCR .....	46

---

---

<b>CAPITULO IV .....</b>	<b>49</b>
<b>DISCUSIÓN, CONCLUSIONES GENERALES Y PERSPECTIVAS .....</b>	<b>49</b>
<b>4.1. DISCUSIÓN GENERAL .....</b>	<b>49</b>
<b>4.1.1. Identificación de dominios conservados y construcción de árbol         filogenético .....</b>	<b>49</b>
<b>4.1.2. Nivel de expresión de genes que codifican para enzimas         metiltransferasas de ADN .....</b>	<b>50</b>
<b>4.2 CONCLUSIONES GENERALES .....</b>	<b>53</b>
<b>4.3 PERSPECTIVAS .....</b>	<b>54</b>
<b>BIBLIOGRAFÍA .....</b>	<b>55</b>

---

## LISTADO DE FIGURAS

<b>Figura 1. 1</b> Mecanismo de metilación del ADN .....	8
<b>Figura 1. 2</b> Metilación de ADN de novo y de mantenimiento .....	9
<b>Figura 1. 3</b> Dominios conservados de genes de metiltransferasas de ADN en <i>A. thaliana</i> .....	10
<b>Figura 1. 4</b> Diseño experimental del presente trabajo.....	21
<b>Figura 1. 5</b> Flujo de trabajo seguido por la empresa Novogene para el ensamble y análisis de transcriptoma de <i>A. angustifolia</i> Haw. ....	24
<b>Figura 2. 1</b> Dominios conservados en secuencias de aminoácidos de metiltransferasas de la familia DRM. ....	34
<b>Figura 2. 2</b> Dominios conservados en secuencias de aminoácidos de metiltransferasas de la familia Met. ....	34
<b>Figura 2. 3</b> Dominios conservados en secuencias de aminoácidos de metiltransferasas de la familia CMT.....	35
<b>Figura 2. 4</b> Árbol filogenético de proteínas de MTasas de ADN.. ....	36
<b>Figura 3. 1</b> Gel de electroforesis para genes de Mtasas de ADN identificados en el transcriptoma de <i>A. angustifolia</i> Haw. ....	44
<b>Figura 3. 2</b> Análisis densitométrico de genes de MTasas de ADN identificados en el transcriptoma de <i>A. angustifolia</i> Haw. ....	44
<b>Figura 3. 3</b> Niveles de expresión de transcritos expresados deferencialmente reportados en el ensamble de transcriptoma de <i>A. angustifolia</i> Haw. ....	45
<b>Figura 3. 5</b> Expresión relativa de los transcritos que codifican para metiltransferasas de ADN de <i>A. angustifolia</i> Haw .....	¡Error! Marcador no definido.

---

---

## LISTADO DE TABLAS

<b>Tabla 2. 1</b> Resultados de calidad de secuenciación del transcriptoma de <i>A. angustifolia</i> Haw.	30
<b>Tabla 2. 2</b> Número total de transcritos y unigenes identificados en el transcriptoma de <i>A. angustifolia</i> Haw.....	30
<b>Tabla 2. 3</b> Datos estadísticos del ensamble de transcriptoma de <i>A. angustifolia</i> Haw. ....	31
<b>Tabla 2. 4</b> Unigenes anotados del transcriptoma de <i>A. angustifolia</i> Haw. ....	31
<b>Tabla 2. 5</b> Lecturas mapeadas al ensamble de novo del transcriptoma de <i>A. angustifolia</i> Haw. ....	32
<b>Tabla 2. 6</b> Resultados de búsquedas de similitud por BLAST de secuencias de metiltransferasas de ADN identificadas en el transcriptoma de <i>A. angustifolia</i> Haw.....	33
<b>Tabla 2. 7</b> Dominios identificados en proteínas de MT-asas de ADN de <i>A. angustifolia</i> Haw....	73

---

## ABREVIATURAS

2,4-D	Ácido 2,4 diclorofenoxiacético
5mC	Metilación en el carbono 5' de la citosina
6-BA	6-bencilaminopurina
A	Fenotipo albino
ADN	Ácido desoxirribonucleico
AFLP	Polimorfismos en la longitud de fragmentos amplificados
AGO 3	Complejo ARGONAUTA 3
AGO 4	Complejo ARGONAUTA 4
AGO 6	Complejo ARGONAUTA 6
ARN	Ácido ribonucleico
ARNip	ARN pequeño de interferencia
ARNm	ARN mensajero
AV	Región albina del tejido variegado
BAH	Dominio de homología bromo adyacente
BCP	1-bromo-3-cloropropano
BLAST	Herramienta de búsqueda por alineamiento básico local
C <sub>3</sub>	Metabolismo C <sub>3</sub>
C <sub>4</sub>	Metabolismo C <sub>4</sub>
ADNc	ADN complementario
CAM	Metabolismo ácido de las crasuláceas
CD	Cromodominio
CMT	CROMOMETILTRANSFERASA
CTAB	Bromuro de hexadeciltrimetilamonio
DCL3	ENDONUCLEASA DICER-LIKE 3
DMT	Dominio metilasa de ADN
Dnmt 1	ADN-METILTRANSFERASA 1
DRM	METILTRANSFERASAS DE DOMINIOS REARREGLADOS
FPKM	Fragmentos por millón de kilobases
G	Fenotipo verde

---

GC	Porcentaje de guaninas-citosinas
Go	Ontología de genes
GV	Región verde del tejido variegado
H1	Histona 1
H3	Histona 3
H3K9	Lisina de la histona 3
H3K9me2	Dimetilación de la lisina nueva en la histona 3
HA	Tejido de hoja albina
HEN1	Enhancer HUA 1
HG	Tejido de hoja verde
HMM	Modelos ocultos de Markov
HPLC	Cromatografía líquida de alta eficiencia
HV	Tejido de fenotipo variegado
KEGG	Enciclopedia de genes y genomas de Kyoto
KNO3	Nitrato de potasio
KO	Ontología de KEGG
KOG	Grupos ortólogos de proteínas de eucariotas
KYP	Enzima KRYPTONITA
MA	Tejido meristemático de plántulas con fenotipo albino
MET1	METILTRANSFERASA DE ADN 1
MG	Tejido meristemático de plántulas con fenotipo verde
mi-ARN	Micro ARN
Mtasas-ADN	Metiltransferasas de ADN
N50	Parámetro de calidad de ensamble N50
N90	Parámetro de calidad de ensamble N90
NaCl	Cloruro de sodio
NCBI	Centro Nacional para la Información Biotecnológica
NH <sub>4</sub> NO <sub>3</sub>	Nitrato de amonio
Nr	Base de datos de secuencias no redundantes del NCBI
Nt	Base de datos de nucleótidos del NCBI

---

PCR	Reacción en cadena de la polimerasa
Pfam	Base de datos de familias y dominios de proteínas
PGDD	Base de datos de duplicación de genomas de plantas
POL IV	Polimerasa de ARN IV
POL V	Polimerasa de ARN V
PVP	Polivinilpolipirrolidona
Q20	Valor Q20 del puntaje de Phred
Q30	Valor Q30 del puntaje de Phred
RAPD	Amplificación aleatoria del ADN polimórfico
RdDM	Metilación de ADN dirigida por ARN
RDR2	Polimerasa dependiente de ARN 2
RFD	Dominio de replicación de foci
RFLP	Polimorfismos en la longitud de los fragmentos de restricción
RT-qPCR	PCR en tiempo real
SAH	S-adenosil-L-homocisteína
SAM	S-adenosil-L-metionina
SSR	Secuencias simples repetidas
SUV4	Supresor de variegación 3-9 homólogo 4
T <sub>m</sub>	Temperatura de alineamiento
V	Fenotipo variegado
VS	Variación somaclonal



---

## RESUMEN

Durante la micropropagación de plantas es posible observar variación fenotípica entre la planta madre y sus hijas, a esta variación se le conoce como variación somaclonal (VS). Este fenómeno puede ser causado por factores genéticos que se describen como cambios en la secuencia del ADN, o por factores epigenéticos en los que participan mecanismos de regulación en los genes que no cambian la secuencia del ADN. En la micropropagación de *Agave angustifolia* Haw. se detectó una VS, en la cual a partir de plántulas verdes emergieron plántulas variegadas y albinas. El albinismo en las plantas es un evento que ocurre de manera poco frecuente en la naturaleza y del cual se sabe poco acerca de su origen molecular, es por esta razón que las variantes somaclonales de *A. angustifolia* Haw. son un modelo de estudio interesante para explorar el albinismo. En estudios anteriores realizados en nuestro laboratorio se encontró una similitud genética mayor al 97% entre las variantes somaclonales de *A. angustifolia* Haw., este hallazgo sugiere que la VS se puede explicar con base a mecanismos epigenéticos. La metilación del ADN es un mecanismo epigenético que sucede cuando las enzimas llamadas metiltransferasas de ADN (mtasas-ADN) transfieren un grupo metilo a la citosina de la cadena del ADN provocando el silenciamiento o disminución de la expresión de los genes. Por lo tanto, la VS de *A. angustifolia* Haw. se podría explicar al detectar cambios en la expresión de genes que codifican para mtsas-ADN entre los fenotipos, debido a que estas estarían regulando genes involucrados en procesos como la biogénesis y desarrollo de los cloroplastos. En el estudio descrito a continuación se utilizaron herramientas bioinformáticas para encontrar secuencias de genes que codifican para mtsas-ADN en el transcriptoma de *A. angustifolia* Haw. y para identificar los dominios conservados que permiten llevar a cabo la función de estos genes. Posteriormente se construyó un árbol filogenético para corroborar la identidad de los genes utilizando las secuencias encontradas en el transcriptoma de *A. angustifolia* Haw. y las secuencias de 15 especies de plantas diferentes. Por último, se seleccionaron algunos genes de mtsas-ADN para analizar su expresión con la técnica de PCR en tiempo real. El análisis reveló una disminución en la expresión de dichos genes en los fenotipos variegado y albino, esta característica podría indicar que existen cambios en la regulación de otros genes que pudieran estar provocando la aparición de los nuevos fenotipos.

---

---

## ABSTRACT

During plants propagation it is possible to observe phenotypic variation between mother plant and the generated new shoots; this variation is known as somaclonal variation (VS). This phenomenon can occur by genetic factors, which are described as changes on DNA sequence, or by epigenetic factors, where regulation mechanisms of genes participate and do not change the DNA sequence. In *Agave angustifolia* Haw. micropropagation VS was detected, in which variegated and albino shoots were emerged from the green plantlets. Plants albinism is a rare event in nature and less is known about its molecular origin. For this reason, *A. angustifolia* Haw. somaclonal variants are an interesting study model to explore albinism. In previous investigations done in our lab there was found a genetic similitude greater than 97% between *A. angustifolia* Haw. somaclonal variants. This finding suggests that this VS can be explained by epigenetic mechanism. DNA methylation is an epigenetic mechanism, which occurs when enzymes known as DNA methyltransferases (DNA-mtases) transfer a methyl group to the cytosines in the DNA chain causing decrease or silencing of genes expression. Therefore, the VS of *A. angustifolia* Haw. could be explained by detecting changes in DNA-mtases coding genes among the phenotypes, because they would be regulating genes involved in process like biogenesis and chloroplast development. In the study described below there were used bioinformatic tools in order to find gene sequences coding for DNA-mtases in the *A. angustifolia* Haw. transcriptome, and also to find the conserved domains, which allow these genes to develop their function. Also, a phylogenetic tree was built in order to corroborate the gene's identity by using the sequences found in *A. angustifolia* Haw. transcriptome and sequences from 15 different plant species. Finally, there were chosen some DNA-mtases genes to analyze their expression using real time PCR technique. This analysis revealed a decreased in some DNA-mtases coding genes in variegated and albino phenotypes. This could indicate that there are changes in the regulation of other genes that could be triggering the emergence of new phenotypes.

---

---

---



### INTRODUCCIÓN

La variación somaclonal (VS) es la variación fenotípica observada entre las diferentes plantas o somaclones obtenidas a partir de un proceso de micropropagación (Bridgen *et al.*, 1994; Sahijram *et al.* 2003; Rica Sánchez-Chiang, 2009; Borges *et al.*, 2021). Esta variación puede ser causada por factores genéticos o epigenéticos determinados por distintas fuentes como el tipo de explante utilizado, su genotipo, el número y duración de los subcultivos, así como la concentración y el tipo de reguladores de crecimiento empleados (Peschke y Phillips, 1992; Tariq y Paszkowski, 2004; Rica Sánchez-Chiang, 2009; Sarmah, 2017).

Durante la micropropagación de *Agave angustifolia* Haw. se observó el surgimiento espontáneo de una VS entre las plántulas verdes (G). A partir de esta VS se obtuvieron dos fenotipos que se distinguen por el color de sus hojas, estos fenotipos son variegado (V) y albino (A) (Duarte-Aké *et al.*, 2016). Se realizó un análisis con marcadores moleculares de polimorfismos en la longitud de fragmentos amplificados (AFLP), encontrando una similitud genética entre los fenotipos de más del 97% (Us-Camas *et al.*, 2017) y una diferencia epigenética relacionada a los ciclos de cultivo y al fenotipo (Duarte-Aké *et al.*, 2016). Duarte-Aké *et al.*, 2016 cuantificó el nivel de metilación global en el ADN de los fenotipos verdes, variegados y albinos en plantas madre y en los nuevos brotes durante cinco subcultivos. Se observó una correlación entre los niveles globales de metilación en el ADN con la aparición de brotes variegados y albinos en los distintos subcultivos.

La metilación del ADN, es un mecanismo epigenético mediado por enzimas llamadas metiltransferasas que juega un papel importante en la expresión de los genes y en la regulación del desarrollo de las plantas (Nie, 2021). Además, existen reportes que describen que los cambios en los patrones de metilación están directamente relacionados con cambios en la expresión de genes que codifican para metiltransferasas (de *novo* y de mantenimiento), y este fenómeno es más frecuente en plantas obtenidas por cultivo *in vitro* que durante su reproducción en campo (Ghosh *et al.*, 2021). La investigación en plantas como por ejemplo *Dendrobium officinale* (una especie de orquídea) (Yu *et al.*, 2021), *Gingko biloba* (Guo *et al.*, 2020) y *Actinidia chinensis* (kiwi) (Zhang *et al.*, 2020) dejan en evidencia el papel específico que juegan estos genes en diferentes etapas de desarrollo y tipo de tejido.

En el presente trabajo se propone evaluar los niveles de expresión de los genes que codifican a metiltransferasas que participan en la metilación *de novo* y de mantenimiento del ADN, en las hojas y meristemas de plántulas de *A. angustifolia* Haw. con fenotipo verde (G), variegado (V) y

albino (A). Para tal fin se realizará una búsqueda con herramientas bioinformáticas de los genes que codifican a metiltransferasas en el transcriptoma de *A. angustifolia* Haw. La expresión diferencial de los genes de metiltransferasas identificados se determinará en los diferentes tejidos analizados mediante RT-qPCR.

## CAPÍTULO I

### ANTECEDENTES

#### 1.1. VARIACIÓN SOMACLONAL

La variación somaclonal (VS) se define como la variación fenotípica determinada por factores genéticos o epigenéticos que se origina durante la micropropagación de plantas obtenidas a partir del cultivo de células, tejidos u órganos vegetales (Bridgen *et al.*, 1994; Sahijram *et al.*, 2003; Rica Sánchez-Chiang, 2009; Azizi *et al.*, 2020). Los factores genéticos incluyen cambios heredables de eliminación, duplicación, inversión y translocación de segmentos de ADN que forman parte de distintos genes, provocando el desarrollo de diferentes niveles de ploidía, y rearrreglos en la estructura de los cromosomas (Sarmah, 2017). Se ha evaluado la presencia de la VS de origen genético en líneas de plantas de interés comercial como *Actinidia deliciosa* (kiwi) (Rajan, R.P., y SINGH, G. (2021), *Coffea arabica* (Azizi *et al.*, 2020), *Elaeis guinensis* (palma aceitera) (G. y A., 2014), *Humulus lupulus* (lúpulo) (Liberatore *et al.*, 2020) y de *Saccharum sp.* (caña de azúcar) (Rashda *et al.*, 2018). Para fines de estos estudios, se han utilizado métodos con marcadores moleculares tales como RFLP (Polimorfismos en la longitud de los fragmentos de restricción), SSR (Secuencias simples repetidas), RAPD (Amplificación aleatoria del ADN polimórfico) y AFLP (Polimorfismos en la longitud de fragmentos amplificados) (Sarmah, 2017).

Por otro lado, los factores epigenéticos actúan como mecanismos de control de expresión génica que incluyen mecanismos de micro-ARN y modificaciones químicas en las histonas y ADN, estas modificaciones no cambian la secuencia del ADN (Tariq y Paszkowski, 2004). En plantas y animales, la hipermetilación de ADN y las modificaciones de histonas como la metilación de la histona H3 en la lisina 9 (H3K9) están relacionadas con el compactamiento de la cromatina que provoca el silenciamiento de los genes (Borg *et al.*, 2021). En adición, se han descrito modificaciones en las histonas tales como acetilación, fosforilación, ubiquitinación y ADP-ribosilación (Bártová *et al.*, 2008; Kim *et al.*, 2009; Vining *et al.*, 2013; Machczyńska *et al.*, 2014). Dichas modificaciones en el ADN e histonas son heredables a través de la mitosis o meiosis y se ha observado que algunas de ellas son mantenidas en múltiples generaciones de células e incluso organismos. Sin embargo, se sabe que dichas modificaciones pueden ser reversibles (Kaepler *et al.*, 2000; Bird 2007; Vranich *et al.*, 2014; Hung *et al.*, 2018).

La aparición de la VS es generalmente espontánea y dependiente del sistema de regeneración con el que se esté trabajando, así como el genotipo, el tipo de tejido, la fuente del explante, la

composición del medio de cultivo, la duración y el número de ciclos de cultivo (Peschke y Phillips, 1992; Rica Sánchez-Chiang, 2009; Sarmah, 2017). Desde un primer punto de vista, la VS es un fenómeno deseable en sistemas de investigación con enfoque de fitomejoramiento cuyo objetivo es la extensión de la variabilidad genética que se encuentra de manera natural en los cultivos y que puede representar una ventaja económica (Tabarez Z E, 1991). En contraste, la VS también puede representar un problema en aquellos sistemas de micropropagación que se utilizan para reproducir plantas élite con características fenotípicas idénticas y muy específicas, por ejemplo, en programas de conservación de germoplasma (Tabarez Z E, 1991; Cardone y Olmos, 2010).

## 1.2. EL ALBINISMO EN PLANTAS

Una de las VS poco comunes en cultivo de explantes foliares, pero muy frecuente en cultivo de anteras, es el albinismo (Abadie *et al.*, 2005; Kumari *et al.* 2009, Zhao *et al.* 2017). El albinismo en las plantas se caracteriza por la pérdida total o parcial de pigmentos en tejidos que son normalmente verdes. Puede ocurrir de manera natural y con poca frecuencia. Sin embargo, es común observar este fenómeno en plantas obtenidas en sistemas de cultivo *in vitro* en los que se realizan cruza para obtener ejemplares híbridos y en las plantas derivadas del uso de tecnología de cultivo de tejidos (Kumari *et al.*, 2009).

Las plantas albinas se caracterizan por la ausencia total de clorofila en tejidos que normalmente son verdes y las plantas variegadas por la ausencia parcial de éste pigmento. Durante un análisis comparativo entre plantas verdes y albinas de *Cephalanthera longifolia* (una especie de orquídea). Abadie *et al.*, (2006) encontraron que el contenido de clorofila en las plantas albinas era significativamente menor ( $1.9 \pm 5 \times 10^{-2}$   $\mu\text{g}/\text{mg}$  de peso fresco) comparado con las plantas verdes ( $2.97 \pm 0.56$   $\mu\text{g}/\text{mg}$  de peso fresco). La clorofila es un pigmento que se requiere durante la fotosíntesis para captar la energía luminosa de la luz solar, y posteriormente utilizarla en la fotólisis del agua; por tal motivo, la ausencia de clorofila en los tejidos de las plantas provoca una muerte prematura de éstas. Sin la ruta metabólica de fotosíntesis, las plantas no pueden crecer autotróficamente y rápidamente agotan las reservas de alimento necesarias para su reproducción y desarrollo (Kumari *et al.*, 2009)

Los estudios reportados sobre el albinismo en las plantas se han enfocado principalmente en observar y analizar los efectos que tiene este fenómeno en la fisiología vegetal. En el cultivo de tejidos (anteras, callo, y embriogénesis) el albinismo puede surgir en explantes de plantas

totalmente normales, con una incidencia de hasta 11 plántulas albinas por una verde en arroz y en algunas variedades de caña puede llegar a observarse que el 40% de las plántulas totales son albinas (Khan *et al.* 2009; Begum *et al.* 2011).

La obtención de plantas a partir del cultivo *in vitro* de microsporas y polen inmaduro es un proceso cuya efectividad en la cebada se ve reducida por la aparición de plántulas albinas (Gajecka *et al.*, 2020). En un estudio reciente se encontró que este fenómeno tiene su origen en el estado de diferenciación de los plastidios de las plantas al momento de iniciar el cultivo, en el 90% de los explantes que producen plantas albinas, la expresión de genes relacionados con la transcripción, traducción y síntesis de almidón era significativamente mayor a los explantes que producen plantas verdes (Gajecka *et al.*, 2020).

Algunas investigaciones con maíz y otros cultivos de cereal como la avena, y trigo, sugieren que la causa del albinismo en estas plantas se debe a mutaciones genéticas que afectan el desarrollo de plastidios y la síntesis de carotenoides (Migneault *et al.*, 2019). Desde el punto de vista epigenético existe un número reducido de investigaciones enfocadas en el albinismo, lo que representa un área de oportunidad importante para investigar.

### **1.3. EPIGENÉTICA**

El biólogo, fisiólogo escocés y fundador de la biología de sistemas Conrad Waddington, introdujo el término epigenética por primera vez en 1940 y lo definió de la siguiente forma: “La epigenética es la rama de la biología que estudia las interacciones entre los genes y sus productos las cuales dan lugar al fenotipo” (Waddington, 1942). Una definición más moderna de la epigenética se refiere al conjunto de cambios heredables en la expresión de los genes que no se deben a cambios en la secuencia del ADN (Holliday, 2006; Bairu *et al.*, 2011).

La epigenética comprende mecanismos de control de expresión génica que en conjunto funcionan como un sistema que puede ser heredado mediante meiosis o mitosis y en adición es reversible. Sin embargo, algunas modificaciones epigenéticas han mostrado ser estables a través de múltiples generaciones, fenómeno al que se le denomina memoria epigenética (Birnbaum y Roudier, 2017). Asimismo, el ambiente con el que interactúan las células puede ser determinante para provocar cambios epigenéticos que pueden verse reflejados en su descendencia (Kaepler *et al.*, 2000; Holliday, 2006; Anastasiadi *et al.*, 2021). Varios autores señalan que, si bien los factores genéticos pueden influir en la aparición de la VS, los factores epigenéticos son la fuente

más frecuente de la variación fenotípica en plantas cultivadas en sistemas de micropropagación (Miguel y Marum, 2011).

