



**UNIVERSIDAD AUTÓNOMA
DE AGUASCALIENTES**

CENTRO DE CIENCIAS BÁSICAS

DOCTORADO EN CIENCIAS BIOLÓGICAS

TESIS:

**Análisis del Transcriptoma en *Mammillaria bombycina* (Quehl) y estudio de los genes Gly
(Glioxalasa).**

PRESENTA:

Carolina Enríquez González

PARA OBTENER EL GRADO DE

DOCTOR EN CIENCIAS EN EL ÁREA DE BIOTECNOLOGÍA VEGETAL

TUTORES

Dr. José Francisco Morales Domínguez

COMITÉ TUTORAL

Dr. Gilberto Alejandro Ocampo Acosta

Dr. Eugenio Martín Pérez Molphe Balch

Dr. Osvaldo Adrián Castellanos Hernández

Aguascalientes, Ags., mayo de 2022

Autorizaciones

M. en C. Jorge Martín Alférez Chávez
PRESENTE

Por medio del presente como **Miembros del Comité Tutorial** designado del estudiante **CAROLINA ENRIQUEZ GÓNZÁLEZ** con ID **687** quien realizó la tesis titulada: **ANÁLISIS DEL TRANSCRIPTOMA EN MAMILLARIA BOMBYCINA Y ESTUDIO DE LOS GENES GLY (GLIOXALASAS)**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia damos nuestro consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que nos permitimos emitir el **VOTO APROBATORIO**, para que ella pueda proceder a imprimirla así como continuar con el procedimiento administrativo para la obtención del grado.

Ponemos lo anterior a su digna consideración y sin otro particular por el momento, le enviamos un cordial saludo,

ATENTAMENTE
"Se Lumen Proferre"

Aguascalientes, Ags., a 31 de mayo de 2022


Dr. Jose Francisco Morales Dominguez
Tutor de tesis


Dr. Gilberto Alejandro Dcampo Acosta
Asesor de tesis


Dr. Eugenio Perez Molphe Bolch
Asesor de tesis


Dr. Osvaldo Adrian Castellanos Hernandez
Asesor de tesis

c.c.p. Interesado
c.c.p. Secretaría Técnica del Proctoma de Posgrado

Fecha de dictaminación dd/mm/aa: 010/06/2022

NOMBRE: CAROLINA ENRIQUEZ GONZÁLEZ ID 687

PROGRAMA: DOCTORADO EN CIENCIAS BIOLÓGICAS LGAC (del posgrado): BIOTECNOLOGÍA Y BIOQUÍMICA VEGETAL Y ANIMAL

TIPO DE TRABAJO: (X) Tesis () Trabajo práctico

TÍTULO: Análisis del Transcriptoma en Mammillaria bombycina (Queh) y estudio de los genes Gly (Glioxalasa).

IMPACTO SOCIAL (señalar el impacto logrado):

CONTRIBUYE AL CONOCIMIENTO Y PRESERVACIÓN D EMAMILLARIA BOMBYCINA

INDICAR SI/NO SEGÚN CORRESPONDA:

Elementos para la revisión académica del trabajo de tesis o trabajo práctico:

- SI El trabajo es congruente con las LGAC del programa de posgrado
- SI La problemática fue abordada desde un enfoque multidisciplinario
- SI Existe coherencia, continuidad y orden lógico del tema central con cada apartado
- SI Los resultados del trabajo dan respuesta a las preguntas de investigación o a la problemática que aborda
- SI Los resultados presentados en el trabajo son de gran relevancia científica, tecnológica o profesional según el área
- SI El trabajo demuestra más de una aportación original al conocimiento de su área
- SI Las aportaciones responden a los problemas prioritarios del país
- NO Generó transferencia del conocimiento o tecnológica

El egresado cumple con lo siguiente:

- SI Cumple con lo señalado por el Reglamento General de Docencia
- SI Cumple con los requisitos señalados en el plan de estudios (créditos curriculares, optativos, actividades complementarias, estancia, predoctoral, etc)
- SI Cuenta con los votos aprobatorios del comité tutorial, en caso de los posgrados profesionales si tiene solo tutor podrá liberar solo el tutor
- NO Cuenta con la carta de satisfacción del Usuario
- SI Coincide con el título y objetivo registrado
- SI Tiene congruencia con cuerpos académicos
- SI Tiene el CVU del Conacyt actualizado
- SI Tiene el artículo aceptado o publicado y cumple con los requisitos institucionales (en caso que proceda)