Los mecanismos epigenéticos se dividen en metilación del ADN, modificaciones de histonas y mecanismos de ARN no codificante (Boquete *et al.*, 2021). Los dos primeros mecanismos influyen en la estructura y compactación de la cromatina y por lo tanto en la accesibilidad de la maquinaria de transcripción a la información genética (Chuang & Jones, 2007; Goldberg *et al.*, 2007; Borg *et al.*, 2021). Mientras que los mecanismos del ARN involucran moléculas de ARN de aproximadamente 22 nucleótidos llamados micro ARN (miARN) que pueden unirse a secuencias de ARN mensajero provocando su degradación o evitando su traducción (Chuang & Jones, 2007; Basso *et al.*, 2019).

El mecanismo epigenético más estudiado y mejor descrito es la metilación del ADN, del cual se ha reportado que en células eucariotas tiene un papel importante en el silenciamiento de elementos móviles de ADN como los transposones, virus y retroelementos (Erdmann & Picard, 2020; Baduel & Colot, 2021; Ghosh *et al.*, 2021), mantenimiento de la integridad del genoma (Sow *et al.*, 2021), impronta parental (Batista & Köhler, 2020), inactivación del cromosoma X en mamíferos (Strehle & Guttman, 2020), regulación del desarrollo (Kumar & Mohapatra, 2021) y participa en la regulación de genes (Zhang, H., *et al.*, 2018; Bednarek & Orłowska, 2020). Existen dos mecanismos para explicar la represión en la expresión de los genes cuando se encuentran metilados. La primera de ellas dicta que esta modificación causa represión transcripcional al interferir directamente con la unión de los factores de transcripción al ADN, esta explicación surge debido a que se han identificado múltiples reguladores de transcripción que no pueden unirse en elementos de reconocimiento metilados (Greenberg & Bourc'his, 2019). El segundo mecanismo se lleva a cabo a través de la misma metilación del ADN, la cual atrae proteínas que se unen específicamente a regiones del ADN metilado. Esta explicación se basa en la identificación de proteínas que contienen un dominio de unión a ADN metilado como las deacetilasas de histonas que pueden provocar la represión transcripcional al establecer un estado compactado de cromatina (Greenberg & Bourchis, 2019).

#### **1.4. METILACIÓN DE ADN EN PLANTAS**

En plantas, la metilación del ADN ocurre en tres sitios o contextos: CG, CHG y CHH (donde H puede ser Adenina, Citosina o Timina) (Chan *et al.*, 2005; De Mendoza *et al.*, 2018). Estas secuencias se encuentran distribuidas principalmente en regiones repetitivas del genoma que

forman parte de transposones, centrómeros, genes ribosomales 5S y 45S, en regiones de promotores y en regiones codificantes de genes altamente expresados (Zilberman *et al.*, 2007; Pikaard y Scheid, 2014; Kumar & Mohapatra, 2021).

Los patrones de metilación y la posición de la cromatina son controlados durante el desarrollo de las plantas o durante la diferenciación de los tejidos (Miguel y Marum, 2011; Law y Jacobsen, 2010). Es posible que el mayor cambio del nivel de metilación en el ADN genómico en el cultivo *in vitro* se debe a la edad del cultivo y al tipo y concentración de reguladores de crecimiento (LoSchiavo *et al.*, 1989; Neelakandan y Wang, 2012; Bobadilla Landey *et al.*, 2013). Para estudiar la variación en el nivel de metilación en el genoma de plantas cultivadas *in vitro* se han utilizado diferentes cultivos como: cebada (Bednarek *et al.*, 2007; Li *et al.*, 2007), cacao (Rodríguez López *et al.*, 2010), chícharo (Smýkal *et al.*, 2007), vid (Schellenbaum *et al.*, 2008), plantago (Kour *et al.*, 2009), doritaenopsis (un género de orquídeas) (Park *et al.*, 2009), *Arabidopsis* (Yu, Z., *et al.*, 2010) y maíz (Yu, X., *et al.*, 2011).

En el genoma de *Arabidopsis thaliana*, algunas regiones inactivas de la heterocromatina (la conformación compacta e inaccesible de la cromatina) incluyendo centrómeros y pericentrómeros están altamente metilados (hasta un 80%) (Cokus *et al.*, 2008). En la eucromatina, se presentan niveles menores de metilación en las citosinas; en estas regiones, los pseudogenes y genes no expresados muestran niveles mayores de metilación que los genes que se expresan activamente, lo que demuestra el papel de la metilación en las citosinas como un mecanismo de silenciamiento transcripcional (Cokus *et al.*, 2008). Del total de genes que se expresan, el 5% están metilados en la región de los promotores y el 33% están metilados en regiones que se transcriben (Cokus *et al.*, 2008). Existen mapas detallados de las metilaciones que pueden encontrarse en *Arabidopsis*, se ha demostrado que los sitios simétricos CG están metilados al 24%, mientras que el 6.7% se encuentran metilados en los contextos CHG, y de los sitios asimétricos CHH están metilados el 1.7% (Cokus *et al.*, 2008). Por último, aproximadamente el 55% de todas las citosinas metiladas en *Arabidopsis* se localizan en contextos CG, el 23% en sitios CHG, y el 22% se encuentran en CHH. Los tres sitios de metilación se encuentran en regiones pericentrométricas que son repetitivas y ricas en transposones, sin embargo, la mayoría de los sitios CG están metilados en el cuerpo de los genes (Ashapkin *et al.*, 2016).

También se ha estudiado el nivel de metilación de los distintos contextos en otras plantas de interés comercial. Por ejemplo, en hojas de *Oryza sativa*, Niederhuth *et al.*, (2016), encontraron

un porcentaje de metilación en CG de 58.4%, en contextos CHG de 31.1% y en CHH 5.1%. Por otro lado, en *Zea mays L.* se identificó un porcentaje de 24%, 6.7% y 1.7% en contextos CG, CHG y CHH respectivamente. La metilación en promedio de *Populus trichocarpa* (álamo negro) es de 43.99% en CG, 29.84% en sitios CHG y 11.57% en CHH (Liang *et al.*, 2019). En *Brassica rapa*, Liu, G., *et al.*, (2018), encontraron una metilación de 39.3%, 15.38% y 5.24% en CG, CHG, y CHH, respectivamente. Estos datos parecen indicar un cierto patrón en la metilación del ADN en las plantas en donde se observa el mayor nivel de metilación en el contexto CG, seguido de CHG y por último CHH. Sin embargo, esta condición no siempre se cumple, por ejemplo, en *Capsicum annuum* se ha observado un contenido de metilación de 19.9% en sitios CG, 30.5% en CHG y 49.6% en CHH (Rawoof *et al.*, 2019).

### Metiltransferasas de ADN

La metilación del ADN es catalizada por enzimas llamadas metiltransferasas de ADN (MTAs-ADN) las cuales participan en la transferencia de un grupo metilo desde una molécula de S-adenosil-L-metionina (SAM) a un átomo del carbono 5 de citosina en el ADN (5mC) (**Figura 1.1**). Como resultado de esta reacción, la SAM es convertida en una molécula de S-adenosil-L-homocisteína (SAH) (Zhang, H., *et al.*, 2018).

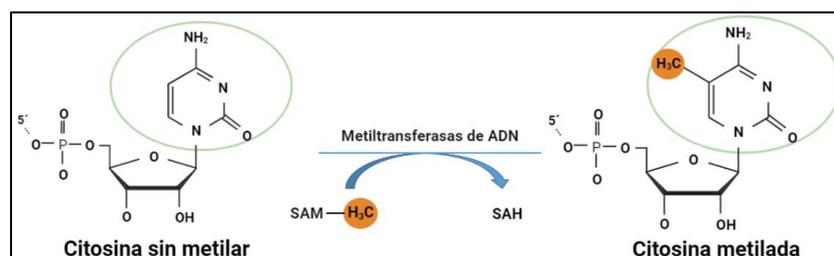


Figura 1. 1 Mecanismo de metilación del ADN. Las enzimas metiltransferasas de ADN transfieren un grupo metilo de la molécula S-adenosil-L-metionina (SAM) a un átomo de carbono 5 de una citosina del ADN. Como resultado la SAM es convertida en S-adenosil-L-homocisteína.

La actividad de las enzimas MTAs-ADN puede ser de metilación *de novo* y de mantenimiento (**Figura 1.2**). En el primer caso, las enzimas catalizan la metilación del carbono 5' en las citosinas que no contienen un grupo metilo formando nuevos patrones; y en el segundo caso, las enzimas se aseguran de que se conserven patrones de metilación preexistentes durante la replicación del ADN (Liu, Z. 2014; Pikaard y Scheid, 2014; Kumar & Mohapatra, 2021).

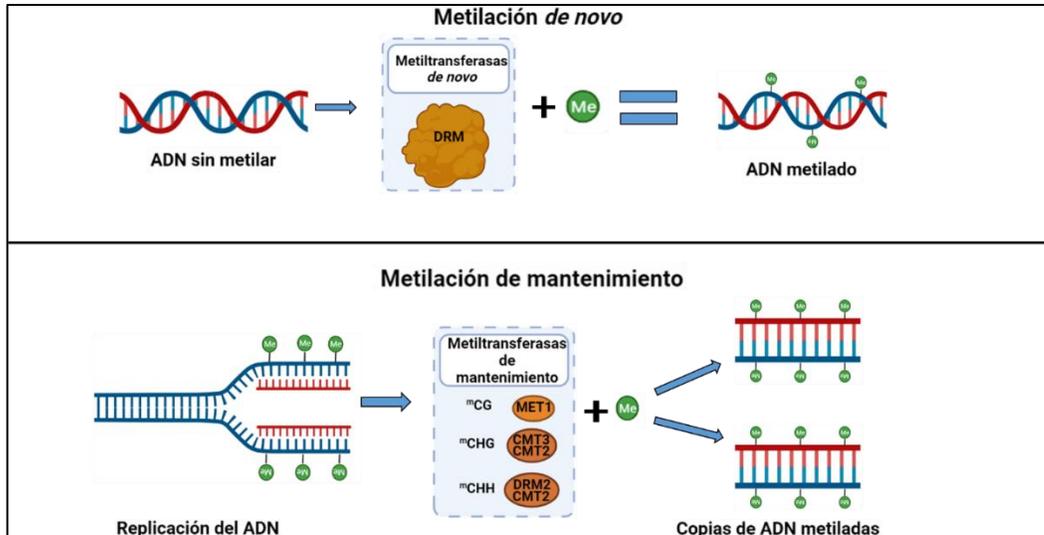


Figura 1. 2 Metilación de ADN de novo y de mantenimiento. La metilación de novo en el ADN es catalizada por las metiltransferasas de dominios rearrglados (DRM). Este tipo de metilación puede ocurrir en cualquier contexto de citosinas en moléculas de ADN que no están metiladas formando nuevos patrones de metilación. Después de la replicación, las metiltransferasas de mantenimiento aseguran que se mantengan los patrones existentes del ADN en las nuevas copias. La metiltransferasa 1 (MET1) participa en la metilación de contextos CG. La metilación en los contextos CHG es llevada a cabo por las enzimas cromometiltransferasas 3 y 2 (CMT3 y CMT2). Mientras que las metiltransferasas DRM2 y CMT2 metilan contextos CHH. C: Cisteína, H: Cualquier nucleótido excepto guanina.

Las enzimas MTAsas-ADN cuentan con dos dominios principales: el dominio *N*-terminal cuya función es reconocer una secuencia específica en el ADN y el dominio catalítico *C*-terminal que participa en la transferencia de un grupo metilo desde la molécula de SAM al carbono 5 del anillo de residuos de citosina (Zhong *et al.*, 2014). Además, las familias de MTAsas-ADN comparten un dominio conservado llamado metilasa de ADN (DMT) (Yaari *et al.*, 2019).

Las MTAsas-ADN en plantas se pueden clasificar en tres familias con base a su homología con las metiltransferasas de mamíferos: Metiltransferasas de Dominios Rearreglados (por sus siglas en inglés DRM), Metiltransferasas 1 (MET1) y Cromometiltransferasas (CMT), de las cuales cada una tiene una función en la metilación *de novo* y de mantenimiento en el ADN (Finnegan y Kovac, 2000b; Ryazanova *et al.*, 2016). La estructura de los dominios de cada familia de MTAsas-ADN es altamente conservada entre todas las especies de plantas (Pavlopoulou y Kossida, 2007). Es importante mencionar que la metilación en las plantas es un sistema complejo en el que existe una comunicación entre los diferentes patrones de metilación, especialmente entre CHG y CHH. De esta manera CMT2 en algunos casos funciona en el mantenimiento de metilaciones en contextos CHG apoyando a CMT3, mientras que CMT3 puede funcionar en conjunto con CMT2

y DRM2 (Stroud *et al.*, 2014). En la **Figura 1.3** se muestra la estructura molecular de cada uno de los tipos de las enzimas MTasas-ADN que se encuentran en *A. thaliana*.

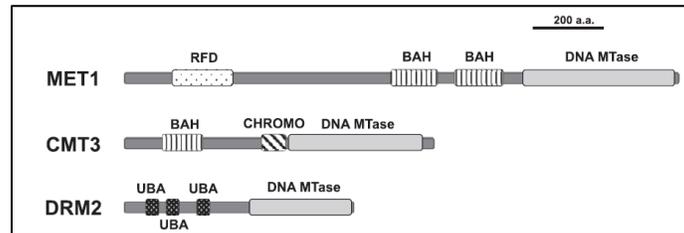


Figura 1. 3 Dominios conservados de genes de metiltransferasas de ADN en *A. thaliana*. Las tres familias de genes de MTasas de ADN comparten el dominio catalítico metilasa de ADN. Además, cada familia tiene dominios específicos. La familia Met contiene dos dominios BAH (de homología bromo adyacente) y un dominio RFD (dominio de replicación de foci). La familia CMT se caracteriza por contener un dominio BAH (de homología bromo adyacente) y un dominio CD (cromodominio). Los genes contenidos en la familia DRM además de presentar un dominio metilasa de ADN (DMT) también contienen al menos un dominio UBA (dominio asociado a ubiquitina). Imagen Tomada y modificada de Maeji y Nishimura, 2018.

### 1.4.1 Metiltransferasas en la metilación *de novo*

#### Metiltransferasas de Dominios Rearreglados (DRM)

A diferencia de la metilación *de novo* de mamíferos que solo ocurre en contextos CG dentro del ADN, la metilación *de novo* en plantas ocurre en todos los contextos: CG, CHH y CHG, la cual es mediada por las metiltransferasas DRM2 de manera dependiente de la vía de metilación de ADN dirigida por ARN (RdDM) (Law y Jacobsen, 2010; Pikaard y Scheid, 2014). Las enzimas DRM presentan homología con las DNMT3 de mamíferos, contienen un dominio de metilasa de ADN y un dominio asociado a ubiquitina (UBA) (Cao *et al.*, 2000b; Henderson *et al.*, 2010; Zhong *et al.*, 2014; Maeji y Nishimura, 2018).

RdDM es una ruta específica en plantas que conduce la metilación *de novo* de ADN y es necesaria para el mantenimiento de metilación en contextos CHH (Stolzenburg *et al.*, 2016; Kumar & Mohapatra, 2021). Para explicar esta ruta se puede dividir en dos fases: la primera de ellas es la fase de biogénesis de un ARN pequeño de interferencia (ARNip), y la segunda consiste en una metilación dirigida. La biogénesis de ARNip es iniciada por una polimerasa de ARN IV (POL IV) única en plantas. Esta enzima genera una cadena sencilla de ARN y posteriormente es

copiada para formar una cadena doble de ARN por la enzima ARN polimerasa dependiente de ARN 2 (RDR2). La cadena resultante es cortada en ARNip de 24 nucleótidos por la endonucleasa Dicer-like 3 (DCL3), el corte de esta enzima produce extremos cohesivos en la posición 3' de los ARNip que posteriormente son metilados por la metiltransferasa de ARN HUA ENHANCER 1 (HEN1). A continuación el ARNip es cargada en ARGONAUTA 4 (AGO4) o ARGONAUTA 6 (AGO6) (Zhong *et al.*, 2014) formando varios complejos de AGO4/6-ARNip. Las cadenas de ARNip guían a AGO4 a secuencias homólogas de ADN, a su vez AGO4 interactúa con la metiltransferasa DRM2 que metila la cadena de ADN. También se sabe que el complejo AGO4-ARNip puede interactuar con la polimerasa V (POL V) y con sus transcriptos para guiar la metilación en secuencias homólogas (Mahfouz, 2010; Kumar & Mohapatra, 2021).

## 1.4.2 Metiltransferasas en la metilación de mantenimiento

### 1.4.2.1. Metiltransferasas (MET)

Las metiltransferasas 1 o MET, son las enzimas homólogas en plantas de la familia de metiltransferasas de mamíferos DNMT1t y su papel principal es el de mantenimiento de grupos metilo en las secuencias CG (Pikaard y Scheid, 2014; Maeji y Nishimura, 2018). En *A. thaliana*, la mutación de *MET1* provoca un decremento en la metilación en los contextos CG (Kankel *et al.*, 2003). La familia Met1 posee un dominio catalítico de metilasa de ADN DMT en el sitio C-terminal. (Maeji y Nishimura, 2018). Estas enzimas además tienen dos dominios de replicación de foci (RFDs) los cuales permiten dirigir a la proteína al sitio de replicación (Maeji y Nishimura, 2018). La enzima MET1 también contiene dos dominios de “homología bromo adyacente” (BAH) entre los dominios RFDs y DMT (**Figura 1.3**). Estos dominios se encuentran comúnmente en proteínas involucradas en la regulación de la cromatina, es posible que participe en una interacción de proteína-proteína con la deacetilasa 6 de histonas para regular el silenciamiento de transposones (Liu *et al.*, 2012; Maeji y Nishimura, 2018).

En *A. thaliana* se han identificado cinco genes de *MET* de los cuales solo se han caracterizado cuatro: *MET1*, *MET2a*, *MET2b* y *MET3* (Jean finnegan y Dennis, 1993; Genger *et al.*, 1999; Prasad Parita, 2018). Los cuatro genes provienen de un gen ancestral a través de eventos de duplicación. La estructura del gen, incluyendo la posición de sus 11 intrones está conservada entre todos los miembros de la familia (Ashapkin *et al.*, 2016). Con base a sus niveles de expresión, *MET1* es el gen que codifica a la enzima más abundante, el nivel de expresión de los genes *MET2a*, *MET2b*, y *MET3* en células de órganos vegetativos es diez veces menor

comparado con *MET1*, mientras que los niveles de expresión de *MET2a* y *MET2b* en el endospermo son similares a los niveles de *MET1* en órganos vegetativos (Ashapkin *et al.*, 2016). El nivel de expresión de *MET3* en el endospermo es cuatro veces menor que *MET2a* y *MET2b*, pero su expresión es varias veces mayor al nivel observado en órganos vegetativos (Ashapkin *et al.*, 2016).

#### 1.4.2.2. Cromometiltransferasas (CMT)

Las cromometiltransferasas (CMT) es una familia de enzimas única en las plantas y su papel es el de mantenimiento de metilación en contextos CHG y CHH (Cao y Jacobsen, 2002; Pavlopoulou y Kossida, 2007). En *Arabidopsis* se encuentran tres genes de esta familia: *CMT1*, *CMT2* y *CMT3*, además del dominio DMT cada uno contiene un cromodominio (CD) que reconoce la dimetilación de lisina en la posición nueve de la histona H3 (H3K9me2) y un dominio BAH (**Figura 1.3**) (Maeji y Nishimura, 2018; Prasad Parita, 2018).

En múltiples estudios se ha encontrado que el gen *CMT1* está incompleto en distintos ecotipos de *Arabidopsis* y por lo tanto no puede codificar una enzima catalíticamente activa. Esto no prueba que el gen carezca de uso en la planta, sin embargo, no se sabe cuál es función biológica (Ashapkin *et al.*, 2016). El mantenimiento de metilaciones en regiones CHH es llevada a cabo por DRM2 o CMT2, dependiendo de la región del genoma (Huettel *et al.*, 2006; Zemach *et al.*, 2013; Liu *et al.*, 2014). DRM2 mantiene grupos metilo en secuencias CHH de regiones donde ocurre RdDM, las cuales se localizan en regiones evolutivamente jóvenes de transposones, y otras secuencias repetitivas en la eucromatina del cromosoma (Huettel *et al.*, 2006; Zemach *et al.*, 2013; Liu *et al.*, 2014). En contraste, la enzima CMT2 cataliza la metilación en secuencias CHH de elementos transponibles de las Histonas 1 (H1) contenidas en la heterocromatina (Jeddeloh, *et al.*, 1999; Zemach *et al.*, 2013).

Por último, en *A. thaliana*, las metilaciones en secuencias CHG (donde H es A, T o C) son mantenidas principalmente por CMT3 y en menor medida por CMT2 en regiones de la heterocromatina que pertenecen a retrotransposones (Zemach *et al.*, 2013; Maeji & Nishimura, 2018). Las enzimas CMT3 puede reconocer histonas H3 cuyas lisinas están metiladas en la posición de lisina 9 (H3K9me2) utilizando los dominios CD y BAH para mantener la metilación en estas regiones (Lindroth *et al.*, 2004; Pikaard y Scheid, 2014; Maeji & Nishimura, 2018). La enzima metiltransferasa de histonas responsable de establecer la metilación en H3K9me2 es la KRYPTONITE (KYP)/SUPPRESSOR OF VARIATION 3-9 HOMOLOG 4 (SUVH4), la cual

puede reconocer y unirse al ADN metilado (Lindroth *et al.*, 2001; Maeji & Nishimura, 2018). Esta característica sugiere entonces que CMT3 y KYP/SUVH4 reforzando la metilación tanto en histona como en ADN de manera redundante (Du *et al.*, 2014; Maeji y Nishimura, 2018).

## 1.6. GÉNERO AGAVE

El género *Agave* forma parte de la familia *Agavaceae*, en esta familia están contenidas diferentes especies de plantas adaptadas a condiciones de aridez. Se han reportado 210 especies de las cuales 159 se encuentran en México por tal motivo este país es considerado centro de origen del *Agave* (García-Mendoza, 2002). Las plantas de este género tienen forma de roseta, hojas suculentas con estomas hundidos y se caracterizan por presentar metabolismo fotosintético CAM (Domínguez *et al.*, 2008). Se componen de hojas que pueden ser anchas o angostas, el centro o piña rica en azúcares que pueden ser utilizadas para la producción de bebidas alcohólicas con ayuda de una fermentación alcohólica y la inflorescencia llamada coloquialmente quiote, la cual nace desde el centro de la planta (Pérez-Zavala *et al.*, 2020). Los agaves son plantas monocárpicas, es decir que solo florecen una vez en su ciclo de vida (Pérez-Zavala *et al.*, 2020). El metabolismo tipo CAM les permite utilizar de dos a cuatro veces más eficientemente el agua que plantas con metabolismo  $C_3$  y  $C_4$ , además tienen la capacidad de sobrevivir más de una temporada sin lluvias y pueden tolerar temperaturas muy bajas y altas (-16.1°C a 61.4°C) (Zhou *et al.*, 2012; Davis *et al.*, 2017). El género *Agave* se caracteriza por tener un genoma complejo, con una longitud de 2940 a 4704 Mpb entre sus especies y con alto nivel de duplicación debido a diferentes niveles de poliploidía (2x,3x,4x,5x,6x) (Zhou *et al.*, 2012).

Las especies dentro del género *Agave* han sido utilizadas para la obtención y producción de alimento, fibras, forraje, medicamento, material de construcción, elaboración de bebidas, uso ornamental y doméstico (García-Herrera *et al.*, 2010). Además, tanto las hojas como el tallo de las especies de *Agave* contienen grandes cantidades de carbohidratos y lignocelulosa por lo que son considerados como una fuente importante de material lignocelulósico para su aprovechamiento en la producción de bioenergía (Davis *et al.*, 2017).

Debido a que el índice de germinación de algunas especies es bajo como en el caso de *A. tequilana* (entre el 4% y 6%) (Ramírez *et al.*, 2000) la forma más común de reproducción consiste en separar los vástagos (hijuelos, retoños) que surgen del rizoma que crece de la planta madre, otros métodos de propagación asexual descritos se basan en la utilización de rizomas, secciones del tallo y por bulbillos de la inflorescencia (García-Herrera *et al.*, 2010).

## 1.7 AGAVE ANGUSTIFOLIA HAW.

La especie de *A. angustifolia* Haw. es utilizada en varios estados de la república mexicana para la elaboración de bebidas alcohólicas como el mezcal o bacanora (Barraza-Morales, *et al.*, 2006). Su distribución abarca zonas de la costa del océano pacífico, desde Sonora hasta Costa Rica y en las costas del océano atlántico desde Costa Rica hasta Tamaulipas; además, se pueden encontrar a una altura de hasta 1500 m sobre el nivel del mar (Gentry, 1982). En la península de Yucatán, esta especie presenta variación morfológica en dependencia de su distribución geográfica; en las dunas costeras se observan plantas pequeñas, en los bosques tropicales caducifolios se encuentran plantas de un tamaño intermedio, y las plantas más grandes de esta región se han encontrado en los bosques tropicales subcaducifolios (Orellana, 2015). *A. angustifolia* Haw. presenta problemas de fecundación debido a que son plantas hermafroditas y producen polen antes de que los estigmas de las mismas plantas estén maduros (Illsley *et al.*, 2004). Además, esta especie presenta floración después un largo ciclo vegetativo (8 a 12 años) con una baja eficiencia en producción de semillas y germinación de las mismas (Arzate-Fernández y Mejía-Franco, 2011). Se ha propagado de manera convencional de forma asexual al separar los hijuelos que surgen de la planta madre, los cuales se desarrollan de rizomas producidos por cultivos establecidos en campo por varios años (Arzate-Fernández y Mejía-Franco, 2011).

En una investigación realizada por Palomino *et al.*, (2017) se llevó a cabo un análisis citogenético para determinar el cariotipo de *A. angustifolia* Haw. de dos cultivares llamados “Cimarrón” y “Lineño” obtenidos de la región Suroeste de Jalisco, México. También, se desarrolló un análisis de citometría de flujo para determinar el tamaño del genoma de estos dos cultivares. Los autores concluyeron que los cultivares “Cimarrón” y “Lineño” contienen un sistema diploide  $2n=2x=60$ , con un número de cromosomas básico de 30, los cuales comprenden de 25 cromosomas pequeños y 5 cromosomas grandes. Por último, el análisis de contenido genético reveló que el genoma de los cultivares tiene una longitud de 4,013 Mpb para el cultivar Cimarrón y 4,002 Mpb para el cultivar Lineño y en términos de distancia el genoma diploide de cada uno mide 102.69  $\mu\text{m}$  y 121.12  $\mu\text{m}$ , respectivamente, lo que indica que existen pocas diferencias en cuanto a su longitud. Estos estudios sugieren que el genoma de *A. angustifolia* Haw. cultivar Cimarrón y Lineño tiene una longitud media en comparación con el genoma de plantas angiospermas (plantas con flores), el cual el genoma más pequeño reportado pertenece a *Genlisea margaretae*

(63.3744 Mpb) y el genoma más grande es de la especie *Paris japónica* (148,880.94 Mpb) (Bennett y Leitch, 2011; Palomino *et al.*, 2017).

### **Variantes fenotípicas de *Agave angustifolia* Haw.**

El doctor Manuel L. Robert, en el Centro de Investigación Científica de Yucatán (CICY) junto con su equipo de trabajo, publicaron en el 2006 un método eficiente y exitoso de micropropagación para diferentes especies y variedades de *Agave* (nueve en total). El método permite la selección de tejidos de alto rendimiento y la rápida propagación de clones de alta calidad con una variabilidad genética muy baja (Robert *et al.*, 2006).

En el laboratorio de epigenética del CICY bajo la dirección de la Dra. Clelia De la Peña se utiliza el mismo método desarrollado por el doctor Manuel L. Robert y su equipo para la reproducción de *A. angustifolia* Haw. Durante la micropropagación de esta especie de *Agave* se observó la aparición espontánea del fenotipo variegado (V) y posteriormente del albino (A) a partir del fenotipo verde (G) (Duarte-Aké *et al.*, 2016). Debido a que el albinismo es un fenómeno difícil de observar en la naturaleza, el estudio de los tres fenotipos (G, V y A) de esta especie representa una oportunidad muy importante para explorar el efecto fisiológico de las plantas y las bases biológicas que explican el albinismo.

A nivel macroscópico, se observaron diferencias morfológicas entre los tres fenotipos, las hojas de las plántulas con fenotipo verde son totalmente verdes y presenta la forma de roseta típica de los *Agaves* (Duarte-Aké *et al.*, 2016). Las plántulas con fenotipo variegado se caracterizan por tener hojas con una región verde y una región albina que se desarrollan desde la parte basal, hasta la punta de las hojas (Duarte-Aké *et al.*, 2016; Us-Camas *et al.*, 2017). Por otro lado, las plántulas albinas tienen una coloración totalmente blanca y el área del meristemo presenta zonas sin diferenciar parecidas a un callo (Duarte-Aké *et al.*, 2016; Us-Camas *et al.*, 2017). Duarte-Aké *et al.*, (2016), observaron diferencias significativas entre los fenotipos utilizando microscopio electrónico de barrido. Las plántulas albinas tienen dos células adyacentes a la estoma en vez de cuatro como en el caso de las plántulas variegadas y verdes, estas cuatro células son características de plantas con metabolismo CAM. También se observó una densidad de estomas mayor al 10% en la zona adaxial y abaxial de las plántulas verdes en comparación con las zonas verde y albina de las plántulas variegadas (Duarte-Aké *et al.*, 2016; Hernández-Castellano *et al.*, 2020).

Asimismo, en el estudio publicado por Duarte-Aké *et al.*, (2016) se cuantificó el contenido de clorofila *a*, clorofila *b* y carotenoides de los tres fenotipos y se demostró que las hojas verdes y la región verde de las plántulas variegadas tienen el contenido más alto de clorofila *a* y carotenoides. El contenido de clorofila *b* fue mayor en las plántulas variegadas que en las verdes, pero con la mitad de contenido de clorofila *a* en comparación con las plántulas con fenotipo verde. Además, las plantas albinas mostraron un contenido muy bajo de clorofila *a*, *b* y carotenoides. Se pudo observar que las regiones albinas de las plántulas variegadas presentan niveles más altos de clorofila *a*, clorofila *b* y carotenoides que las plántulas albinas (Hernández-Castellano *et al.*, 2020).

Por otro lado, en el trabajo publicado por Us-Camas *et al.*, (2017), se realizó un análisis molecular utilizando marcadores moleculares AFLP con la finalidad de conocer si existía variabilidad genética entre los fenotipos. En este estudio se observó una variabilidad genética menor al 3% entre los fenotipos, este hallazgo sugiere de manera importante que la aparición de los fenotipos variegado y albino podría explicarse con base a mecanismos epigenéticos. Asimismo, se realizó un análisis de metilación global utilizando tejido de las hojas del fenotipo G, del fenotipo A y el tejido de la zona verde (GV) y albina (AV) de las hojas de plantas con fenotipo variegado (Us-Camas *et al.*, 2017; Duarte-Aké *et al.*, 2016). Para este análisis se utilizaron técnicas de HPLC, en donde se comprobó que, aunque no se encontró una diferencia importante entre los niveles de metilación del ADN genómico de los tres fenotipos de 3 meses de cultivo (Us-Camas *et al.*, 2017), la formación de nuevos brotes variegados y albinos estaba fuertemente ligada a una hipometilación en el ADN (Duarte-Aké *et al.*, 2016). Se ha reportado que las metilaciones en genes específicos que son necesarios para la biogénesis de plastidios pueden estar involucrados en la aparición de fenotipos variegado y albino (Wang, Q., *et al.*, 2015; Wang, Q., *et al.*, 2016). Por esta razón, no se puede descartar el papel de las metilaciones puntuales en el genoma de los fenotipos y sería muy interesante conocer cuáles son los genes que están expresados al alta y a la baja, como consecuencia de una regulación epigenética mediada por metilación en el ADN, y que estén provocando la variación encontrada entre los fenotipos de *A. angustifolia* Haw.

## 1.8 JUSTIFICACIÓN

El albinismo en plantas es un fenómeno que se presenta de manera poco frecuente en la naturaleza, lo cual dificulta su investigación. En plantas, la mayoría de los estudios se han enfocado en estudiar los efectos en la fisiología de las mismas y poco se sabe aún del origen y las bases biológicas que explican este evento. Las técnicas que se han utilizado para explicar el albinismo, implican la inducción de mutaciones en las plantas que muchas veces provocan una baja supervivencia o efectos pleiotrópicos. Durante la micropropagación de *Agave angustifolia* Haw. se observó la aparición espontánea de plantas con fenotipo variegado y albino; asimismo, es posible mantenerlas vivas en condiciones *in vitro* por prolongados periodos de tiempo. En estudios anteriores en nuestro laboratorio se analizó el contenido genético de los tres fenotipos y se concluyó que presentan una variación genética casi nula, lo que sugiere una explicación epigenética detrás de este fenómeno. También, se encontró que las plantas variegadas tenían menos metilación en su ADN que las verdes, y las albinas tenían menos que las variegadas. Sin embargo, actualmente no existe información sobre el papel de las enzimas metiltransferasas de ADN, contemplando los mecanismos de metilación *de novo* y mantenimiento, a través de los cuales pudiera estar ocurriendo este fenómeno. En el presente trabajo, se propone primero identificar el número y tipo de metiltransferasas de ADN en *A. angustifolia* y analizar por medio de RT-qPCR los niveles de expresión de los genes que codifican para las metiltransferasas de ADN tanto en el meristemo como en las hojas de las plantas con fenotipo verde, variegado y albino. Se espera que los niveles de expresión encontrados en estos genes puedan proveer las respuestas sobre cuáles son las enzimas metiltransferasas de ADN involucradas en la metilación *de novo* o de mantenimiento que se encuentran participando en los diferentes tejidos de estas variantes somaclonales.

## **1.9 PREGUNTA DE INVESTIGACIÓN**

¿Cuál es el patrón de expresión de los genes que codifican para metiltransferasas de ADN en el tejido del meristemo y hoja de los fenotipos G, V y A?

## 1.10 HIPÓTESIS

Los niveles de expresión de los genes que codifican para las enzimas metiltransferasas de ADN varían en los fenotipos verde, variegado y albino en los tejidos de hoja y meristemo de *A. angustifolia* Haw. Esta diferencia de expresión podría estar relacionada a la hipometilación en las plantas albinas.

### 1.11 OBJETIVO GENERAL

Determinar la relación existente entre los niveles de expresión de genes que codifican para metiltransferasas de ADN y los fenotipos somaclonales variegado y albino de *A. angustifolia* Haw.

### 1.12 OBJETIVOS ESPECÍFICOS

- Identificar las secuencias de genes que codifican para metiltransferasas de ADN en el transcriptoma de *A. angustifolia* Haw.
- Corroborar la identidad de las secuencias encontradas en el transcriptoma mediante búsquedas de similitud y la construcción de un árbol filogenético
- Evaluar los niveles de expresión diferencial de genes que codifican para metiltransferasas de ADN en plantas de fenotipo verde, variegado y albino, en hoja y meristemo, así como la sección albina y la verde de las hojas variegadas.

1.13 DISEÑO EXPERIMENTAL

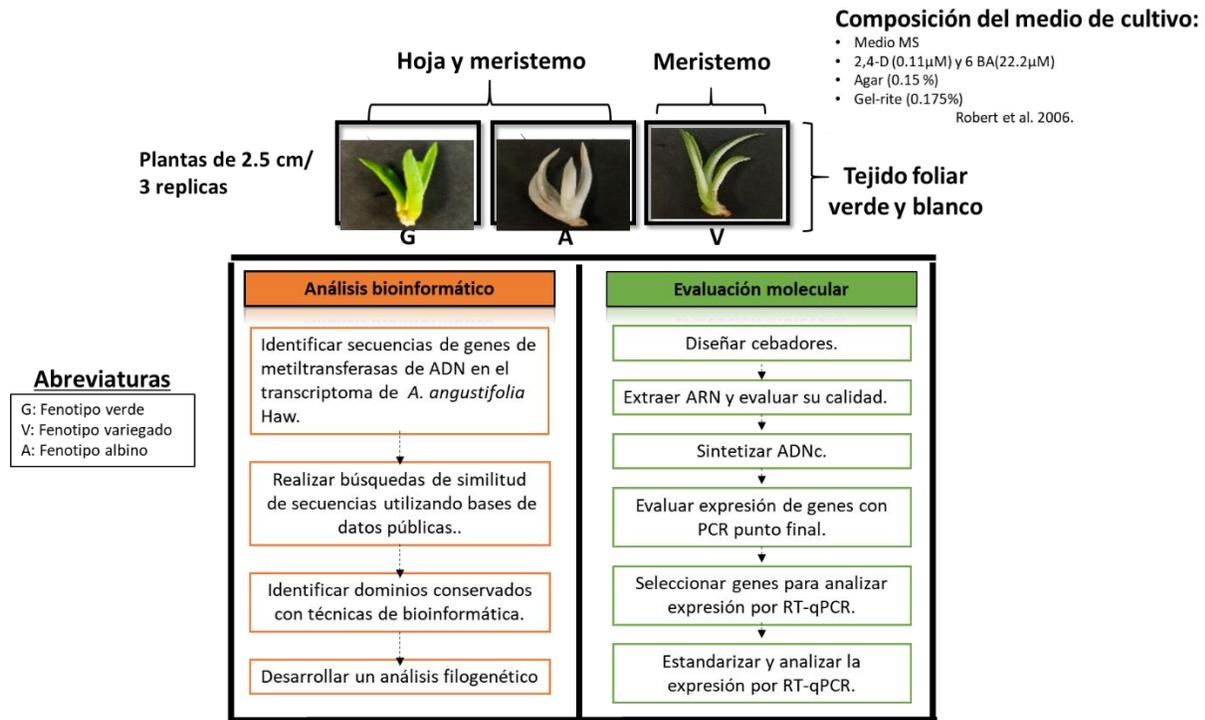


Figura 1. 4 Diseño experimental del presente trabajo. Variantes de *A. angustifolia* Haw. evaluadas: G: verde, V: variegada, A: albina.



---

## CAPÍTULO II

### IDENTIFICACIÓN DE SECUENCIAS DE GENES DE METILTRANSFERASAS DE ADN EN *A. ANGUSTIFOLIA* HAW.

#### 2.1. INTRODUCCIÓN

La bioinformática es una disciplina que se apoya de las matemáticas y recursos computacionales para desarrollar bases de datos, algoritmos, técnicas de estadística para analizar y resolver preguntas biológicas en áreas de investigación como la biología molecular, genética, fisiología y biotecnología de plantas (Marturano, 2009; Bilotta *et al.*, 2018).

Recientemente, las aplicaciones de la bioinformática dentro de la biotecnología de plantas se han expandido con la aparición de técnicas experimentales que permiten obtener información relevante a partir de genomas, transcriptomas, proteomas y metabolomas de diferentes especies de plantas (Gomez-Casati *et al.*, 2018). Esta información es depositada y organizada sistemáticamente en bases de datos públicas y privadas tales como NCBI, PFAM y EMBL-EBI para facilitar su acceso (Anamika *et al.*, 2019). Las herramientas de bioinformática como el algoritmo de búsqueda llamado *Blast* permiten buscar rápidamente secuencias nucleotídicas y proteínas de diferentes especies de organismos para posteriormente realizar un análisis comparativo con programas como Hmmer, que utilizan métodos probabilísticos para comparar secuencias similares y de este modo predecir *in silico* la posible función y estructura de nuevos genes (Zhao *et al.*, 2020).

Hoy en día, es posible aprovechar las tecnologías de secuenciación masiva para realizar análisis filogenéticos basados en algoritmos tales como *Neighbor joining* (Saitou & Nei, 1987) y *Maximum likelihood* (le Cam, 1990) con la finalidad de conocer las relaciones ancestrales entre las distintas especies de organismos dentro del árbol de la vida, determinar las relaciones entre familias de genes, conocer la historia de poblaciones y clasificar secuencias metagenómicas, genes, elementos regulatorios y ARN no codificante (Yang y Rannala, 2012; Young y Gillung, 2020).

Actualmente el grupo de epigenética del CICY dirigido por la Dra. Clelia De la Peña, cuenta con el transcriptoma de los fenotipos verdes y albinos en los tejidos de hoja y meristemo. En este capítulo se describen datos relevantes de la calidad de secuenciación y construcción del transcriptoma de *A. angustifolia* Haw. Se detalla, además, las herramientas de bioinformática y

las bases de datos empleadas para la identificación y clasificación de secuencias de genes que codifican para metiltransferasas de ADN dentro del transcriptoma.

## 2.2. MATERIALES Y MÉTODOS

A continuación, se describe la estrategia que se utilizó para el ensamble de transcriptoma, anotación funcional de genes y análisis de expresión de genes diferencialmente expresados en el transcriptoma de *A. angustifolia* Haw. El ensamble y análisis del transcriptoma se realizó en la compañía Novogene (<https://en.novogene.com/about/about-novogene/#>). En la **Figura 2.5** se muestra el flujo de trabajo utilizado por la empresa Novogene para ensamble y análisis del transcriptoma. También se describe la metodología y herramientas utilizadas para conocer la ubicación e identidad de los dominios conservados presentes en las secuencias de genes identificados y por último se desarrollan los pasos empleados para la construcción de un árbol filogenético utilizando las secuencias de aminoácidos de las metiltransferasas de *A. angustifolia* Haw. y secuencias de proteínas de otras especies de plantas reportadas en otros estudios.

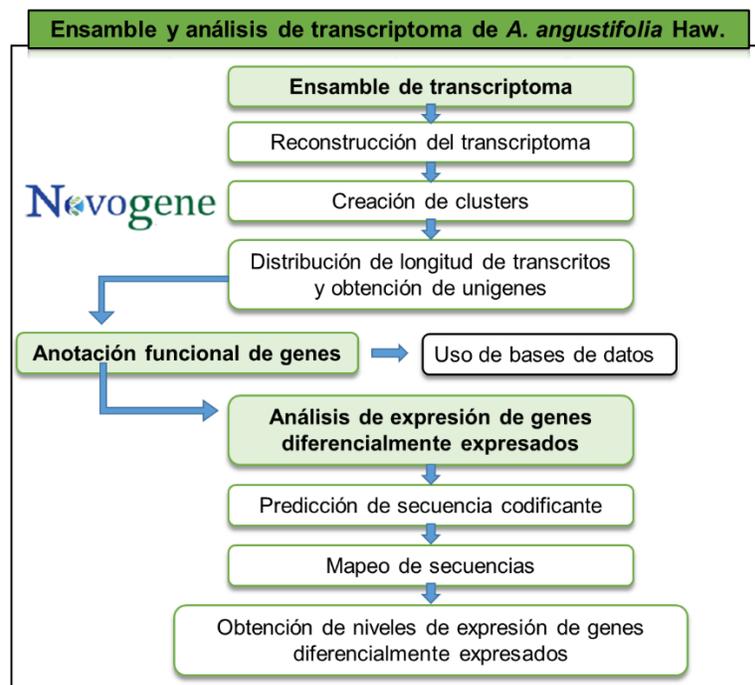


Figura 1. 5 Flujo de trabajo para el ensamble y análisis de transcriptoma de *A. angustifolia* Haw.

### 2.2.1. Ensamble de transcriptoma

El transcriptoma del fenotipo verde (G) y albino (A), tanto del meristemo (M) como de la hoja (H), se envió a secuenciar y ensamblar en la empresa Novogene (<https://en.novogene.com/services/research-services/transcriptome-sequencing/>). El análisis de las secuencias se realizó utilizando secuenciadores con tecnología Illumina.

#### 2.2.1.1. Reconstrucción del transcriptoma

El proveedor Novogene realizó un ensamble *de novo* utilizando el programa Trinity desarrollado por Broad Institute y la Universidad Hebrea de Jerusalem (Grabherr *et al.*, 2013).

#### 2.2.1.2. Creación de clusters

Una vez obtenidos los *contigs* a partir del ensamble, se agruparon y se organizaron en *clusters* por el método llamado “Hierarchical Clustering” para eliminar redundancia entre los transcritos. Para este paso se utilizó el programa *Corset* (versión 1.05) (Davidson y Oshlack., 2014).

#### 2.2.1.3. Distribución de longitud de transcritos y obtención de unigenes

Durante este paso, Novogene seleccionó los contigs o transcritos más largos del filtrado realizado en el programa *Corset*. La longitud máxima de los transcritos fue de 15,535 pb los cuales posteriormente fueron anotados y analizados.

#### 2.2.1.4. Anotación funcional de genes

Para la anotación de la función de los genes que forman el transcriptoma de *A. angustifolia Haw.* en los fenotipos verde y albino, se utilizaron siete bases de datos diferentes (Nr, Nt, Pfam, KOG, Swiss-Prot, KEGG, KO, GO), los cuales se describen a continuación:

**Nr:** Es la base de datos de secuencias no redundantes de nucleótidos del *NCBI*, esta base de datos incluye información de GenBank, PDB (Protein Data Bank), SwissProt, PIR (Protein Information Resource), PRF (Protein Research Foundation) entre otras. <https://www.ncbi.nlm.nih.gov/>

**Nt:** La base de datos de nucleótidos del *NCBI* a su vez contiene secuencias de nucleótidos del Genbank, EMBL y DDBJ. <https://www.ncbi.nlm.nih.gov/>

**Pfam:** (Protein family) (Mistry *et al.*, 2021): Es una base de datos comprendida por familias y dominios de proteínas, cada una de las cuales está representada por alineamientos múltiples de secuencias y modelos ocultos de Markov (HMMs). Las proteínas generalmente están compuestas por una o más regiones funcionales, llamadas dominios. Por lo tanto, la identificación

de los dominios que forman a las proteínas puede proveer información sobre la función biológica de la misma. Pfam también genera grupos superiores de entradas llamadas clanes. Un clan está formado por una colección de entradas que están relacionadas por un cierto grado de similitud de secuencias, estructura o perfil de HMM. Los datos presentados en Pfam están basados en información de proteomas de UniProt. Pfam también se puede utilizar para buscar alineamientos con secuencias y accesos de otras bases de datos como Uniport y NCBI.

**KOG** (euKaryotic Orthologus Groups) (Bioinformatics *et al.*, 2003): Es una base de datos que contiene relaciones de genes ortólogos de organismos eucarióticos. De acuerdo con su relación evolutiva, esta base de datos divide los genes homólogos en diferentes *clusters* ortólogos.

**Swiss-Prot**: Es una base de datos formada por secuencias de proteínas e información sobre su función de manera libre. Tiene dos secciones principales: UniProtKB/Swiss-Prot contiene secuencias anotadas y revisadas manualmente, y UniProtKB/TrEMBL que contiene secuencias anotadas automáticamente y que no están revisadas (Bairoch & Apweiler, 1997).