En caso de Tesis por artículos científicos publicados:

- SI Aceptación o Publicación de los artículos según el nivel del programa
- SI El estudiante es el primer autor
- SI El autor de correspondencia es el Tutor del Núcleo Académico Básico
- SI En los artículos se ven reflejados los objetivos de la tesis, ya que son producto de este trabajo de investigación.
- SI Los artículos integran los capítulos de la tesis y se presentan en el idioma en que fueron publicados
- SI La aceptación o publicación de los artículos en revistas indexadas de alto impacto

Con base a estos criterios, se autoriza se continúen con los trámites de titulación y programación del examen de grado

SI
No

FIRMAS

Elaboró:

* NOMBRE Y FIRMA DEL CONSEJERO SEGÚN LA LGAC DE ADSCRIPCIÓN:

DRA. ILIANA ERNESTINA MEDINA RAMÍREZ

NOMBRE Y FIRMA DEL SECRETARIO TÉCNICO:

DRA. ALMA LILIAN GUERRERO BARRERA

* En caso de conflicto de intereses, firmará un revisor miembro del NAB de la LGAC correspondiente distinto al tutor o miembro del comité tutorial, asignado por el Decano

Revisó:

NOMBRE Y FIRMA DEL SECRETARIO DE INVESTIGACIÓN Y POSGRADO:

DRA. HAYDÉE MARTÍNEZ RUVALCABA

Autorizó:

NOMBRE Y FIRMA DEL DECANO:





M. en C. JORGE MARTÍN ALÍEZ CHÁVEZ

Nota: procede el trámite para el Depto. de Apoyo al Posgrado

En cumplimiento con el Art. 105C del Reglamento General de Docencia que a la letra señala entre las funciones del Consejo Académico: ... Cuidar la eficiencia terminal del programa de posgrado y el Art. 105F las funciones del Secretario Técnico, llevar el seguimiento de los alumnos.

Article

De Novo Transcriptome of *Mammillaria bombycina* (Cactaceae) under In Vitro Conditions and Identification of Glyoxalase Genes

Carolina Enriquez-González ¹ , Cristina Garcidueñas-Piña ¹, Osvaldo Adrián Castellanos-Hernández ² , Sergio Enriquez-Aranda ³, Abraham Loera-Muro ⁴ , Gilberto Ocampo ⁵, Eugenio Pérez-Molphe Balch ¹ and José Francisco Morales-Domínguez ^{1,*} 

- ¹ Centro de Ciencias Básicas, Departamento de Química, Universidad Autónoma de Aguascalientes, Av. Universidad 940, Aguascalientes 20100, Mexico; al687@edu.uaa.mx (C.E.-G.); cristina.garcidueñas@edu.uaa.mx (C.G.-P.); eperezmb@gmail.com (E.P.-M.B.)
- ² Centro de Investigación en Biología Molecular Vegetal, Centro Universitario de la Ciénega, Universidad de Guadalajara, Av. Universidad, 1115, Linda Vista, Ocotlán 47810, Mexico; ocnoscr@gmail.com
- ³ Centro de Ciencias Básicas, Departamento de Sistemas de la Información, Universidad Autónoma de Aguascalientes, Av. Universidad 940, Aguascalientes 20100, Mexico; sergio.enriquez@edu.uaa.mx
- ⁴ CONACyT Centro de Investigaciones Biológicas del Noroeste S.C., Av. Instituto Politécnico Nacional 195, Playa Palo de Santa Rita Sur, La Paz 23096, Mexico; aloera@cibnor.mx
- ⁵ Centro de Ciencias Básicas, Departamento de Biología, Universidad Autónoma de Aguascalientes, Av. Universidad 940, Aguascalientes 20100, Mexico; gilberto.ocampo@edu.uaa.mx
- * Correspondence: francisco.morales@edu.uaa.mx