**KEGG** (Kyoto Encyclopedia of Genes and Genome): Es una base de datos que sirve para entender funciones complejas y herramientas de los sistemas biológicos, como la célula y los ecosistemas, desde un enfoque molecular, especialmente secuencias generadas a partir de secuenciación de genomas y otras tecnologías experimentales (Kanehisa *et al.*, 2021).

**KO** (Kegg orthology) Combina cada sistema de anotación de la base de datos KEGG en un solo resultado (Kanehisa *et al.*, 2021).

**GO** (Gene Ontology): Es un consorcio de información que provee información acerca de las proteínas y ARN no codificante generado a partir de genes contenidos en organismos vivos. Clasifica la información con base a diferentes atributos del producto de los genes: proceso biológico en el que está involucrado, función molecular y componente celular en el que se puede ubicar (Carbon *et al.*, 2021).

Los parámetros utilizados por la empresa Novogene para las búsquedas de unigenes en las diferentes bases de datos fueron los siguientes:

**NT**: BLAST del NCBI versión 2.2.28+, con valor e-value de 1e-5.

**NR, SwissProt, KOG:** Diamond 0.8.22. Para NR y SwissProt se utilizó un e-value de  $1e-5$  y para KOG  $1e-3e$ .

**PFAM:** Se utilizó para predecir la estructura de los dominios de las proteínas. El programa que se usó fue HMMER versión 3.0, con valor e-value de 0.01.

**GO:** Se utilizó BlastGO v2.5 (Gotz *et al.* 2008) y scripts contruidos por Novogene, con un valor e-value de  $1e-6$ .

**KEGG:** Servidor de anotación automática de KEGG, valor de e-value  $1e-10$

## **2.2.2. Análisis de expresión de genes diferencialmente expresados**

### **2.2.2.1. Predicción de secuencia codificante**

Para los unigenes que fueron identificados durante la búsqueda por BLAST de acuerdo con las bases de datos NR y Swissprot, se extrajo la secuencia de nucleótidos y se tradujo a secuencia de aminoácidos utilizando codones estándar. Los unigenes que no tuvieron resultados en la búsqueda por BLAST fueron analizados con el programa ESTScan versión 3.0.3 (Lottaz *et al.*, 2003) usando los parámetros por *default* para predecir la región codificante e identificar la dirección de las secuencias. El programa ESTScan está basado en Modelos Ocultos de Markov.

### **2.2.2.2. Mapeo de secuencias**

Para el mapeo de secuencias se trabajó con el transcriptoma ensamblado *de novo* y filtrado por el programa Corset para mapear las lecturas obtenidas de secuenciación con el programa RSEM versión 1.2.26 (Li & Dewey, 2011) y los parámetros por *default*.

### **2.2.2.3. Genes diferencialmente expresados en el transcriptoma**

El análisis se realizó con el programa DESeq versión 1.10.1 (Love *et al.*, 2014), el cual se basa en la distribución binomial negativa. Además, se realizó un filtrado para eliminar aquellos genes diferencialmente expresados que tuvieran un valor de probabilidad  $p\text{-value} < 0.05$

## **2.2.3. Identificación de genes de metiltransferasas de ADN en el transcriptoma**

Se utilizó el programa Transdecoder v 5.5.0 (<https://github.com/TransDecoder/TransDecoder/releases>) sin cambios en los parámetros para traducir las secuencias de nucleótidos del transcriptoma de *A. angustifolia* Haw. y obtener secuencias de aminoácidos a partir de los marcos abiertos de lectura más largos. Para trabajar

con Transdecoder se utilizó el servidor de bioinformática “Galaxy” versión 20.09 (Afgan *et al.*, 2018).

Posteriormente, se hizo una revisión de los resultados generados en el reporte del transcriptoma de la identificación de genes en diferentes bases de datos y se seleccionaron aquellas secuencias que fueron reconocidas como posibles genes de metiltransferasas de ADN. Para corroborar estos resultados, se utilizó el programa HMMER con el objetivo de identificar secuencias en el transcriptoma que tengan el dominio catalítico de los genes de metiltransferasas de ADN. Para este paso, se generó un perfil de alineamiento de los dominios catalíticos metilasa de ADN (código PF00145) reportados de plantas en la base de datos Pfam (Mistry *et al.*, 2021). Posteriormente, se utilizó este alineamiento y el transcriptoma traducido en aminoácidos para aislar solamente aquellas secuencias que contengan el dominio catalítico de las proteínas de metiltransferasas de ADN. Para este paso se trabajó con el programa HMMER versión 3.3 (Wheeler & Eddy, 2013). A continuación, las secuencias identificadas se analizaron mediante búsquedas de similitud con el algoritmo BLAST del Centro Nacional para la Información Biotecnológica (NCBI).

#### **2.2.4. Identificación de dominios conservados de metiltransferasas de ADN en el transcriptoma ensamblado**

Se llevaron a cabo búsquedas en la base de datos Pfam versión 34 (Mistry *et al.*, 2021) y la base de datos de dominios conservados (CDD) del NCBI <https://www.ncbi.nlm.nih.gov/>, para identificar la posición y longitud de los dominios conservados de los genes de metiltransferasas de ADN identificados. Por último, se utilizó el programa Genious 2021.1 (<https://www.geneious.com/>) para ilustrar la identidad y ubicación de los dominios conservados identificados en las proteínas de metiltransferasas de ADN obtenidas del transcriptoma ensamblado de *A. angustifolia* Haw.

#### **2.2.5. Árbol filogenético de proteínas de metiltransferasas de ADN**

Como método de corroboración para la clasificación de las secuencias de genes de metiltransferasas de ADN se construyó un árbol filogenético con las secuencias de aminoácidos de los genes identificados en el transcriptoma de *A. angustifolia* Haw. y las secuencias de aminoácidos de genes ortólogos de otras 14 especies de plantas reportadas en el trabajo desarrollado por Moglia *et al.*, (2019). En la presente tesis, para la elaboración del árbol filogenético, se utilizaron secuencias de proteínas de metiltransferasas de ADN contenidas en el

genoma secuenciado de 14 especies de plantas diferentes, cuatro monocotiledóneas (*O. sativa*, *Z. mays*, *Sorghum bicolor*, *Brachypodium distachyon*) y 10 dicotiledóneas (*A. thaliana*, *Solanum melongena*, *Cynara cardunculus*, *Glycine max*, *Solanum lycopersicum*, *Fragaria x anassa*, *Solanum tuberosum*, *Salvia miltiorrhiza*, *Ricinus communis*, y *P. trichocarpa*). Por último, se añadieron también al análisis las secuencias de metiltransferasas de ADN de *Asparagus officinalis* reportadas en la base de datos de Genbank. Para el alineamiento múltiple de secuencias de las proteínas se utilizó el programa *Muscle* (Edgar, 2004) con el método UPGMA y los parámetros restantes sin modificar. Posteriormente, el árbol filogenético se construyó con el programa MEGA 7 versión 7.0 (Kumar *et al.*, 2016) con el método estadístico *Neighbor-joining* y la prueba de Bootstrap con 1000 iteraciones. Asimismo, se utilizó el modelo de sustitución p-distance y pairwise deletion para el procesamiento de “Gaps”.

## **2.3. RESULTADOS**

### **2.3.1. Análisis de calidad de lecturas de secuenciación del transcriptoma de *A. angustifolia* Haw.**

Para la búsqueda de las secuencias que codifican para metiltransferasas de ADN fue necesario revisar el transcriptoma de *A. angustifolia* con que cuenta el laboratorio de epigenética de las plantas verdes y albinas. Para ello, primero se tuvo que verificar la calidad de lecturas de secuenciación para las réplicas de muestras de hoja (H) y meristemo (M) de fenotipo verde (G) y albino (A), los cuales obtuvieron porcentajes elevados de Q20, entre 95.47% y 96.96%. Estos valores representan que las bases que forman las lecturas tienen una detección correcta del 99% o mayor. Además, al observar el valor Q30, se identificó que del 89.29% al 92.31% de las bases en las lecturas tienen una detección correcta mayor al 99.9%. Por último, el porcentaje de contenido de GC en las lecturas de secuenciación, se encuentra entre el 49.2% al 51.36% dependiendo de la muestra en la **Tabla 2.1**.

Tabla 2. 1 Resultados de calidad de secuenciación del transcriptoma de *A. angustifolia* Haw.

Muestra	Lecturas crudas (millones)	Lecturas limpias (millones)	Bases limpias (Gb)	Q20 (%)	Q30 (%)	GC (%)
HG1	141	134	20.1	95.83	89.96	50.46
HG2	121	117	17.6	95.47	89.29	50.72
HG3	125	118	17.8	96.3	90.94	50.91
MG1	123	118	17.8	95.69	89.71	49.27
MG2	120	114	17.1	96.32	91.06	49.2
MG3	126	116	17.5	96.34	91.03	49.2
HA1	125	119	17.9	96.41	91.29	49.72
HA2	126	119	18	96.5	91.44	49.9
HA3	159	149	22.4	96.58	91.57	51.36
MA1	128	122	18.4	96.2	90.89	49.76
MA2	121	117	17.6	96.96	92.31	49.48
MA3	125	120	18.1	96.91	92.22	50.05

Muestra: Indica el nombre del tejido utilizado para el análisis, HG: hoja de la planta verde, MG: meristemo de la planta verde, HA: hoja de la planta albina, MA: meristemo de la planta albina. Bases limpias: Número de bases limpias multiplicadas por la longitud de lecturas mostradas en unidades de Gb. Q20: Porcentaje de bases cuyo porcentaje de detección correcta fue mayor al 99%. Q30: Porcentaje de bases cuyo porcentaje de detección correcta fue mayor al 99.9%. GC: Porcentaje de contenido de guaninas y citosinas en el número total de bases.

### 2.3.2. Distribución de longitud de transcritos y obtención de unigenes del transcriptoma de *A. angustifolia* Haw.

Para obtener los unigenes a partir de los *contigs*, se seleccionaron los transcritos más largos del filtrado realizado en el programa *Corset*, los cuales posteriormente fueron anotados y analizados. Durante el ensamble se obtuvieron 270,409 transcritos, de los cuales se obtuvieron 270,163 unigenes. En la **Tabla 2.2** se muestra el número total de unigenes y transcritos obtenidos, así como los rangos de longitudes para cada uno.

Tabla 2. 2 Número total de transcritos y unigenes identificados en el transcriptoma de *A. angustifolia* Haw.

Intervalo de longitud de secuencias	200-500pb	500pb-1kpb	1-2kpb	>2kpb	Total
Número de transcritos	105,974	84,735	54,930	24,770	270,409
Número de unigenes	104,729	48,734	54,930	24,770	270,163

200-500b: Número de secuencias de una longitud de entre 200 a 500pb; 500pb-1kpb; 1-2kpb y mayores a 2kpb

Con el análisis del ensamble del transcriptoma se pudo observar que la longitud mínima de los transcritos fue de 201 pb mientras que la longitud máxima fue de 15,535 pb (**Tabla 2.3**). El promedio de los unigenes secuenciados fue de 904 pb.

Tabla 2. 3 Datos estadísticos del ensamble de transcriptoma de *A. angustifolia* Haw.

	Longitud mínima	Longitud promedio	Longitud media	Longitud máxima	N50	N90	Nucleótidos totales (millones)
<b>Transcritos</b>	201	904	615	15,535	1,282	417	244.33
<b>Unigenes</b>	201	904	615	15,535	1,282	417	244.26

### 2.3.3. Anotación funcional de genes identificados en el transcriptoma de *A. angustifolia* Haw.

De los 270,163 unigenes encontrados, se logró anotar 141,541 (52.39%) y siendo el 7.12% encontrados en todas las bases de datos utilizadas. En la **Tabla 2.4** se sintetiza el porcentaje de unigenes que fueron detectados en cada una de las bases de datos utilizadas. Se pudo observar que la base de datos SwissProt fue la que anotó un mayor número de unigenes, representando el 33.93% mientras que con KO solo se pudo anotar el 16.44%.

Tabla 2. 4 Unigenes anotados del transcriptoma de *A. angustifolia* Haw.

Base de datos	Número de unigenes	Porcentaje (%)
NR	117,256	43.4
NT	75,885	28.08
KO	4,428	16.44
SwissProt	91,669	33.93
Pfam	90,297	33.42
GO	90,821	33.61
KOG	38,204	14.14

Para la anotación funcional de genes se utilizaron siete bases de datos, NR: base de datos de secuencias de nucleótidos no redundante (NCBI), NT: base de datos de nucleótidos (NCBI), KO: Ontología de KEGG, SwissProt: base de datos de proteínas, Pfam: familia de proteínas, GO: ontología de genes, KOG: grupos ortólogos de eucariotas. Número de unigenes: número de unigenes totales identificados. Porcentaje: porcentaje de unigenes anotados en las distintas bases de datos.

### 2.3.4. Mapeo de lecturas de secuenciación del transcriptoma de *A. angustifolia* Haw.

Se realizó un mapeo de las lecturas de secuenciación (**Tabla 2.5**) utilizando como referencia el transcriptoma ensamblado *de novo* y filtrado con el programa *Corset*. Del total de muestras, se logró mapear del 64.71% al 67.9% de las lecturas de secuenciación al transcriptoma ensamblado y usado como referencia.

Tabla 2. 5 Lecturas mapeadas al ensamble de novo del transcriptoma de *A. angustifolia* Haw.

Nombre de la muestra	Lecturas totales (millones)	Total de lecturas mapeadas (millones)
HG1	134.18	90.16 (67.19%)
HG2	117.11	75.78 (64.71%)
HG3	118.81	77.81 (65.49%)
MG1	118.71	79.09 (66.93%)
MG2	114.16	76.48 (66.99%)
MG3	116.73	77 (65.97%)
HA1	119.33	81.44 (68.24%)
HA2	119.99	82.26 (68.56%)
HA3	149	101.25 (67.96%)
MA1	122.69	82.3 (67.08%)
MA2	117.511	77.6 (66.04%)
MA3	120.34	80.97 (67.29%)

HG: tejido de hoja verde, MG: tejido meristemo de la planta verde, HA: hoja de la planta albina, MA: meristemo de la planta albina.

### 2.3.5. Búsqueda de identidad de secuencias del transcriptoma de *A. angustifolia* Haw. con el algoritmo BLAST del NCBI

Una vez que ya se tenía toda la información del transcriptoma, se procedió a buscar la identidad de secuencias y para ello se utilizaron las secuencias de genes que codifican para metiltransferasas de ADN en el transcriptoma de *A. angustifolia* Haw. Se utilizó la herramienta bioinformática BLAST que se encuentra instalado en la base de datos del Centro Nacional para la Información Biotecnológica (NCBI). Los mejores resultados de búsqueda coincidieron con genes de la especie *Asparagus officinalis* que pertenecen a las tres familias de metiltransferasas de ADN: DRM, MET y CMT. Se logró identificar 10 genes en total (**Tabla 2.6**): tres genes *DRM2* con un porcentaje de similitud de alrededor de 66% a 70%, un gen *DRM3* con un porcentaje de similitud de 60.38%. Por otro lado, solo una secuencia presentó un porcentaje de similitud de 80.83% con un gen *MET1B* de *A. officinalis*. Además, tres secuencias fueron identificadas como

genes *CMT2* y dos como *CMT3*, presentando un porcentaje de similitud de entre 84% y 89% en el primer caso y de aproximadamente 59% hasta 67% en el caso de los genes *CMT3*. Los resultados tuvieron un valor de E de 0 a  $6E-51$  y una cobertura de secuencia en el gen de referencia que varía de 57% hasta 89% aproximadamente

Tabla 2. 6. Resultados de búsquedas de similitud por BLAST de secuencias de metiltransferasas de ADN identificadas en el transcriptoma de *A. angustifolia* Haw.

Genes DRM2						
ID Transcriptoma	Gen	Gen ortólogo	Valor de E	Cobertura	%De similitud	#De accesión
Cluster-46512.96201	AaH DRM2a_201	DRM2- <i>Asparagus officinalis</i>	0	75%	76.65%	XP_020260564.1
Cluster-46512.110739	AaH DRM2b_739	DRM2- <i>Asparagus officinalis</i>	0	73%	76.65%	XP_020260564.1
Cluster-46512.142275	AaH DRM2c_275	DRM2- <i>Asparagus officinalis</i>	0	75%	77.40%	XP_020260564.1
Genes DRM3						
Cluster-46512.161086	AaH DRM3a_086	DRM23- <i>Asparagus officinalis</i>	0	84%	60.38%	XP_020254982.1
Genes CMT2						
Cluster-46512.160638	AaH CMT2a_638	CMT2-Isoforma X2- <i>Asparagus officinalis</i>	0	73%	70.50%	XP_010908353.1
Cluster-46512.159175	AaH CMT2b_175	CMT2-Isoforma X1- <i>Asparagus officinalis</i>	0	78%	69.17%	XP_010908353.1
Cluster-46512.167523	AaH CMT2c_523	CMT2-Isoforma X1- <i>Asparagus officinalis</i>	0	67%	66.61%	XP_010908353.1
Genes CMT3						
Cluster-46512.85248	AaH CMT3a_248	CMT3 Isoforma X3- <i>Asparagus officinalis</i>	0	85%	67.53%	XP_020251487.1
Cluster-46512.141003	AaH CMT3b_003	CMT3- <i>Asparagus officinalis</i>	0	65%	59.40%	XP_020246348.1
Gen MET						
Cluster-46512.117080	AaH MET1_080	MET1B- <i>Asparagus officinalis</i>	0	91%	80.83%	XP_020269746.1

### 2.3.6. Identificación de dominios conservados en secuencias de genes de metiltransferasas de ADN

Para la identificación de dominios conservados en las secuencias de aminoácidos de metiltransferasas encontradas en el transcriptoma de *A. angustifolia* Haw., se realizaron búsquedas de similitud en las bases de datos públicas Pfam y CD domains (NCBI). En la **Figura 2.1** se observa la ubicación de los dominios conservados de las secuencias de *A. angustifolia* Haw. que pertenecen a la familia DRM. Cada una de las secuencias contiene un dominio UBA (rectángulo naranja) en el extremo N-terminal y el dominio catalítico metilasa de ADN (rectángulo verde) en una ubicación muy similar del extremo C-terminal, además, la longitud de las secuencias varía en longitud de 576 a 617 aminoácidos. En el **Anexo 1** se detalla la longitud y ubicación de los dominios identificados, así como la base de datos en la que fueron identificados.

En el transcriptoma de *A. angustifolia* Haw., se logró identificar una secuencia que pertenece a la familia MET (**Figura 2.2**). Esta secuencia contiene dos dominios RFD (rectángulos grises), dos dominios BAH (rectángulos azules) en la región N-terminal, asimismo presenta un dominio

catalítico de metilasa de ADN (rectángulo verde) en la región C-terminal. La longitud de la secuencia es de 1548 aminoácidos.

Por último, se identificaron cinco secuencias de metiltransferasas de ADN que contienen los dominios característicos de la familia CMT (**Figura 2.3**). Cada una de las secuencias de aminoácidos contiene un dominio BAH (rectángulo azul) y un dominio CD (rectángulo amarillo) en el extremo *N*-terminal y el dominio catalítico metilasa de ADN (rectángulo verde) en el extremo C-terminal. La longitud de las secuencias varía de 698 a 928 aminoácidos.

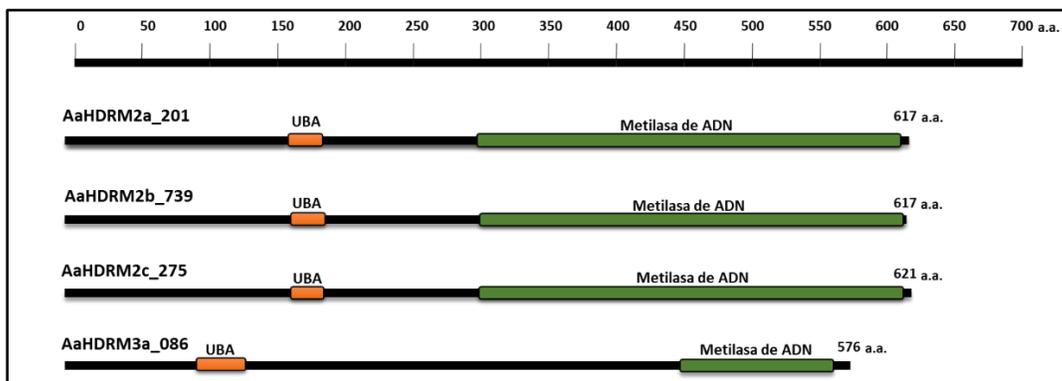


Figura 2. 1 Dominios conservados en secuencias de aminoácidos de metiltransferasas de la familia DRM. UBA: Dominio asociado a ubiquitina, Metilasa\_ADN: Dominio catalítico metilasa de ADN. En el lado derecho de cada secuencia se muestra su longitud real en aminoácidos.

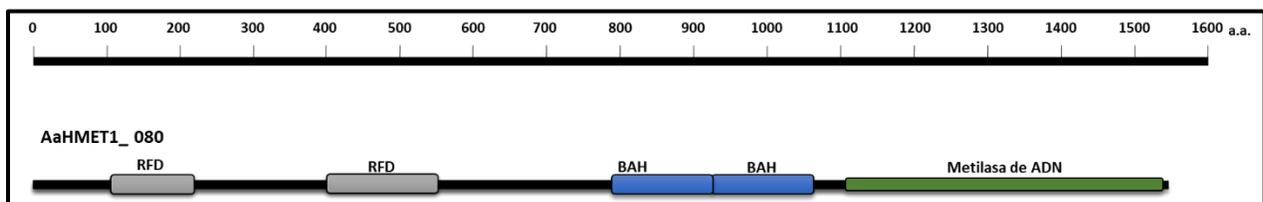


Figura 2. 2 Dominios conservados en secuencias de aminoácidos de metiltransferasas de la familia Met. BAH: Dominio conservado de homología bromoadyacente. RFD: Dominios conservados de replicación de foci. Metilasa\_ADN: Dominio catalítico metilasa de ADN.

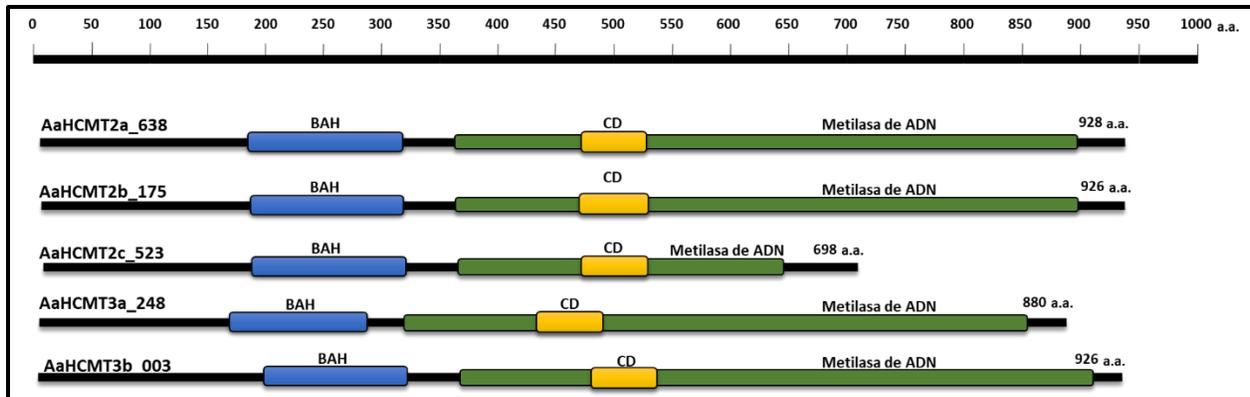


Figura 2. 3 Dominios conservados en secuencias de aminoácidos de metiltransferasas de la familia CMT. BAH: Dominio conservado de homología bromoadyacente. CD: Cromodominio. Metilasa\_ADN: Dominio catalítico metilasa de ADN.

### 2.3.7. Análisis filogenético de las secuencias identificadas en el transcriptoma

En la **Figura 2.5** se muestra el árbol filogenético obtenido de las secuencias de aminoácidos de los genes identificados en el transcriptoma de *A. angustifolia* Haw., las secuencias de aminoácidos de metiltransferasas de ADN en 14 especies reportadas en el trabajo de Moglia *et al.*, (2019) y las secuencias de metiltransferasas de *Asparagus officinalis*. Durante el análisis, se formaron tres clados principales que corresponden a las tres familias de metiltransferasas de ADN: MET, DRM y CMT, cada uno de los clados principales tiene un soporte del 100% para la prueba de bootstrap de 1,000 iteraciones. El total de las secuencias identificadas en el transcriptoma de *A. angustifolia* Haw. se encuentran en el clado que pertenece a la familia de metiltransferasas de ADN esperada. Además, las proteínas AaH\_MET1\_080, AaH\_DRM2a\_201, AaH\_DRM2b\_739, AaH\_DRM2c\_275, AaH\_DRM3d\_086, AaH\_CMT3a\_248, AaH\_CMT3b\_003 se organizan cerca de las secuencias de aminoácidos de proteínas ortólogas de *A. officinalis*. Sin embargo, las proteínas AaH\_CMT2a\_638, AaH\_CMT2b\_175, AaH\_CMT2c\_523, se distribuyen en el mismo clado que las proteínas ortólogas CMT1 de la especie *Brachypodium distachyon*, CMT1 de la especie *Sorghum bicolor* y CMT2 de la especie *O. sativa*.

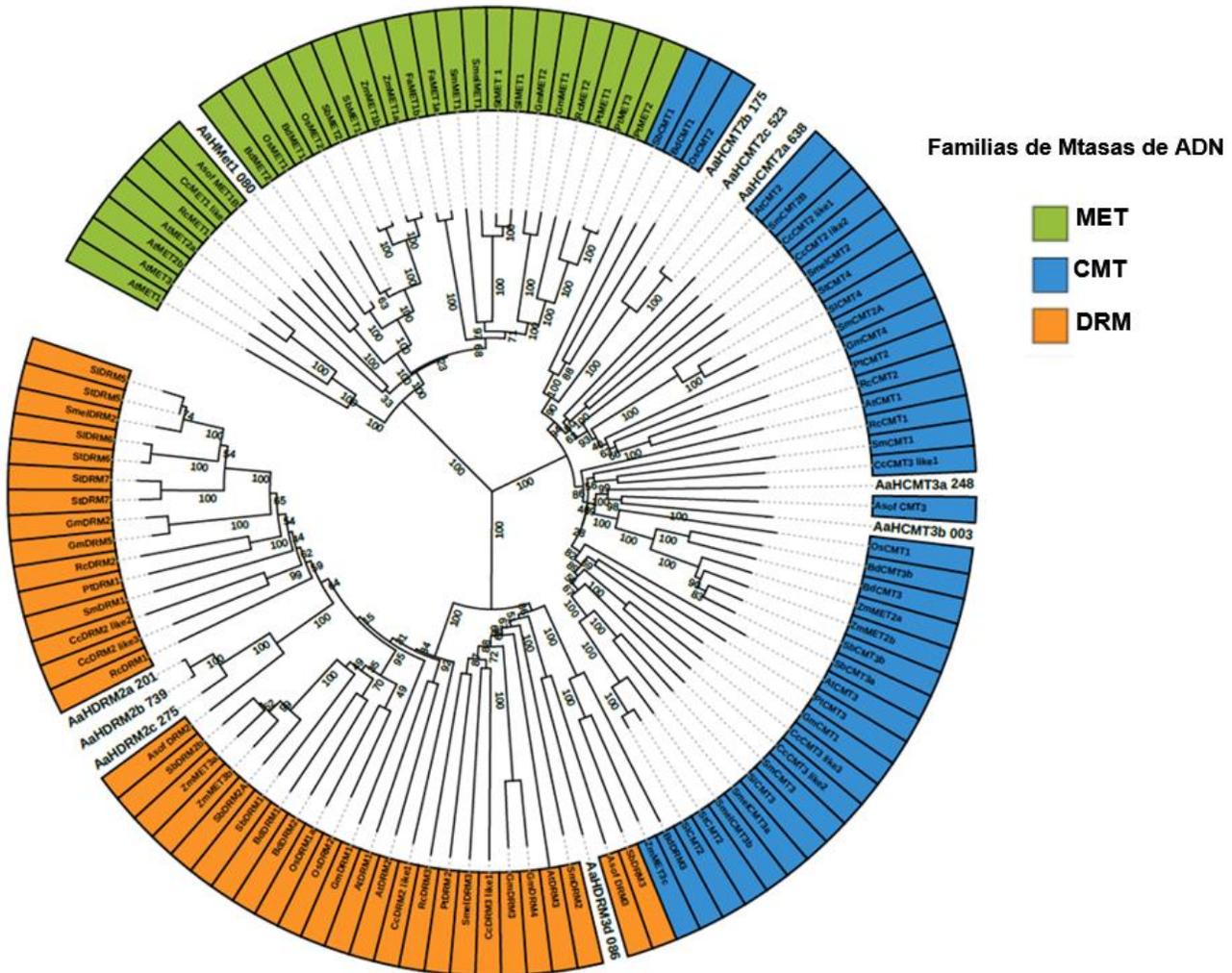


Figura 2. 4 Árbol filogenético de proteínas de MTasas de ADN. Se construyó utilizando el método de neighbor-joining. Los colores (verde, morado y naranja) indican los tres clados principales obtenidos para las familias de MTasas de ADN: MET, CMT y DRM, respectivamente. El soporte de las ramas se muestra con porcentaje de bootstrap para 1,000 réplicas. AaH: *Agave angustifolia* Haw., Asof: *Asparagus officinalis*, Bd: *Brachypodium distachyon*, Cc: *Cynara cardunculus*, Fa: *Fragaria x anassa*, Gm: *Glycine max*, Os: *Oryza sativa*, Pt: *Populus trichocarpa*, Rc: *Ricinus communis*, Sm: *Salvia miltiorrhiza*, Sl: *Solanum lycopersicum*, Smel: *Solanum melongena*, St: *Solanum tuberosum*, Sb: *Sorghum bicolor*, Zm: *Zea mays*.

## CAPÍTULO III

### ANÁLISIS DE EXPRESIÓN DE GENES DE METILTRANSFERASAS DE ADN

#### 3.1. INTRODUCCIÓN

En estudios anteriores realizados por Us-Camas *et al.*, (2017), se comprobó que la variación genética entre los fenotipos verde, variegado y albino de *A. angustifolia* Haw., era menor al 4%, concluyendo que la variación fenotípica observada en el modelo de estudio podría deberse a factores epigenéticos en vez de genéticos. Por otro lado, Duarte-Aké *et al.*, (2016) encontró que durante la micropropagación del fenotipo verde se observaba una hipometilación justo antes de que emergiera el fenotipo variegado y este a su vez disminuía la metilación en el ADN dando lugar al fenotipo albino. De igual manera, existen reportes que sugieren que las metilaciones en genes específicos que son necesarios para la biogénesis de plastidios, pueden estar involucrados en la aparición de fenotipos variegado y albino (Q. Wang *et al.*, 2015; Q. M. Wang *et al.*, 2016). Con base a estos estudios se propone que la variación fenotípica encontrada en *A. angustifolia* Haw. se puede explicar por una regulación epigenética mediada por metiltransferasas de ADN en genes específicos dando lugar a los fenotipos variegados y albinos.

Con el objetivo de conocer el nivel de expresión de los genes de metiltransferasa de ADN que fueron encontrados en el transcriptoma de *A. angustifolia* Haw., en este capítulo de la tesis se presentan los resultados de expresión evaluados por RT-qPCR de cuatro genes que codifican para metiltransferasas de ADN en los fenotipos G, V y A en los meristemos y hojas. Además, se detallan los análisis de expresión diferencial en la zona albina y la zona verde de la hoja variegada.

## 3.2. MATERIALES Y MÉTODOS

### 3.2.1. Diseño de cebadores

Se diseñaron cebadores para amplificar secuencias específicas de genes que codifican para enzimas de metiltransferasas de ADN, para ello, se utilizaron las secuencias encontradas en el transcriptoma de *A. angustifolia* Haw. Para el diseño de los cebadores se contempló que la región amplificada contuviera un fragmento de la región del dominio catalítico de cada gen de metiltransferasa de ADN y otro fragmento por fuera del dominio. Además, dicha región tiene una longitud esperada de 150pb a 200pb. Se utilizó la herramienta del NCBI de PRIMER-BLAST con las recomendaciones de diseño aportadas por la Unidad de Síntesis y Secuenciación de ADN del Instituto de Biotecnología, UNAM. En la **Tabla 3.1** se reporta la secuencia de los cebadores diseñados

Tabla 3.1 Secuencias de cebadores diseñados para el análisis de expresión de genes que codifican para las enzimas metiltransferasas de ADN de *A. angustifolia* Haw (AaH).

Gen	Secuencia (5'--->3')	Tm °C	Tamaño de fragmento esperado (bp)
AaH_MET1B_080	MET1B_F: TGCCCTTGGTATGAACCAG	59.67	199
	MET1B_R: TTGCCCATGTACCACCACAT	59.59	
AaH_DRM2a_201	DRM2a_F: GAGAATGTCGCCCTAGCTCC	59.97	172
	DRM2b_R: TTGGAGGGATCGGTAGGAGG	60.11	
AaH_DRM2b_739	DRM2b_F: AAACAACTCCGCTCGTGGT	58.4	150
	DRM2b_R: CACCATAAGTAAGCAATGTC	54.3	
AaH_DRM2c_275	DRM2c_F: TGTGCTGCCGCTAGAAAAAG	56.4	150
	DRM2c_R: AAGATGTGTGCGGCTATCCC	56.4	
AaH_DRM3a_086	DRM3a_F: ATGCTTGAGGTGAACACCGA	58.93	135
	DRM3a_R: AAGCATACTGATGTCTGGGGG	60.11	
AaH_CMT2a_638	CMT2a_F: TTCGGGAAGTACCATGTGA	58.95	153
	CMT2a_R: ACCCGAGCTCCTAAGCAAAG	59.75	
AaH_CMT2b_175	CMT2b_F: CCGGAACAAGACAGATCCT	59.03	197
	CMT2b_R: ATCAAAAGGGGGCCATCTCC	59.74	
AaH_CMT2c_523	CMT2c_F: TACAACCAAACGACTGCAACT	58.37	176
	CMT2c_R: GAGGACCCCCACATAGAACA	58.42	
AaH_CMT3a_248	CMT3a_F: TACAACCAAACGACTGCAACT	58.37	176
	CMT3a_R: GAGGACCCCCACATAGAACA	58.42	
AaH_CMT3b_003	CMT3b_F: TTGTTCTCAGGGTGTGGTGC	60.47	134
	CMT3b_R: TCTGGGTGATTGTGCCTGAG	59.67	
Gen 18S	18S_F: GCTACCACATCCAAGGAAG	60.7	135
	18S_R: TCGTTAAGGGATTAGATTGT	57.8	

### 3.2.2. Condiciones de cultivo y características de las plántulas

Las plantas de cada fenotipo, G, V y A, fueron cultivadas en cajas Magenta en grupos de seis plantas por fenotipo. Se utilizaron 50 ml de medio de cultivo para micropropagación MS-B reportado por Robert *et al.*, (2006), el cual está formulado con medio MS (Murashige y Skoog, 1962) modificado en la concentración de nitrógeno con 10mM de KNO<sub>3</sub> y 5mM de NH<sub>4</sub>NO<sub>3</sub>. En adición, este medio es suplementado con reguladores de crecimiento, con una concentración a 0.11µM de 2,4-D (ácido 2,4 diclorofenoxiacético) y 4.43 µM de 6-BA (6-benciladenina). Para solidificar el medio se utilizó agar (0.2%) y gelrite (0.2%). Las plantas se mantuvieron en un cuarto de cultivo con una temperatura de 27 ± 2°C, bajo condiciones de fotoperiodo de 12 h, con una intensidad luminosa de 40 µmol m<sup>-2</sup>s hasta que alcanzaron una altura de 2.5 cm aproximadamente. Las hojas de las plántulas V fueron diseccionadas utilizando un bisturí en las zonas verde (GV) y albina (AV) para realizar los análisis de expresión.

### 3.2.3. Extracción de ARN

Se extrajo ARN total de cada muestra (HG, HA, MG, MV, MA, GV y AV) por triplicado. Para este paso se empleó el método de TRI Reagent siguiendo las indicaciones del fabricante (Sigma). Primeramente, se maceraron 100 mg de muestra con nitrógeno nítrico con ayuda de un mortero y pistilo. Se agregó 1ml de TRI Reagent y se mezcló en vortex por unos segundos. Posteriormente, se incubó la muestra por 5 min a temperatura ambiente. A continuación, se agregaron 100 µl de BCP(1-bromo-3-cloropropano) y se mezcló en vortex por unos segundos. Para continuar, se incubaron las muestras por 15 min a temperatura ambiente. Se centrifugaron las muestras a 12,000xg por 15 min a 4°C. Después, se transfirió la fase acuosa a un tubo limpio de 1.5 ml y se añadió 0.5 ml de isopropanol frío (almacenado a -20°C). Se mezcló manualmente de forma ligera y se incubó la muestra a temperatura ambiente por 10 min. Posteriormente, se centrifugó a 12,000xg por 8 min a 4°C. A continuación, se descartó la fase acuosa volteando el tubo, conservando la pastilla de ARN en el tubo. Se lavó la pastilla con 1 ml de etanol frío al 75%. Se centrifugaron los tubos a 7500xg por 5 min a 4°C. Se descartó el sobrenadante y se dejó secar el contenido por 10min. Por último, cada muestra de ARN se re suspendió con 30 µl de agua ultra pura y se incubó a 55°C por 15 min. Las muestras de ARN se almacenaron a -80°C. Se verificó la integralidad del ARN en un gel de agarosa a 1.5% y se cuantificó su concentración en un nanodrop.

### 3.2.4. Síntesis de ADNc

Para la síntesis de ADNc se utilizó el kit de SuperScript IV siguiendo las indicaciones del proveedor (Thermo Fisher Scientific). Se utilizó 1 µg de ARN de cada muestra, se agregó 1 µl de Oligo d(T) 50 mM, 1 µl de dNTPs (10mM) y se aforó con agua ultra pura hasta alcanzar 13 µl. A continuación, la muestra se mantuvo por 5 min a 65°C para desnaturalizar el ARN y se colocó en hielo. Se añadió a la mezcla 4 µl de buffer de reacción incluido en el kit, 1 µl de DTT (100mM), 1 µl de ARNasa y 2 µl de la enzima transcriptasa inversa. Posteriormente, la mezcla de reacción se incubó a 42°C por una hora. Por último, se incubó a 80°C por 5 minutos. El ADNc se almacenó a -20°C hasta su uso.

### 3.2.5. PCR punto final

Se realizó un ensayo de PCR punto final para corroborar que los cebadores amplificaban correctamente la región de interés. Se utilizó el protocolo DreamTaq del fabricante ThermoFisher para PCR punto final. Los reactivos utilizados para un volumen final de 25 µl son los siguientes: 2.5 µl de buffer DreamTaq, 0.5 µl de cebador forward a 10 mM, 0.5 µl de cebador reverse a 10 mM, 1 µg de ADNc y 0.625 µl de polimerasa DreamTaq. Por último, se ajustó el volumen de reacción con agua ultra pura hasta obtener un volumen final de 25 µl. El programa utilizado para la técnica de PCR punto final fue el siguiente: desnaturalización inicial a 95°C por 3 min, a continuación, se utilizaron 25 ciclos a 95°C por 30 seg, una Tm específica para cada juego de cebadores (**Tabla 3.1**) por 30 seg y 72°C por un minuto para la extensión final. Se corrió un gel de agarosa al 1.5% para observar el tamaño y verificar que no existiera amplificaciones inespecíficas.

### 3.2.6. Determinación del nivel de expresión por densitometría de genes de metiltransferasas de ADN de *A. angustifolia* Haw.

Con el objetivo de conocer de manera preliminar la expresión de los genes de metiltransferasas de ADN, se corrió un gel de electroforesis utilizando el gen ribosomal 18S como gen normalizador y testigo de carga, con los productos de amplificación de PCR de cada juego de cebadores y con muestras de los siete tejidos: hoja verde (HG) y meristemo verde (MG), hoja albina (HA) y meristemo albino (MA), zona verde y zona albina de hoja variegada (VA) y el meristemo del fenotipo variegado (MV). Posteriormente, se realizó un análisis densitométrico con el programa ImageJ.

### 3.2.7. Elección de genes de metiltransferasas de ADN para analizar por RT- qPCR

Después de analizar el gel de electroforesis con el producto de amplificación de PCR de los 10 genes de metiltransferasas de ADN, se eligieron cuatro genes expresados diferencialmente y que fueran representativos de las tres familias de metiltransferasas MET, DRM y CMT. Los genes elegidos fueron *AaH\_MET1B\_080*, *AaH\_DRM2a\_201*, *AaH DRM2b\_739*, *AaHCMT3a\_248*.

### 3.2.8. Análisis de expresión de metiltransferasas a partir del reporte del transcriptoma

Para corroborar los datos de expresión de metiltransferasas de ADN obtenidos a partir de RT-qPCR, se graficaron los niveles de expresión en FPKM (Fragmentos por millón) reportados en el transcriptoma de los cuatro genes elegidos para su análisis. La unidad FPKM se utiliza para análisis de expresión ARN-seq donde la secuenciación es de cadena doble (pair-end), ya que toma en cuenta que 2 fragmentos pueden pertenecer a la misma lectura y por lo tanto no se cuenta dos veces. Para obtener los FPKM, se cuenta el número total de lecturas en una muestra y se divide ese número entre 1 millón, la cifra que resulta se conoce como factor de escala “por millón”. Posteriormente se divide el número de lecturas entre el factor de escala “por millón”, por último, el resultado se divide entre la longitud del gen en kilobases.

### 3.2.9. Medición de niveles de expresión de metiltransferasas de ADN por RT-qPCR

El ARN total de cada fenotipo (G, V y A) se extrajo usando el método TRI Reagent siguiendo las especificaciones del proveedor (Sigma). Se verificó la integridad del ARN en un gel de agarosa al 1.5%. Para la síntesis del ADNc se utilizó el kit SuperScript III First-Strand Synthesis System (Invitrogen) comenzando con 1 µg de ARN total hasta un volumen final de 25µl. Para las reacciones de RT-qPCR se utilizó 0.125 µl de cada cebador (5 µM), 6.25 µl de PCR SYBR Master Mix (Applied biosystems) y una concentración de 20 ng/µl de ADNc en un volumen final de reacción de 12.5 µl. Para la amplificación se utilizó el siguiente procedimiento: Desnaturalización inicial de 95°C por 10 min, seguido de 40 ciclos a 95°C por 40 seg, una T<sub>m</sub> específica de cada juego de cebadores (**Tabla 3.1**) por 40 seg y 72°C por 1 min. Como gen normalizador se utilizó el gen ribosomal 18S siguiendo el programa de amplificación: desnaturalización inicial a 95°C por 10 min, seguido de 40 ciclos a 95°C por 40 seg, 62°C por 40 seg y 72°C por 2 min.

### 3.3. RESULTADOS

#### 3.3.1. Determinación del nivel de expresión por densitometría de genes de metiltransferasas de ADN de *A. angustifolia* Haw.

Se hizo el análisis densitométrico de las bandas amplificadas por PCR punto final para todos los genes (**Figura 3.1**). Se logró identificar un único transcrito de la familia *MET*, *AaH\_MET1B\_080* (**Figura 3.1**). La expresión de este transcrito fue muy similar en hoja verde y albina, así como en la zona albina de la hoja variegada. Sin embargo, su expresión fue cincuenta por ciento mayor en la zona verde de la hoja variegada. En adición, su expresión fue muy semejante entre el meristemo verde y meristemo albino. Por último, el gen *AaH\_MET1B\_080* tuvo una expresión aproximadamente sesenta por ciento mayor en meristemo variegado que en meristemo verde y albino.

En la familia de genes *DRM*, *AaH\_DRM2a\_201*, mostró un nivel de expresión similar en muestras de hoja verde en comparación con muestras de hoja albina, así como en muestras de tejido verde y albino de plantas variegadas (**Figura 3.2**). La expresión del transcrito en los meristemos verde y albino fue muy similar, sin embargo, la expresión en el meristemo variegado fue 50% mayor. El gen *AaH\_DRM2b\_739* se expresó al doble en hoja albina en comparación con hoja verde, además el gen tuvo una expresión similar en la zona verde y albina de plantas con fenotipo variegado, esta expresión fue 3 veces mayor que la expresión en hoja verde. Por otro lado, se encontró una expresión tres veces mayor en meristemo verde en comparación con meristemo albino, y su expresión en meristemo variegado fue 1.5 veces mayor que en meristemo verde y por lo tanto 4.5 veces mayor que en meristemo albino. Por otro lado, *AaH\_DRM2c\_275*, mostró una expresión 50% mayor en hoja albina en comparación con hoja verde. Además, se observó que este transcrito se expresa de igual manera en la hoja albina y en la zona albina de hoja variegada. Sin embargo, en la zona verde se expresa 50% más. En los tejidos meristemáticos, *AaH\_DRM2c\_275* se expresó 50% más en meristemo verde que en meristemo albino, por último, el gen se expresó aproximadamente 30% más en meristemo variegado que en meristemo verde. En el caso del transcrito identificado como *AaH\_DRM3a\_086*, se encontró que su expresión fue muy similar entre hoja verde y hoja albina. De manera interesante se observó que la expresión del gen fue aproximadamente 50% menor en la zona albina de hojas en plantas con fenotipo variegado, además, la expresión de *AaH\_DRM3a\_086* fue aproximadamente 1.5 veces mayor en la zona verde de hojas variegadas que en hojas verdes y albinas.

En el caso de la familia CMT, el transcrito *AaH\_CMT2a\_638* mostró una expresión similar en hojas con fenotipo verde y albino, así como en la zona albina de hojas variegadas (**Figura 3.2**). Sin embargo, en la zona verde de hojas con fenotipo variegado, *AaH\_CMT2a\_638* mostró una expresión 1.5 veces mayor. Por otro lado, en los tejidos meristemáticos, este transcrito se expresó de manera muy parecida en meristemo verde y meristemo albino, mientras que el meristemo variegado fue de aproximadamente 60% más que en meristemo verde y albino. Para el caso de *AaH\_CMT2b\_175* se observó una expresión muy similar en los tejidos de hoja verde y albina. Sin embargo, en la zona verde de hojas variegadas, *AaH\_CMT2b\_175* tuvo una expresión 50% que, en el tejido de hoja verde y albina, además en la zona albina de hoja variegada se expresó tres veces menos. Por otro lado, *AaH\_CMT2c\_523* presentó una expresión similar en el tejido foliar verde, albino y en la zona albina de hoja variegada, sin embargo, la zona verde de hoja variegada se expresó cincuenta por ciento más. Además, *AaH\_CMT2c\_523* tuvo una expresión muy similar en meristemo verde y meristemo albino, por otro lado, el gen tuvo una expresión 50% mayor en el meristemo variegado. Otro de los genes analizados por densitometría en gel de electroforesis fue *AaH\_CMT3a\_248*, su amplificación fue muy similar en hoja verde y albina, así como en la zona albina de hoja variegada (**Figura 3.1**), sin embargo, su expresión fue 50% mayor en la zona verde de hoja variegada (**Figura 3.2**). En los tejidos meristemáticos *AaH\_CMT3a\_248* tuvo una expresión 50% mayor en meristemo verde que en meristemo albino y en meristemo variegado su expresión fue 60% mayor que en meristemo verde. El último transcrito por analizar de la familia CMT fue *AaH\_CMT3b\_003*, el cual se expresó de manera muy similar en la zona albina de hoja variegada y en el tejido de hoja verde, por otro lado, el gen se expresó de forma semejante en hoja albina y en la zona verde de hoja variegada, esta expresión fue cincuenta por ciento mayor en comparación con la zona albina de hoja variegada y el tejido de hoja verde. En adición, se observó una expresión similar del gen *AaH\_CMT3b\_003* en los tejidos de meristemo verde y albino, mientras que su expresión en meristemo variegado fue 70% mayor.

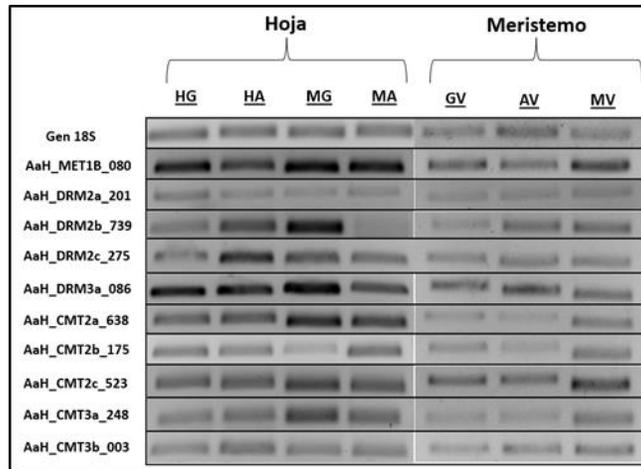


Figura 3. 1 Gel de electroforesis para genes de Mtasas de ADN identificados en el transcriptoma de *A. angustifolia* Haw. Para este análisis se realizó un PCR con los cebadores diseñados a partir de las secuencias del transcriptoma.

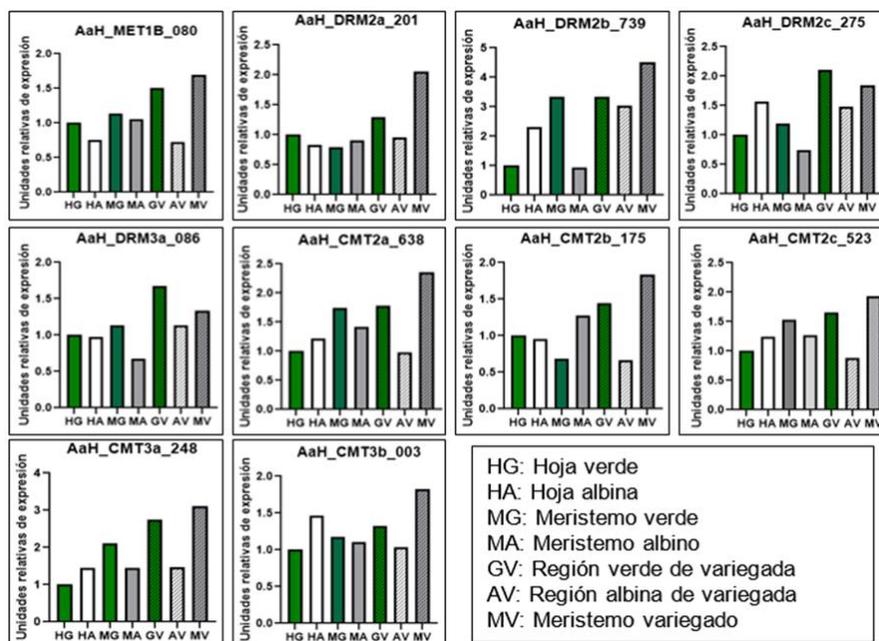


Figura 3. 2 Análisis densitométrico de genes de MTasas de ADN identificados en el transcriptoma de *A. angustifolia* Haw. El análisis se realizó con el programa ImageJ. Se utilizó el gen ribosomal 18S como gen normalizador y testigo de carga. Se utilizaron muestras de ARN de tejidos de hoja y meristemo de los fenotipos verde, variegado y albino.

### 3.3.2. Niveles de expresión de genes de metiltransferasas de ADN reportados en el transcriptoma *A. angustifolia* Haw.

Con la finalidad de corroborar los datos de expresión obtenidos por RT-qPCR de los cuatro genes de metiltransferasas de ADN seleccionados, primeramente, se graficaron los datos de FPKMs (Fragmentos por millón de kilobases) reportados en el transcriptoma de *A. angustifolia* Haw. de estos transcritos en la **Figura 3.3**.

Se observó que en general el nivel de expresión de *AaH\_MET1B\_080*, *AaH\_DRM2b\_739* y *AaH\_CMT3a\_248* fue más alto en las hojas con fenotipo verde que en las hojas con el fenotipo albino y solamente el gen *Aa\_HDRM2a\_201* tuvo una expresión más alta en el fenotipo albino (**Figura 3.3**). En el caso del tejido foliar, *AaHMET1B\_080*, *AaH\_DRM2b\_739* tuvieron una expresión similar en el fenotipo verde y albino y *AaH\_DRM2a\_201* se expresó aproximadamente dos veces más en el tejido albino en comparación con el tejido verde. Por último, *AaH\_CMT3a\_248* se expresó cinco veces más en el tejido verde que en el tejido albino. Por otro lado, *AaH\_MET1B\_080*, y *AaH\_CMT3a\_248* tuvieron una expresión aproximadamente tres veces mayor en el meristemo de plantas con fenotipo verde con respecto al meristemo de plantas con fenotipo albino. *AaH\_DRM2a\_201* y *AaH\_DRM2b\_739* se expresaron de manera muy similar en el meristemo de ambos fenotipos.

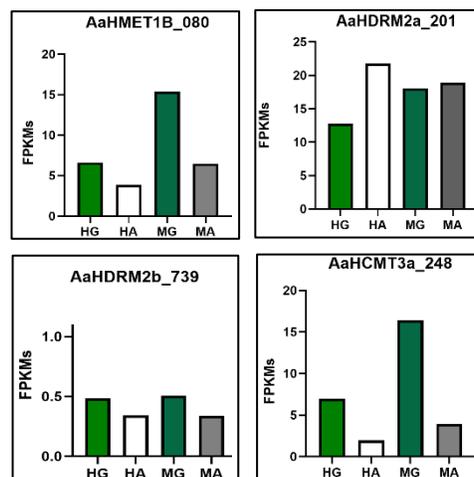


Figura 3. 3 Niveles de expresión de transcritos expresados diferencialmente reportados en el ensamble de transcriptoma de *A. angustifolia* Haw. Los datos fueron tomados directamente del reporte generado durante el análisis de expresión del transcriptoma de *A. angustifolia*.

### 3.3.3. Análisis de expresión de genes que codifican para metiltransferasas de ADN por RT-qPCR

El análisis de expresión por RT-qPCR indica que el nivel de expresión de los genes fue significativamente mayor en las hojas con fenotipo verde que en las hojas con fenotipo albino (**Figura 3.4**). Este patrón se repite en la expresión de los genes en el meristemo de plantas con fenotipo verde que es mayor a la expresión presente en el meristemo de plantas albinas. Los transcritos *AaH\_MET1\_080*, *AaH\_DRM2a\_201* y *AaH\_DRM2b\_739*, se expresaron aproximadamente dos veces más en las hojas con fenotipo verde que en las hojas con fenotipo albino. La diferencia más importante se observó en el gen *AaH\_CMT3a\_248*, el cual se expresó cinco veces más en el tejido de hojas con fenotipo verde con respecto a las hojas con fenotipo albino.

En los cuatro transcritos presentados en la **Figura 3.4** se detectó un aumento importante de más del doble en la expresión en el meristemo verde con respecto al meristemo albino. Nuevamente, *AaH\_CMT3a\_248* tuvo la diferencia de expresión más importante debido a que se expresó cinco veces más en el meristemo verde con respecto al meristemo albino. Además, los cuatro genes analizados tuvieron una expresión similar entre el meristemo albino y meristemo variegado. De forma interesante, se detectó una disminución de más del cincuenta por ciento en la expresión de los genes en las muestras de tejido foliar y meristemático de plantas con fenotipo variegado con respecto a las muestras restantes. *AaH\_MET1\_080*, *AaH\_DRM2a\_201* y *AaH\_CMT3a\_248* se expresaron de manera muy semejante tanto en la zona verde como en la albina de las hojas con fenotipo variegado, Sin embargo, se logró detectar un aumento del doble de expresión del gen *AaH\_DRM2b\_739* en la zona albina con respecto a la zona verde de las hojas variegadas.

Los patrones de expresión de los genes de metiltransferasas de ADN observados en el análisis por RT-qPCR fueron muy similares a los reportados en el transcriptoma de *A. angustifolia* Haw., sin embargo, los niveles de expresión obtenidos por densitometría, difieren en algunos casos con los encontrados por RT-qPCR y los reportados en el transcriptoma.

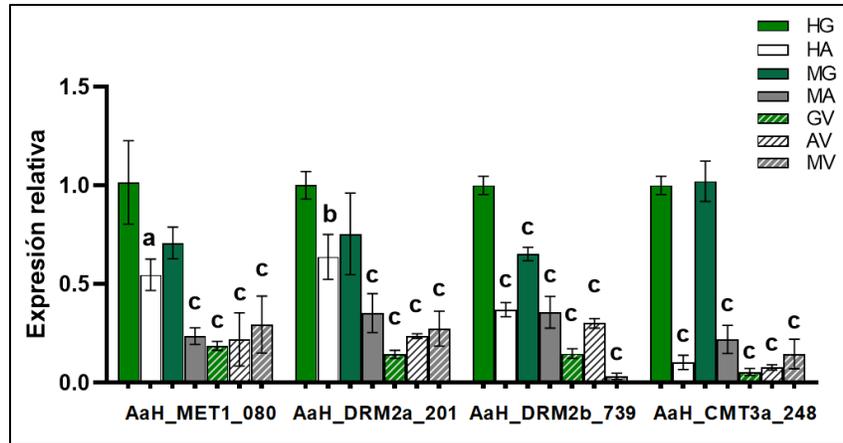


Figura 3. 4 Expresión relativa de los transcritos que codifican para metiltransferasas de ADN de *A. angustifolia* Haw. (AaH\_MET1, AaH\_DRM2a, AaH\_DRM2b, AaH\_CMT3a) por RT-qPCR. AaH: *Agave angustifolia* Haw., MET1: Metiltransferasa 1, DRM: Metiltransferasa de dominios rearrreglados, CMT: Cromometiltransferasa, HG: Hoja verde, HA: Hoja albina, MG: Meristemo verde, MA: Meristemo albino, GV: Región verde de planta variegada, AV: Región albina de planta variegada, MV: Meristemo variegado. Las diferencias significativas entre el tejido HG y los demás tejidos fue obtenida mediante la prueba estadística de Tukey utilizando diferentes valores de significancia (a $P$ ≤0.05, b $P$ ≤0.01, c $P$ ≤0.0001, respectivamente).



---

## CAPITULO IV

### DISCUSIÓN, CONCLUSIONES GENERALES Y PERSPECTIVAS

#### 4.1. DISCUSIÓN GENERAL

##### 4.1.1. Identificación de dominios conservados y construcción de árbol filogenético

Actualmente, las características y funciones de los genes de metiltransferasas de ADN en *Arabidopsis thaliana* han sido estudiadas ampliamente (Pavlopoulou & Kossida, 2007; Guo *et al.*, 2020). Asimismo, en los últimos cuatro años, se han identificado y caracterizado genes de metiltransferasas de ADN en plantas de interés comercial como: *Camelia sinensis* (planta de té) (Zhu, *et al.*, 2020), *Cynara scolymus* (alcachofa) (Gianoglio *et al.*, 2017), *Gingko biloba* (Guo *et al.*, 2020), *Actinidia chinensis* (kiwi) (Zhang *et al.*, 2020), *Ricinus communis* (ricino) (Victoria *et al.*, 2018) y *Solanum melongena* (berenjena) (Moglia *et al.*, 2019). Sin embargo, no existen estudios previos sobre las características y niveles de expresión de genes de metiltransferasas de ADN presentes en *A. angustifolia* Haw.

La búsqueda de dominios conservados en las secuencias de aminoácidos de metiltransferasas de ADN de *A. angustifolia* Haw. reveló que los diez genes poseen una región regulatoria y una catalítica (**Figuras 2.1, 2.2 y 2.3**). El análisis sugiere que el dominio metilasa de ADN es altamente conservado en las secuencias, mientras que la región reguladora contiene dominios diferentes característicos de cada familia de metiltransferasas de ADN.

De acuerdo con la disposición de las secuencias en el árbol filogenético (**Figura 2.4**), se forman tres clados principales que pertenecen a las familias de metiltransferasas MET, CMT, DRM de manera consistente con otros estudios (Gianoglio *et al.*, 2017; Moglia *et al.*, 2019). Además, la ubicación de las secuencias de proteínas de metiltransferasas de ADN de *A. angustifolia* Haw. corresponde con los resultados de búsquedas de similitud de secuencia, confirmando su clasificación basada en la identificación de sus dominios conservados. Por último, la mayoría de las secuencias homólogas de metiltransferasas de ADN se distribuyeron en el clado correspondiente a su familia, esta característica sugiere la conservación de su función como metiltransferasa de mantenimiento o *de novo* (Moglia *et al.*, 2019).

#### 4.1.2. Nivel de expresión de genes que codifican para enzimas metiltransferasas de ADN

El nivel de expresión de los genes de metiltransferasas de ADN de *A. angustifolia* Haw. presentaron un patrón muy similar en los datos reportados en el transcriptoma (**Figura 3.3**) y en los datos obtenidos de durante el análisis de expresión por RT-qPCR (**Figura 3.4**), con lo cual se corrobora los datos reportados en el ensamble del transcriptoma de *A. angustifolia* Haw. Sin embargo, los niveles de expresión obtenidos a partir del análisis por densitometría de gel de electroforesis (**Figuras 3.1 y 3.2**) fueron un poco distintos a los reportados en el ensamble del transcriptoma de *A. angustifolia* Haw (**Figura 3.3**) y los niveles de expresión detectados con la técnica de RT-qPCR (**Figura 3.4**). Esto podría deberse a una pérdida de muestra por evaporación durante el proceso de PCR punto final o al tiempo de cargar en el gel de electroforesis. Con lo que se concluye que la validación de los datos del transcriptoma por RT-qPCR son más precisos que el uso de PCR punto final.

En estudios anteriores realizados por Duarte-Aké *et al.*, (2016) se analizó el estado de metilación global de los tres fenotipos de *A. angustifolia* Haw. durante cinco ciclos de cultivo. Se observó que las plantas madre tenían diferencias en la metilación global, la planta variegada mostró 1% menos metilación que la planta verde, mientras que la planta albina contiene 2% menos metilación global que la planta verde. Esta diferencia en la metilación coincide con la disminución en la expresión de los genes de metiltransferasas identificada en el presente trabajo, en la que se observa menos expresión en los genes de las familias MET, DRM y CMT de las plantas albinas y variegadas en comparación con la planta verde. Por otro lado, se encontró que los brotes variegados emergían de las plantas verdes al tercer ciclo de cultivo coincidiendo con una hipometilación observada en las plantas madres. Los fenotipos variegados observados en plantas son principalmente debidos a mutaciones, silenciamiento por mecanismos de ARN, actividad de transposones y metilación de promotores (Liu, G., *et al.*, 2020). En el estudio realizado por Díaz-Martínez *et al.*, (2012), se describe que el fenotipo variegado de *A. tequilana* generado en la propagación asexual y en micropropagación está relacionado con cambios en los patrones de metilación del ADN en específico en secuencias CCGG. En *A. angustifolia* Haw. Se encontró una mayor expresión en genes que codifican para las MET1, las responsables en metilar en los sitios CCGG, en las plantas verdes (**Figura 3.4**), lo cual puede estar relacionado a una mayor metilación en estas plantas. En la micropropagación de *Clivia miniata* var. Variegata, es posible obtener plantas con el mismo contenido genético, pero con tres fenotipos diferentes: verdes, plantas variegadas que se caracterizan por poseer hojas con zonas verdes y amarillas

que se extienden desde la base hasta la punta y plantas amarillas. En un estudio realizado por Wang *et al.*, (2016) se encontraron evidencias de que el cuerpo del gen *CPSAR1* (GTPasa de cloroplasto asociada a secreción Ras1) en las hojas amarillas y *YCF2* (Proteína de ensamblaje del fotosistema 1) en la zona amarilla de las plantas variegadas se encuentran hipermetilados en sitios CG causando su represión. Los productos de ambos genes participan en la biogénesis de cloroplastos (Wang *et al.*, 2016). Además, el silenciamiento de *CPSAR1* por ADN de transferencia resulta en embriones albinos, fotosintéticamente incompetentes y que no llegan a un estado de maduración (García, *et al.*, 2010).

Se ha reportado la aparición natural de plantas con fenotipo albino en *O. Sativa*. Se encontró que este fenotipo es causado por cambios en los patrones metilación en las secuencias CG en un gen llamado *OsAK1* que codifica para una cinasa de adenilato. Consistentemente con el fenotipo observado, el contenido de clorofila *a*, clorofila *b* y clorofila total es mucho menor que en el fenotipo verde. Además, utilizando técnicas de microscopia de transmisión electrónica se observaron pocos cloroplastos indiferenciados y con un tamaño mucho menor en plantas albinas que plantas con fenotipo verde (Wei *et al.*, 2017). En los fenotipos variegados y albinos de *A. angustifolia* Haw. También se ha encontrado un bajo contenido no solamente de clorofila *a*, clorofila *b* y clorofila total sino también de carotenoides (Duarte-Aké *et al.*, 2016; Us-Camas *et al.*, 2017). También, Hernández-Castellano *et al.*, (2020) observaron un número importante de proplastidios en las plantas albinas de *A. angustifolia* Haw.

La mutante albina de trigo M-1376 se caracteriza en su fenotipo por tener hojas blancas, rojas y verdes. En un estudio realizado por Song *et al.*, (2018) se observó, mediante técnicas de microscopía, que las células del mesófilo en hojas verdes estaban ordenadas y contenían cloroplastos de manera abundante, las hojas rojas algunas células del mesófilo ordenadas y aún contenían cloroplastos. Sin embargo, las hojas albinas contenían células del mesófilo desorganizadas y pocos cloroplastos. Además, se comparó la estructura de los cloroplastos en las hojas utilizando microscopia de transmisión de electrones y se encontró que los cloroplastos de las hojas verdes contenían granas con las membranas de los tilacoides intactos, por el contrario, los pocos cloroplastos encontrados en las hojas con fenotipo albino contenían tilacoides sin membrana desarrollada. Por su parte, se analizó también el contenido de metilación en muestras de hojas con fenotipo verde y albino utilizando técnicas de marcadores moleculares MSAP (amplificación de polimorfismos sensible a metilaciones) y el perfil de metilación mostró una hipermetilación en los genes *psaA* y *psbD* en las plantas albinas. Ambos genes codifican para proteínas esenciales que forman parte del fotosistema 1 (PSI) y fotosistema 2 (PSII), a su

vez PSI y PSII, junto con la ATPasa y el citocromo b6f se ubican en la membrana de los tilacoides de los cloroplastos y participan en la fotosíntesis (Song *et al.*, 2018; Evans & Clarke, 2019). Asimismo, se evaluó el nivel de expresión de los genes *psaA* y *psbD* con la técnica de RT-qPCR y descubrieron que su expresión fue mínima en hojas albinas con respecto a hojas con fenotipo verde. Por consiguiente, es probable que los genes *psaA* y *psbD* sean susceptibles a una modulación por mecanismos de metilación en el ADN. En la presente tesis, se pudo observar un aumento en la expresión de los cuatro genes evaluados en la zona albina de las hojas variegadas con respecto a la zona verde (**Figura 3.4**). Es posible que las enzimas metiltransferasas de ADN están regulando negativamente la expresión de los genes *psaA* y *psbD* así como otros genes específicos y como consecuencia no se logre una biogénesis completa de cloroplastos en la zona albina de las hojas variegadas. Futuras investigaciones serán necesarias para comprobar esta teoría.

---

## 4.2 CONCLUSIONES GENERALES

- Se identificaron 10 transcritos que codifican para metiltransferasas de ADN en el transcriptoma de *A. angustifolia* Haw. utilizando herramientas de bioinformática, los cuales son secuencias únicas y completas.
- Se encontró la identidad y homología de las 10 secuencias de metiltransferasas de ADN realizando búsquedas de similitud en bases de datos públicas. Cada una de las secuencias presentan similitud con genes que codifican para metiltransferasas de la especie *Asparagus officinalis* de las familias MET, DRM y CMT.
- Para cada una de las secuencias de metiltransferasa se logró determinar la ubicación del dominio metilasa de ADN en el lado N-terminal y los dominios de regulación en el lado C-terminal. Cada uno de los dominios tienen una secuencia completa y concuerdan con el tipo de gen esperado.
- Se construyó un árbol filogenético utilizando las secuencias de aminoácidos de metiltransferasas de ADN identificadas en el transcriptoma de *A. angustifolia* Haw. y secuencias de proteínas de otras 15 especies de diferentes niveles evolutivos. El árbol resultante está formado por tres principales clados que pertenecen a las familias de metiltransferasas MET, DRM y CMT. Las secuencias de *A. angustifolia* Haw., se ubicaron en el clado esperado de acuerdo al tipo de gen y se agruparon cerca de secuencias de las especies *Asparagus officinalis*, *Brachypodium distachyon*, *Sorghum bicolor* y *Oryza sativa*.
- El análisis de expresión de los genes que codifican para metiltransferasas de ADN en *A. angustifolia* Haw. reveló diferencias entre los tres fenotipos. Se observó una menor expresión en los genes *MET1*, *DRM2*, *DRM3* y *CMT2b* en las hojas albinas en comparación con las hojas verdes, lo cual confirma que la hipometilación reportada en las plantas albinas está relacionada con la baja expresión de las metiltransferasas de ADN.

### 4.3 PERSPECTIVAS

- Analizar la expresión de los genes importantes en los fotosistemas como *psaA* y *psbD*.
- Realizar ensayos con inhibidores de metiltransferasas de ADN para conocer el efecto en la expresión en genes que codifican para metiltransferasas de ADN y genes relacionados con la biogénesis de cloroplastos.
- Dado que la metilación del ADN es un sistema complejo y dinámico mediado por enzimas de metilación llamadas metiltransferasas de ADN y enzimas que retiran grupos metilo llamadas demetilasas de ADN, sería interesante conocer el nivel de expresión de estas últimas en los tejidos de hoja y meristemo en plantas de los tres fenotipos de *A. angustifolia* Haw. utilizando análisis de PCR en tiempo real.
- Utilizar técnicas de edición genómica como CRISPR-Cas para identificar cuál de las metiltransferasas tiene un papel preponderante en las plantas albinas.

## BIBLIOGRAFÍA

- Davidson, N. M., & Oshlack, A. (2014). Corset: Enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biology*, 15(7), 1–14. <https://doi.org/10.1186/s13059-014-0410-6>
- Abadie, J. C., Püttsepp, Ü., Gebauer, G., Faccio, A., Bonfante, P., & Selosse, M. A. (2006). *Cephalanthera longifolia* (Neottieae, Orchidaceae) is mixotrophic: A comparative study between green and nonphotosynthetic individuals. *Canadian Journal of Botany*, 84(9), 1462–1477. <https://doi.org/10.1139/B06-101>
- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Ech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>
- Anamika, Mehta, S., Singh, B., Patra, A., & Islam, Md. A. (2019). Databases: A Weapon from the Arsenal of Bioinformatics for Plant Abiotic Stress Research. In *Recent Approaches in Omics for Plant Resilience to Climate Change*. [https://doi.org/10.1007/978-3-030-21687-0\\_7](https://doi.org/10.1007/978-3-030-21687-0_7)
- Anastasiadi, D., Venney, C. J., Bernatchez, L., & Wellenreuther, M. (2021). Epigenetic inheritance and reproductive mode in plants and animals. *Trends in Ecology and Evolution*, 36(12), 1124–1140. <https://doi.org/10.1016/j.tree.2021.08.006>
- Arzate-Fernández, A. M., & Mejía-Franco, R. (2011). Capacidad embriogénica de callos inducidos en ejes embrionarios cigóticos de *Agave angustifolia* Haw. *Revista Fitotecnia Mexicana*, 34(2), 101–106.
- Ashapkin, V. v., Kutueva, L. I., & Vanyushin, B. F. (2016). Plant DNA methyltransferase genes: Multiplicity, expression, methylation patterns. *Biochemistry (Moscow)*, 81(2), 141–151. <https://doi.org/10.1134/S0006297916020085>
- Azizi, P., Hanafi, M. M., Sahebi, M., Harikrishna, J. A., Taheri, S., Yassoralipour, A., & Nasehi, A. (2020). Epigenetic changes and their relationship to somaclonal variation: A need to monitor the micropropagation of plantation crops. *Functional Plant Biology*, Jiménez 2005. <https://doi.org/10.1071/FP19077>

- Baduel, P., & Colot, V. (2021). The epiallelic potential of transposable elements and its evolutionary significance in plants. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 376, Issue 1826). <https://doi.org/10.1098/rstb.2020.0123>
- Bairoch, A., & Apweiler, R. (1997). The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucleic Acids Research*, 25(1), 31–36. <https://doi.org/10.1093/nar/25.1.31>
- Bairu, M. W., Aremu, A. O., & van Staden, J. (2011). Somaclonal variation in plants: Causes and detection methods. *Plant Growth Regulation*, 63(2), 147–173. <https://doi.org/10.1007/s10725-010-9554-x>
- Barraza-Morales, A., Lorenzo Sánchez-Teyer, F., Robert, M., Esqueda, M., & Gardea, A. (2006). Variabilidad genética en *Agave angustifolia* Haw. de la Sierra Sonorense, México, determinada con marcadores AFLP. *Artículo Científico Rev. Fitotec. Mex*, 29(1), 1–8. <https://www.redalyc.org/pdf/610/61029101.pdf>
- Bártová, E., Krejčí, J., Harničarová, A., Galiová, G., & Kozubek, S. (2008). Histone modifications and nuclear architecture: A review. *Journal of Histochemistry and Cytochemistry*, 56(8), 711–721. <https://doi.org/10.1369/jhc.2008.951251>
- Basso, M. F., Ferreira, P. C. G., Kobayashi, A. K., Harmon, F. G., Nepomuceno, A. L., Molinari, H. B. C., & Grossi-de-Sa, M. F. (2019). MicroRNAs and new biotechnological tools for its modulation and improving stress tolerance in plants. *Plant Biotechnology Journal*,
- Batista, R. A., & Köhler, C. (2020). Genomic imprinting in plants-revisiting existing models. *Genes & Development*, 34(1–2), 24–36. <https://doi.org/10.1101/gad.332924.119>
- Bednarek, P. T., Orłowska, R., Koebner, R. M. D., & Zimny, J. (2007). Quantification of the tissue-culture induced variation in barley (*Hordeum vulgare* L.). *BMC Plant Biology*, 7, 1–9. <https://doi.org/10.1186/1471-2229-7-10>
- Begum, M., Islam, M., Miah, M., Hossain, M., & Islam, N. (2011). Production of Somaclone In vitro for Drought Stress Tolerant Plantlet Selection in Sugarcane (*Saccharum officinarum* L.). *The Agriculturists*, 9(1–2), 18–28. <https://doi.org/10.3329/agric.v9i1-2.9475>
- Bennett, M. D., & Leitch, I. J. (2011). Nuclear DNA amounts in angiosperms: Targets, trends and tomorrow. *Annals of Botany*, 107(3), 467–590. <https://doi.org/10.1093/aob/mcq258>

- Bilotta, M., Tradigo, G., & Veltri, P. (2018). Bioinformatics data models, representation and storage. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vols. 1–3, Issue 1). Elsevier Ltd. <https://doi.org/10.1016/B978-0-12-809633-8.20410-X>
- Bioinformatics, B., Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. v, Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. v, Vasudevan, S., Wolf, Y. I., Yin, J. J., & Natale, D. A. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, 4(4), 1–14.
- Bird, A. (2007). Perceptions of epigenetics. *Nature*, 447(7143), 396–398. <https://doi.org/10.1038/nature05913>
- Birnbaum, K. D., & Roudier, F. (2017). Epigenetic memory and cell fate reprogramming in plants. *Regeneration*, 4(1), 15–20. <https://doi.org/10.1002/reg2.73>
- Bobadilla Landey, R., Cenci, A., Georget, F., Bertrand, B., Camayo, G., Dechamp, E., Herrera, J. C., Santoni, S., Lashermes, P., Simpson, J., & Etienne, H. (2013). High Genetic and Epigenetic Stability in *Coffea arabica* Plants Derived from Embryogenic Suspensions and Secondary Embryogenesis as Revealed by AFLP, MSAP and the Phenotypic Variation Rate. *PLoS ONE*, 8(2). <https://doi.org/10.1371/journal.pone.0056372>
- Boquete, M. T., Muyle, A., & Alonso, C. (2021). Plant epigenetics: phenotypic and functional diversity beyond the DNA sequence. *American Journal of Botany*, 108(4), 553–558. <https://doi.org/10.1002/ajb2.1645>
- Borg, M., Jiang, D., & Berger, F. (2021). Histone variants take center stage in shaping the epigenome. *Current Opinion in Plant Biology*, 61. <https://doi.org/10.1016/j.pbi.2020.101991>
- Borges, F., Donoghue, M. T. A., LeBlanc, C., Wear, E. E., Tanurdžić, M., Berube, B., Brooks, A., Thompson, W. F., Hanley-Bowdoin, L., & Martienssen, R. A. (2021). Loss of Small-RNA-Directed DNA Methylation in the Plant Cell Cycle Promotes Germline Reprogramming and Somaclonal Variation. *Current Biology*, 31(3), 591-600.e4. <https://doi.org/10.1016/j.cub.2020.10.098>
- Bridgen, P. M. P., Skirvin, R. M., Mcpheeters, K. D., Norton, M., Veilleux, R. E., Grosser, J. W., Bridgen, M. P., Skirvin, R. M., Mcpheeters, K. D., Norton, M., & Drive, G. (1994). Workshop Papers and Authors Sources and Frequency of Somaclonal Variation. 29(1990), 1232–1237.

- Cao, X., & Jacobsen, S. E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proceedings of the National Academy of Sciences of the United States of America*, 99(SUPPL. 4), 16491–16498. <https://doi.org/10.1073/pnas.162371599>
- Cao, X., Springer, N. M., Muszynski, M. G., Phillips, R. L., Kaeppler, S., & Jacobsen, S. E. (2000a). Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. *Proceedings of the National Academy of Sciences of the United States of America*, 97(9), 4979–4984. <https://doi.org/10.1073/pnas.97.9.4979>
- Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L. P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang, X., Mushayahama, T., ... Elser, J. (2021). The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research*, 49(D1), D325–D334. <https://doi.org/10.1093/nar/gkaa1113>
- Cardone, S., & Olmos, S. (2010). *Biotecnología y mejoramiento vegetal II*. Capítulo 5. January.
- Chuang, J. C., & Jones, P. A. (2007). Epigenetics and microRNAs. *Pediatric Research*, 61(5 PART 2 SUPPL.), 24–29. <https://doi.org/10.1203/pdr.0b013e3180457684>
- Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., & Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature*, 452(7184), 215–219. <https://doi.org/10.1038/nature06745>
- Davis, S. C., Kuzmick, E. R., Niechayev, N., & Hunsaker, D. J. (2017). Productivity and water use efficiency of *Agave americana* in the first field trial as bioenergy feedstock on arid lands. *GCB Bioenergy*, 9(2), 314–325. <https://doi.org/10.1111/gcbb.12324>
- DChan, S. W. L., Henderson, I. R., & Jacobsen, S. E. (2005). Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature Reviews Genetics*, 6(5), 351–360. <https://doi.org/10.1038/nrg1601>
- de Mendoza, A., Bonnet, A., Vargas-Landin, D. B., Ji, N., Hong, F., Yang, F., Li, L., Hori, K., Pflueger, J., Buckberry, S., Ohta, H., Rosic, N., Lesage, P., Lin, S., & Lister, R. (2018). Recurrent

---

acquisition of cytosine methyltransferases into eukaryotic retrotransposons. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-03724-9>

Díaz-Martínez, M., Nava-Cedillo, A., Guzmán-López, J. A., Escobar-Guzmán, R., & Simpson, J. (2012). Polymorphism and methylation patterns in *Agave tequilana* Weber var. "Azul" plants propagated asexually by three different methods. *Plant Science*, 185–186, 321–330. <https://doi.org/10.1016/j.plantsci.2012.01.004>

Domínguez, M., González Jiménez, Ma., Rosales Gómez, C., Quiñones Valles, C., Delgadillo Díaz de León, S., Mireles Ordaz, S., & Pérez Molphe Balch, E. (2008). El cultivo *in vitro* como herramienta para el aprovechamiento, mejoramiento y conservación de especies del género *Agave*. *Investigación y Ciencia: De La Universidad Autónoma de Aguascalientes*, 41, 53.

Du, J., Johnson, L. M., Groth, M., Feng, S., Hale, C. J., Li, S., Vashisht, A. A., Gallego-Bartolome, J., Wohlschlegel, J. A., Patel, D. J., & Jacobsen, S. E. (2014). Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Molecular Cell*, 55(3), 495–504. <https://doi.org/10.1016/j.molcel.2014.06.009>

Duarte-Aké, F., Castillo-Castro, E., Pool, F. B., Espadas, F., Santamaría, J. M., Robert, M. L., & De-la-Peña, C. (2016). Physiological differences and changes in global DNA methylation levels in *Agave angustifolia* Haw. albino variant somaclones during the micropropagation process. *Plant Cell Reports*, 35(12), 2489–2502. <https://doi.org/10.1007/s00299-016-2049-0>

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>

Erdmann, R. M., & Picard, C. L. (2020). RNA-directed DNA Methylation. In *PLoS Genetics* (Vol. 16, Issue 10). <https://doi.org/10.1371/journal.pgen.1009034>

Evans, J. R., & Clarke, V. C. (2019). The nitrogen cost of photosynthesis. *Journal of Experimental Botany*, 70(1), 7–15. <https://doi.org/10.1093/jxb/ery366>

Finnegan, E. J., & Kovac, K. A. (2000b). Plant DNA methyltransferases. *Plant Molecular Biology*, 43(2–3), 189–201. [https://doi.org/10.1007/978-94-011-4183-3\\_5](https://doi.org/10.1007/978-94-011-4183-3_5)

Gajecka, M., Marzec, M., Chmielewska, B., Jelonek, J., Zbieszczak, J., & Szarejko, I. (2020). Plastid differentiation during microgametogenesis determines green plant regeneration in barley

microspore culture. *Plant Science*, 291(October 2019), 110321.  
<https://doi.org/10.1016/j.plantsci.2019.110321>

García, C., Khan, N. Z., Nannmark, U., & Aronsson, H. (2010). The chloroplast protein CPSAR1, dually localized in the stroma and the inner envelope membrane, is involved in thylakoid biogenesis. *Plant Journal*, 63(1), 73–85. <https://doi.org/10.1111/j.1365-313X.2010.04225.x>

García-Herrera, E. J., Méndez-Gallegos, S. J., & Talavera-Magaña, D. (2010). El genero *Agave* spp. en México: principales usos de importancia socioeconómica y agroecológica. *Revista Salud Pública y Nutrición, Edición Especial*, 5(73), 109–129.

García-Mendoza, A. (2002). Distribution of *Agave* (Agavaceae) in Mexico. *Cactus and Succulent Journal*, 74(4), 177–187.

Genger, R. K., Kovac, K. A., Dennis, E. S., Peacock, W. J., & Finnegan, E. J. (1999). Multiple DNA methyltransferase genes in *Arabidopsis thaliana*. *Plant Molecular Biology*, 41(2), 269–278. <https://doi.org/10.1023/A:1006347010369>

Gentry, H. S. (1982). *Agaves of Continental North America*. University of Arizona Press, Tucson, AZ.

Ghosh, A., Igamberdiev, A. U., & Debnath, S. C. (2021). Tissue culture-induced DNA methylation in crop plants: a review. *Molecular Biology Reports*, 48(1), 823–841. <https://doi.org/10.1007/s11033-020-06062-6>

Gianoglio, S., Moglia, A., Acquadro, A., Comino, C., & Portis, E. (2017). The genome-wide identification and transcriptional levels of DNA methyltransferases and demethylases in globe artichoke. *PLoS ONE*, 12(7), 1–17. <https://doi.org/10.1371/journal.pone.0181669>

Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. *Cell*, 128(4), 635–638. <https://doi.org/10.1016/j.cell.2007.02.006>

Gomez-Casati, D. F., Busi, M. v., Barchiesi, J., Peralta, D. A., Hedin, N., & Bhadauria, V. (2018). Applications of bioinformatics to plant biotechnology. *Current Issues in Molecular Biology*. <https://doi.org/10.21775/CIMB.027.089>

Grabherr, M. G. ; Brian J. Haas, Moran Yassour Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir

- 
- Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W., N., & Friedman, and A. R. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>. Trinity
- Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, 20(10), 590–607. <https://doi.org/10.1038/s41580-019-0159-6>
- Guo, X., Xie, Q., Li, B., & Su, H. (2020). Molecular characterization and transcription analysis of DNA methyltransferase genes in tomato (*Solanum lycopersicum*). *Genetics and Molecular Biology*, 43(1), 1–10. <https://doi.org/10.1590/1678-4685-GMB-2018-0295>
- Henderson, I. R., Deleris, A., Wong, W., Zhong, X., Chin, H. G., Horwitz, G. A., Kelly, K. A., Pradhan, S., & Jacobsen, S. E. (2010). The De novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in *Arabidopsis thaliana*. *PLoS Genetics*, 6(10), 1–11. <https://doi.org/10.1371/journal.pgen.1001182>
- Hernández-Castellano, S., Garruña-Hernández R., Us-Camas, R., Kú-Gonzalez, A., & De-la-Peña, C. (2020). *Agave angustifolia* albino plantlets lose stomatal physiology function by changing the development of the stomatal complex due to a molecular disruption. *Molecular Genetics and Genomics*, 295(3), 787–805. <https://doi.org/10.1007/s00438-019-01643-y>
- Holliday, R. (2006). Epigenetics: A historical overview. *Epigenetics*, 1(2), 76–80. <https://doi.org/10.4161/epi.1.2.2762>
- Huettel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A. J. M., & Matzke, M. (2006). Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO Journal*, 25(12), 2828–2836. <https://doi.org/10.1038/sj.emboj.7601150>
- Hung, Y. H., Liu, F., Zhang, X. Q., Xiao, W., & Hsieh, T. F. (2018). Sexual and non-sexual reproduction: Inheritance and stability of epigenetic variations and consequences for breeding application. In *Advances in Botanical Research* (Vol. 88). Elsevier Ltd. <https://doi.org/10.1016/bs.abr.2018.09.002>

- Illsley Granich, Catarina. *et al.* 2004. Manual de manejo campesino de magueyes mezcaleros. Grupo de Estudios Ambientales A.C., CONABIO. Rainforest Alliance. Fundación Ford. SSS Sanzekan Tinemi.
- Jean finnegan, E., & Dennis, E. S. (1993). Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Research*, 21(10), 2383–2388. <https://doi.org/10.1093/nar/21.10.2383>
- Jeddeloh, J. A., Stokes, T. L., & Richards, E. J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genetics*, 22(1), 94–97. <https://doi.org/10.1038/8803>
- Kaeppler, S. M., Kaeppler, H. F., & Rhee, Y. (2000). Epigenetic aspects of somaclonal variation in plants. *Plant Molecular Biology*, 43, 179–188.
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Research*, 49(D1), D545–D551. <https://doi.org/10.1093/nar/gkaa970>
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1), D457–D462. <https://doi.org/10.1093/nar/gkv1070>
- Kankel, M. W., Ramsey, D. E., Stokes, T. L., Flowers, S. K., Haag, J. R., Jeddeloh, J. A., Riddle, N. C., Verbsky, M. L., & Richards, E. J. (2003). *Arabidopsis* MET1 cytosine methyltransferase mutants. *Genetics*, 163(3), 1109–1122.
- Khan, I. A., Dahot, M. U., Seema, N., Yasmeen, S., Bibi, S., Raza, G., Khatri, A., & Naqvi, M. H. (2009). Direct regeneration of sugarcane plantlets: A tool to unravel genetic heterogeneity. *Pakistan Journal of Botany*, 41(2), 797–814. <https://doi.org/10.13140/RG.2.2.28819.22563>
- Kim, J. K., Samaranayake, M., & Pradhan, S. (2009). Epigenetic mechanisms in mammals. *Cellular and Molecular Life Sciences*, 66(4), 596–612. <https://doi.org/10.1007/s00018-008-8432-4>
- Kour, G., Kour, B., Kaul, S., & Dhar, M. K. (2009). Genetic and epigenetic instability of amplification-prone sequences of a novel B chromosome induced by tissue culture in *Plantago lagopus* L. *Plant Cell Reports*, 28(12), 1857–1867. <https://doi.org/10.1007/s00299-009-0789-9>

- Kumar, S., & Mohapatra, T. (2021). Dynamics of DNA Methylation and Its Functions in Plant Growth and Development. *Frontiers in Plant Science*, 12(May). <https://doi.org/10.3389/fpls.2021.596236>
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Kumari, M., Clarke, H. J., Small, I., & Siddique, K. H. M. (2009). Albinism in plants: A major bottleneck in wide hybridization, androgenesis and doubled haploid culture. *Critical Reviews in Plant Sciences*, 28(6), 393–409. <https://doi.org/10.1080/07352680903133252>
- Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 11(3), 204–220. <https://doi.org/10.1038/nrg2719>
- le Cam, L. (1990). Maximum Likelihood: An introduction. *International Statistical Review*, 25(6), 27–28. <https://doi.org/10.1255/nirn.1476>
- Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 323. <https://doi.org/https://doi.org/10.1186/1471-2105-12-323>
- Li, Xiaoling, Yu, X., Wang, N., Feng, Q., Dong, Z., Liu, L., Shen, J., & Liu, B. (2007). Genetic and epigenetic instabilities induced by tissue culture in wild barley (*Hordeum brevisubulatum* (Trin.) Link). *Plant Cell, Tissue and Organ Culture*, 90(2), 153–168. <https://doi.org/10.1007/s11240-007-9224-5>
- Liang, L., Chang, Y., Lu, J., Wu, X., Liu, Q., Zhang, W., Su, X., & Zhang, B. (2019). Global methylomic and transcriptomic analyses reveal the broad participation of DNA methylation in daily gene expression regulation of *Populus trichocarpa*. *Frontiers in Plant Science*, 10(February), 1–10. <https://doi.org/10.3389/fpls.2019.00243>
- Liberatore, C. M., Rodolfi, M., Beghè, D., Fabbri, A., Ganino, T., & Chiancone, B. (2020). *In vitro* leaf-derived organogenesis and somaclonal variant detection in *Humulus lupulus* L. *In Vitro Cellular and Developmental Biology - Plant*, 56(6), 865–874. <https://doi.org/10.1007/s11627-020-10088-7>

- Lindroth, A. M., Cao, X., Jackson, J. P., Zilberman, D., McCallum, C. M., Henikoff, S., & Jacobsen, S. E. (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science*, 292(5524), 2077–2080. <https://doi.org/10.1126/science.1059745>
- Liu, G., Li, C., Yu, H., Tao, P., Yuan, L., Ye, J., Chen, W., Wang, Y., Ge, P., Zhang, J., Zhou, G., Zheng, W., Ye, Z., & Zhang, Y. (2020). GREEN STRIPE, encoding methylated TOMATO AGAMOUS-LIKE 1, regulates chloroplast development and Chl synthesis in fruit. *New Phytologist*, 228(1), 302–317. <https://doi.org/10.1111/nph.16705>
- Liu, G., Xia, Y., Liu, T., Dai, S., & Hou, X. (2018). The DNA methylome and association of differentially methylated regions with differential gene expression during heat stress in *Brassica rapa*. *International Journal of Molecular Sciences*, 19(5). <https://doi.org/10.3390/ijms19051414>
- Liu, X., Yu, C. W., Duan, J., Luo, M., Wang, K., Tian, G., Cui, Y., & Wu, K. (2012). HDA6 Directly interacts with DNA methyltransferase MET1 and maintains transposable element silencing in Arabidopsis. *Plant Physiology*, 158(1), 119–129. <https://doi.org/10.1104/pp.111.184275>
- Liu, Z. W., Shao, C. R., Zhang, C. J., Zhou, J. X., Zhang, S. W., Li, L., Chen, S., Huang, H. W., Cai, T., & He, X. J. (2014). The SET Domain Proteins SUVH2 and SUVH9 Are Required for Pol V Occupancy at RNA-Directed DNA Methylation Loci. *PLoS Genetics*, 10(1). <https://doi.org/10.1371/journal.pgen.1003948>
- LoSchiavo, F., Pitto, L., Giuliano, G., Torti, G., Nuti-Ronchi, V., Marazziti, D., Vergara, R., Orselli, S., & Terzi, M. (1989). DNA methylation of embryogenic carrot cell cultures and its variations as caused by mutation, differentiation, hormones and hypomethylating drugs. *Theoretical and Applied Genetics*, 77(3), 325–331. <https://doi.org/10.1007/BF00305823>
- Lottaz, C., Iseli, C., Jongeneel, C. v., & Bucher, P. (2003). Modeling sequencing errors by combining Hidden Markov models. *Bioinformatics*, 19(SUPPL. 2). <https://doi.org/10.1093/bioinformatics/btg1067>
- Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550 (2014). <https://doi.org/10.1186/s13059-014-0550-8>
- Machczyńska, J., Orłowska, R., Mańkowski, D. R., Zimny, J., & Bednarek, P. T. (2014). DNA methylation changes in triticale due to *in vitro* culture plant regeneration and consecutive

---

reproduction. *Plant Cell, Tissue and Organ Culture*, 119(2), 289–299.  
<https://doi.org/10.1007/s11240-014-0533-1>

Maeji, H., & Nishimura, T. (2018). Epigenetic Mechanisms in Plants. In *Advances in Botanical Research* (Vol. 88). Elsevier Ltd. <https://doi.org/10.1016/bs.abr.2018.09.014>

Mahfouz, M. M. (2010). RNA-directed DNA methylation: Mechanisms and functions. *Plant Signaling and Behavior*, 5(7), 806–816. <https://doi.org/10.4161/psb.5.7.11695>

Marturano, A. (2009). Bioinformatics and Ethics Editorial: *Bioethics*, 23(7).  
<https://doi.org/10.1111/j.1467-8519.2009.01764.x>

Mgbeze, G. C., & Iserhienrhien, A. (2014). Somaclonal variation associated with oil palm (*Elaeis guineensis* Jacq.) clonal propagation: A review. *African Journal of Biotechnology*, 13(9), 989-997

Migneault, A., Sandhu, H., McCord, P., Zhao, D., & Erickson, J. (2019). Albinism in sugarcane: significance, research gaps, and potential future research developments. *Sugar Tech*, 21(3), 536–541. <https://doi.org/10.1007/s12355-018-0668-1>

Miguel, C., & Marum, L. (2011). An epigenetic view of plant cells cultured *in vitro*: Somaclonal variation and beyond. *Journal of Experimental Botany*, 62(11), 3713–3725.  
<https://doi.org/10.1093/jxb/err155>

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49(D1), D412–D419.  
<https://doi.org/10.1093/nar/gkaa913>

Moglia, A., Gianoglio, S., Acquadro, A., Valentino, D., Milani, A. M., Lanteri, S., & Comino, C. (2019). Identification of DNA methyltransferases and demethylases in *Solanum melongena* L., and their transcription dynamics during fruit development and after salt and drought stresses. *PLoS ONE*, 14(10), 1–18. <https://doi.org/10.1371/journal.pone.0223581>

Murashige, T., & Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum*, 15, 474–497.

- Neelakandan, A. K., & Wang, K. (2012). Recent progress in the understanding of tissue culture-induced genome level changes in plants and potential applications. *Plant Cell Reports*, 31(4), 597–620. <https://doi.org/10.1007/s00299-011-1202-z>
- Nie, W. F. (2021). DNA methylation: From model plants to vegetable crops. *Biochemical Society Transactions*, 49(3), 1479–1487. <https://doi.org/10.1042/BST20210353>
- Niederhuth, C. E., Bewick, A. J., Ji, L., Alabady, M. S., Kim, K. do, Li, Q., Rohr, N. A., Rambani, A., Burke, J. M., Udall, J. A., Egesi, C., Schmutz, J., Grimwood, J., Jackson, S. A., Springer, N. M., & Schmitz, R. J. (2016). Widespread natural variation of DNA methylation within angiosperms. *Genome Biology*, 17(1), 1–19. <https://doi.org/10.1186/s13059-016-1059-0>
- Orellana, R. (2015). Algunos aspectos ecológicos de los *Agaves* de la Península de Yucatán. October 1985.
- Palomino, G., Martínez, J., Romero, P., Barba-González, R., & Rodríguez-Garay, B. (2017). Nuclear genome size and karyotype analysis of *Agave angustifolia* Haw. “Cimarron” and “Lineño” (Asparagales, Asparagaceae). *Caryologia*, 70(2), 93–101. <https://doi.org/10.1080/00087114.2017.1279931>
- Park, S. Y., Murthy, H. N., Chakrabarthy, D., & Paek, K. Y. (2009). Detection of epigenetic variation in tissue-culture-derived plants of *Doritaenopsis* by methylation-sensitive amplification polymorphism (MSAP) analysis. *In Vitro Cellular and Developmental Biology - Plant*, 45(1), 104–108. <https://doi.org/10.1007/s11627-008-9166-6>
- Pavlopoulou, A., & Kossida, S. (2007). Plant cytosine-5 DNA methyltransferases: Structure, function, and molecular evolution. *Genomics*, 90(4), 530–541. <https://doi.org/10.1016/j.ygeno.2007.06.011>
- Pérez-Zavala, M. de L., Hernández-Arzaba, J. C., Bideshi, D. K., & Barboza-Corona, J. E. (2020). Agave: a natural renewable resource with multiple applications. In *Journal of the Science of Food and Agriculture* (Vol. 100, Issue 15, pp. 5324–5333). <https://doi.org/10.1002/jsfa.10586>
- Peschke, V. M., & Phillips, R. L. (1992). Genetic implications of somaclonal variation in plants. *Advances in Genetics*, 30(C), 41–75. [https://doi.org/10.1016/S0065-2660\(08\)60318-1](https://doi.org/10.1016/S0065-2660(08)60318-1)

- Pikaard, C. S., & Scheid, O. M. (2014). Epigenetic regulation in plants. *Cold Spring Harbor Perspectives in Biology*, 6(12), 1–31. <https://doi.org/10.1101/cshperspect.a019315>
- Prasad Parita, A. (2018). Insight into the DNA methyltransferases in plants and animals. *International Journal of Current Research in Life Sciences*. <https://doi.org/10.23959/sfdorj-1000023>
- Rajan, R. P., & Singh, G. (2021). A review on application of somaclonal variation in important horticulture crops. *plant cell biotechnology and molecular biology*, 22(35-36), 161-175. Retrieved from <https://www.ikppress.org/index.php/PCBMB/article/view/6319>
- Ramirez M. R.; L.A. Parra; F.I. Armenta; J.Borodanenko y J.L. Barrera. 2000. Germinación de semillas y cultivo in Vitro de brotes de *Agave Tequilana* WEBER, Var. Azul. Inc. Memorias del XVIII Congreso nacional de fitogenética. Irapuato, Gto. Mex. Pp 344.
- Rashda, N., Muhammad, A., Siddra, I., & Muhammad, S. (2018). Induction of somaclonal variation in selected drought sensitive group of sugarcane (*Saccharum officinarum*). *International Journal of Agriculture and Biology*, 20(4), 777–783.
- Rawoof, A., Ahmad, I., & Ramchiary, N. (2019). Epigenome Landscape in Capsicum Genome. 187–199. [https://doi.org/10.1007/978-3-319-97217-6\\_11](https://doi.org/10.1007/978-3-319-97217-6_11)
- Rica Sánchez-Chiang, C. (2009). Técnicas moleculares para la detección de variantes somaclonales. *Agronomía Mesoamericana*, 20(1), 135–151. <http://www.redalyc.org/articulo.oa?id=43711514015>
- Robert, M. L., Herrera-Herrera, J. L., Castillo, E., Ojeda, G., & Herrera-Alamillo, M. A. (2006). An efficient method for the micropropagation of *Agave* species. *Methods in Molecular Biology* (Clifton, N.J.), 318(1), 165–178. <https://doi.org/10.1385/1-59259-959-1:165>
- Rodríguez López, C. M., Wetten, A. C., & Wilkinson, M. J. (2010). Progressive erosion of genetic and epigenetic variation in callus-derived cocoa (*Theobroma cacao*) plants. *New Phytologist*, 186(4), 856–868. <https://doi.org/10.1111/j.1469-8137.2010.03242.x>
- Ryazanova, A. Yu., Abrasimova, L. A., Oretskaya, T. S., & Kubareva, E. A. (2016). Diverse Domains of (Cytosine-5)-DNA Methyltransferases: Structural and Functional Characterization. *IntechOpen, i(tourism)*, 13. <https://doi.org/http://dx.doi.org/10.5772/57353>

- Sahijram, L., Soneji, J. R., & Bollamma, K. T. (2003). Analyzing somaclonal variation in micropropagated bananas (*Musa* spp.). *In Vitro Cellular and Developmental Biology - Plant*, 39(6), 551–556. <https://doi.org/10.1079/IVP2003467>
- Saitou, N., & Nei, M. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol*, 4(4), 406–425.
- Sarmah, D. (2017). Somaclonal variation and its' application in ornamentals plants. *International Journal of Pure & Applied Bioscience*, 5(2), 396–406. <https://doi.org/10.18782/2320-7051.2762>
- Schellenbaum, P., Mohler, V., Wenzel, G., & Walter, B. (2008). Variation in DNA methylation patterns of grapevine somaclones (*Vitis vinifera* L.). *BMC Plant Biology*, 8, 1–10. <https://doi.org/10.1186/1471-2229-8-78>
- Smykal, P., Valledor, L., Rodríguez, R., & Griga, M. (2007). Assessment of genetic and epigenetic stability in long-term in vitro shoot culture of pea (*Pisum sativum* L.). *Plant Cell Reports*, 26(11), 1985–1998. <https://doi.org/10.1007/s00299-007-0413-9>
- Song, Y., Tang, H., Ke, X., Guo, J., Zhang, S., Wang, J., Niu, N., Ma, S., Zhao, H., & Zhang, G. (2018). Chloroplast structure and DNA methylation polymorphisms in an albino mutant of wheat (*Triticum aestivum*) cv. Xinong 1376. *Crop and Pasture Science*, 69(4), 362–373. <https://doi.org/10.1071/CP17471>
- Sow, M. D., le Gac, A. L., Fichot, R., Lanciano, S., Delaunay, A., le Jan, I., Lesage-Descauses, M. C., Citerne, S., Caius, J., Brunaud, V., Soubigou-Taconnat, L., Cochard, H., Segura, V., Chaparro, C., Grunau, C., Daviaud, C., Tost, J., Brignolas, F., Strauss, S. H., ... Maury, S. (2021). RNAi suppression of DNA methylation affects the drought stress response and genome integrity in transgenic poplar. In *New Phytologist* (Vol. 232, Issue 1). <https://doi.org/10.1111/nph.17555>
- Stolzenburg, S., Goubert, D., & Rots, M. G. (2016). DNA Methyltransferases - Role and Function. *Springer International*, 945, 475–490. <https://doi.org/10.1007/978-3-319-43624-1>

- Strehle, M., & Guttman, M. (2020). Xist drives spatial compartmentalization of DNA and protein to orchestrate initiation and maintenance of X inactivation. *Current Opinion in Cell Biology*, 64(Xci), 139–147. <https://doi.org/10.1016/j.ceb.2020.04.009>
- Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D. J., & Jacobsen, S. E. (2014). Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nature Structural and Molecular Biology*, 21(1), 64–72. <https://doi.org/10.1038/nsmb.2735>
- Tabares Z., Eddie; Pachón G., Jorge; Roca, William M. 1991. Variación somaclonal y su aplicación al mejoramiento de cultivos. In: Roca, William M.; Mroginski, Luis A. (eds.). *Cultivo de tejidos en la agricultura: Fundamentos y aplicaciones*. Centro Internacional de Agricultura Tropical (CIAT), Cali, CO. p. 339-359.
- Tariq, M., & Paszkowski, J. (2004). DNA and histone methylation in plants. *Trends in Genetics*, 20(6), 244–251. <https://doi.org/10.1016/j.tig.2004.04.005>
- Us-Camas, R., Castillo-Castro, E., Aguilar-Espinosa, M., Limones-Briones, V., Rivera-Madrid, R., Robert-Díaz, M. L., & De-la-Peña, C. (2017). Assessment of molecular and epigenetic changes in the albinism of *Agave angustifolia* Haw. *Plant Science*, 263(June), 156–167. <https://doi.org/10.1016/j.plantsci.2017.07.010>
- Victoria, D., Alik, K., Venetia, K., Georgios, M., & Zoe, H. (2018). Spatial and temporal expression of cytosine-5 DNA methyltransferase and DNA demethylase gene families of the *Ricinus communis* during seed development and drought stress. *Plant Growth Regulation*, 84(1), 81–94. <https://doi.org/10.1007/s10725-017-0323-y>
- Vining, K., Pomraning, K. R., Wilhelm, L. J., Ma, C., Pellegrini, M., Di, Y., Mockler, T. C., Freitag, M., & Strauss, S. H. (2013). Methylome reorganization during *in vitro* dedifferentiation and regeneration of *Populus trichocarpa*. *BMC Plant Biology*, 13(1), 1. <https://doi.org/10.1186/1471-2229-13-92>
- Vranych, C. v., Rivero, M. R., Merino, M. C., Mayol, G. F., Zamponi, N., Maletto, B. A., Pistoresi-Palencia, M. C., Touz, M. C., & Rópolo, A. S. (2014). SUMOylation and deimination of proteins: Two epigenetic modifications involved in *Giardia encystation*. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843(9), 1805–1817. <https://doi.org/10.1016/j.bbamcr.2014.04.014>

- Waddington, C. H. (2012). The epigenotype. 1942. *International Journal of Epidemiology*, 41(1), 10–13. <https://doi.org/10.1093/ije/dyr184>
- Wang, Q. M., Wang, L., Zhou, Y., Cui, J., Wang, Y., & Zhao, C. (2016). Leaf patterning of *Clivia miniata* var. *variegata* is associated with differential DNA methylation. *Plant Cell Reports*, 35(1), 167–184. <https://doi.org/10.1007/s00299-015-1877-7>
- Wang, Q., Zhang, S., Chen, C., Wang, H., Guo, Z., & Li, D. (2015). An improved Tilt Derivative Method Combined with the MMTs E Using in Fracture System Identification for the SWIR 49° E Area. *ASEG Extended Abstracts*, 2015(1), 1–4. <https://doi.org/10.1071/aseg2015ab284>
- Wei, X., Song, X., Wei, L., Tang, S., Sun, J., Hu, P., & Cao, X. (2017). An epiallele of rice *AK1* affects photosynthetic capacity. *Journal of Integrative Plant Biology*, 59(3), 158–163. <https://doi.org/10.1111/jipb.12518>
- Wheeler, T. J., & Eddy, S. R. (2013). Nhmmer: DNA homology search with profile HMMs. *Bioinformatics*, 29(19), 2487–2489. <https://doi.org/10.1093/bioinformatics/btt403>
- Yaari, R., Katz, A., Domb, K., Harris, K. D., Zemach, A., & Ohad, N. (2019). RdDM-independent de novo and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs. *Nature Communications*, 10(1), 1–10. <https://doi.org/10.1038/s41467-019-09496-0>
- Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: Principles and practice. In *Nature Reviews Genetics*. <https://doi.org/10.1038/nrg3186>
- Young, A. D., & Gillung, J. P. (2020). Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, 45(2), 225–247. <https://doi.org/10.1111/syen.12406>
- Yu, X., Li, X., Zhao, X., Jiang, L., Miao, G., Pang, J., Qi, X., & Liu, B. (2011). Tissue culture-induced genomic alteration in maize (*Zea mays*) inbred lines and F1 hybrids. *Annals of Applied Biology*, 158(3), 237–247. <https://doi.org/10.1111/j.1744-7348.2011.00458.x>
- Yu, Z., Haberer, G., Matthes, M., Rattei, T., Mayer, K. F. X., Gierl, A., & Torres-ruiz, R. A. (2010). Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. 107(41), 17809–17814. <https://doi.org/10.1073/pnas.1000852107>

- Zemach, A., Kim, M. Y., Hsieh, P. H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S. L., & Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*, 153(1), 193–205. <https://doi.org/10.1016/j.cell.2013.02.033>
- Zhang Yingying, Christian Rohde, Tierling Sascha, Stamerjohanns Heinrich, Reinhardt Richard, Walter Jorn, J. A. (2009). DNA methylation analysis by bisulfite conversion, cloning, and sequencing of individual clones. *DNA Methylation*, 1, 177–187. <https://doi.org/10.1007/978-1-59745-522-0>
- Zhang, H., Lang, Z., & Zhu, J. K. (2018). Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology*, 19(8), 489–506. <https://doi.org/10.1038/s41580-018-0016-z>
- Zhang, Y., He, X., Zhao, H., Xu, W., Deng, H., Wang, H., Wang, S., Su, D., Zheng, Z., Yang, B., Grierson, D., Wu, J., & Liu, M. (2020). Genome-wide identification of DNA methylases and demethylases in Kiwifruit (*Actinidia chinensis*). *Frontiers in Plant Science*, 11(September), 1–11. <https://doi.org/10.3389/fpls.2020.514993>
- Zhao, J., Morozova, N., Williams, L., Libs, L., Avivi, Y., & Grafi, G. (2001). Two phases of chromatin decondensation during dedifferentiation of plant cells. Distinction between competence for cell fate switch and a commitment for S phase. *Journal of Biological Chemistry*, 276(25), 22772–22778. <https://doi.org/10.1074/jbc.M101756200>
- Zhao, P., Wang, K., Zhang, W., Liu, H. Y., Du, L. P., Hu, H. R., & Ye, X. G. (2017). Comprehensive analysis of differently expressed genes and proteins in albino and green plantlets from a wheat anther culture. *Biologia Plantarum*, 61(2), 255–265. <https://doi.org/10.1007/s10535-016-0662-y>
- Zhao, Y., Wang, J., Chen, J., Zhang, X., Guo, M., & Yu, G. (2020). A literature review of gene Function prediction by modeling gene ontology. In *Frontiers in Genetics*. <https://doi.org/10.3389/fgene.2020.00400>
- Zhong, S., Fei, Z., Chen, Y. R., Zheng, Y., Huang, M., Vrebalov, J., McQuinn, R., Gapper, N., Liu, B., Xiang, J., Shao, Y., & Giovannoni, J. J. (2013). Single-base resolution methylomes of tomato

fruit development reveal epigenome modifications associated with ripening. *Nature Biotechnology*, 31(2), 154–159. <https://doi.org/10.1038/nbt.2462>

Zhong, X., Du, J., Hale, C. J., Gallego-Bartolome, J., Feng, S., Vashisht, A. A., Chory, J., Wohlschlegel, J. A., Patel, D. J., & Jacobsen, S. E. (2014). Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell*, 157(5), 1050–1060. <https://doi.org/10.1016/j.cell.2014.03.056>

Zhou, W. Z., Zhang, Y. M., Lu, J. Y., & Li, J. F. (2012). Construction and evaluation of normalized cDNA libraries enriched with full-length sequences for rapid discovery of new genes from sisal (*Agave sisalana* perr.) different developmental stages. *International Journal of Molecular Sciences*, 13(10), 13150–13168. <https://doi.org/10.3390/ijms131013150>

Zhu, C., Zhang, S., Zhou, C., Chen, L., Fu, H., Li, X., Lin, Y., Lai, Z., & Guo, Y. (2020). Genome-wide investigation and transcriptional analysis of cytosine-5 DNA methyltransferase and DNA demethylase gene families in tea plant (*Camellia sinensis*) under abiotic stress and withering processing. *PeerJ*, 2020(1). <https://doi.org/10.7717/peerj.8432>

Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., & Henikoff, S. (2007). Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, 39(1), 61–69. <https://doi.org/10.1038/ng1929>

## Anexo 1. Dominios identificados en proteínas de Metiltransferasas de ADN

Tabla 2. 7 Dominios identificados en proteínas de MT-asas de ADN de *A. angustifolia* Haw. Se describe la longitud en aminoácidos de las secuencias identificadas en el transcriptoma, además se muestran los dominios identificados en las bases de datos Pfam y dominios conservados de NCBI (CDD). Ubicación: contiene las coordenadas en aminoácidos donde se encuentra el dominio identificad. Por último, se incluye el número de accesión de los dominios identificados en cada una de las bases de datos.

ID del transcriptoma	Nombre del gen	Longitud de la proteína (a.a)	Dominios en PFAM	Ubicación(a.a)	Número de accesión	Dominios en CDD	Ubicación (a.a.)	Número de accesión
Cluster-46512.96201	AaH DRM2a_201	617	MtADN	299-613	CL0063	MtADN	301-611	cl30459
Cluster-46512.96201	AaH DRM2a_201	617	UBA	162-188	CL0214	UBA	164-188	cl21463
Cluster-46512.110739	AaH DRM2b_739	617	MtADN	300-612	CL0063	MtADN	301-611	cl30459
Cluster-46512.110739	AaH DRM2b_739	617	UBA	163-189	CL0214	UBA	164-188	cl21463
Cluster-46512.142275	AaH DRM2c_275	621	MtADN	303-617	CL0063	MtADN	305-615	cl30459
Cluster-46512.142275	AaH DRM2c_275	621	UBA	164-462	CL0214	UBA	166-190	cl21463
Cluster-46512.161086	AaH DRM3a_086	576	MtADN	454-571	CL0063	MtADN	451-574	cl30459
Cluster-46512.161086	AaH DRM3a_086	576	UBA	100-130	CL0214	UBA	96-132	cl21463
Cluster-46512.160638	AaH CMT2a_638	928	MtADN	356-891	CL0063	MtADN	356-890	cl30459
Cluster-46512.160638	AaH CMT2a_638	928	CHROMO	465-520	CL0049	CHROMO	464-520	cd18635
Cluster-46512.160638	AaH CMT2a_638	928	BAH	182-308	-	BAH	180-311	cl02608
Cluster-46512.159175	AaH CMT2_175	926	MtADN	355-887	CL0063	MtADN	355-888	cl30459
Cluster-46512.159175	AaH CMT2b_175	926	CHROMO	463-518	CL0049	CHROMO	462-518	cd18635
Cluster-46512.159175	AaH CMT2b_175	926	BAH	180-308	-	BAH	179-310	cl02608
Cluster-46512.167523	AaH CMT2c_523	698	MtADN	356-693	CL0063	MtADN	356-634	cl30459
Cluster-46512.167523	AaH CMT2c_523	698	BAH	182-309	-	BAH	180-311	cl02608
Cluster-46512.167523	AaH CMT2c_523	698	CHROMO	465-520	CL0049	CHROMO	464-520	cd18635
Cluster-46512.85248	AaH CMT3a_248	880	MtADN	315-844	CL0063	MtADN	313-843	cl30459
Cluster-46512.85248	AaH CMT3a_248	880	BAH	165-282	cl02608	BAH	164-283	cl02608
Cluster-46512.85248	AaH CMT3a_248	880	CHROMO	425-481	CL0049	-	-	-
Cluster-46512.141003	AaH CMT3b_003	926	MtADN	359-903	CL0063	MtADN	361-901	cl30459
Cluster-46512.141003	AaH CMT3b_003	926	BAH	196-315	cl02608	BAH	195-315	cl02608
Cluster-46512.141003	AaH CMT3b_003	926	CHROMO	476-532	CL0049	CHROMO	476-532	cd18635
Cluster-46512.117080	AaH MET1B_080	1548	MtADN	1110-1539	CL0063	MtADN	1108-1540	cl30459
Cluster-46512.117080	AaH MET1B_080	1548	BAH	753-887	cl02608	BAH	750-888	cl02608
Cluster-46512.117080	AaH MET1B_080	1548	BAH	924-1065	cl02608	BAH	924-1065	cl02608
Cluster-46512.117080	AaH MET1B_080	1548	RFD	104-240	-	RFD	107-220	pfam12047
Cluster-46512.117080	AaH MET1B_080	1548	RFD	399-554	-	RFD	400-553	pfam12047