Citation: Enriquez-González, C.; Garcidueñas-Piña, C.; Castellanos-Hernández, O.A.; Enriquez-Aranda, S.; Loera-Muro, A.; Ocampo, G.; Pérez-Molphe Balch, E.; Morales-Domínguez, J.F. De Novo Transcriptome of *Mammillaria bombycina* (Cactaceae) under In Vitro Conditions and Identification of Glyoxalase Genes. *Plants* **2022**, *11*, 399. <https://doi.org/10.3390/plants11030399>

Academic Editor: Thomas Nussbaumer

Received: 12 November 2021
Accepted: 18 January 2022
Published: 31 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: *Mammillaria bombycina* is a cactus distributed in the central region of Mexico. Cactaceae have the particularity of surviving drought and high temperatures, which is why in vitro propagation studies have been carried out successfully to preserve this species and use it as a study model in cacti. In this contribution, a de novo transcriptome of *M. bombycina* was produced under in vitro conditions for the identification and expression of genes related to abiotic stress. Samples were sequenced using an Illumina platform, averaging 24 million clean readings. From assembly and annotation, 84,975 transcripts were generated, 55% of which were unigenes. Among these, the presence of 13 isoforms of genes belonging to glyoxalase I, II and III were identified. An analysis of the qRT-PCR expression of these genes was performed under in vitro and ex vitro conditions and dehydration at 6 and 24 h. The highest expression was observed under greenhouse conditions and dehydration at 24 h, according to the control. The de novo assembly of the *M. bombycina* transcriptome remains a study model for future work in cacti.

Keywords: Cactaceae; Glyoxalases; D-lactate; methylglyoxal; RNA-Seq

1. Introduction

Mammillaria bombycina is located in the mountainous area between the states of Jalisco and Aguascalientes, Mexico in the subtropical scrub, and in the low deciduous forest of this region [1]. It is generally small, globose, with abundant thorns and very colorful flowers (Figure 1), which makes it highly attractive [2]. Overexploitation, overgrazing, and climate change have considerably reduced its spread and natural distribution [3], which is why *M. bombycina* has been classified as endangered by Mexican authorities in NOM-059. -SEMARNAT-2010 [4]. As measures to preserve this species, a large number of in vitro propagation protocols have been developed (Figure 1a) as well as a germplasm bank [5]. *M. bombycina* seedlings grown in vitro can adapt to an ex vitro medium, and their cultures can be maintained under greenhouse conditions (Figure 1b), enabling studies of the identification and expression of genes related to abiotic stress to be developed [6] and research related to its metabolism to continue.

Agradecimientos

Al finalizar este trabajo tan arduo y lleno de dificultades es para mí un verdadero placer utilizar este espacio para expresar mis más sinceros agradecimientos a los miembros de mi comité tutorial: Al Dr. Francisco Morales Domínguez por guiarme, compartir su conocimiento y darme su apoyo incondicional no sólo para realizar este proyecto, si no a lo largo de todos los años que tenemos de conocernos, por sus consejos, amistad, y comprensión en los momentos difíciles. Al Dr. Gilberto Ocampo por todo su apoyo y comprensión, por compartir su conocimiento y experiencia, por depositar en mí su confianza y financiar parte de este proyecto, ya que sin su apoyo este trabajo no se hubiera podido lograr. Al Dr. Eugenio Pérez Molphe-Balch por su comprensión, paciencia, por compartir conmigo su experiencia y conocimiento, por su dedicación, compromiso a lo largo de todos los años que tenemos de conocernos. Al Dr. Osvaldo Adrián Castellanos Hernández por compartir su experiencia y conocimiento, su paciencia y apoyo incondicional.

Al Dr. Sergio Enríquez por formar parte del proyecto, compartir su conocimiento y guiarme en toda la parte Bioinformática.

A la Dra. Cristina Garcidueñas Piña y el Dr. Abraham Loera Muro por compartir su conocimiento y experiencia al realizar algunos experimentos en colaboración al proyecto.

A mi compañera y amiga Mariana Muñoz Carrillo, por acompañarme y darme soporte a lo largo de este camino.

A mis compañeros del Laboratorio de Biología Molecular de plantas, Lili, Dani y Mario Arzate, por su amistad y por siempre apoyarme y darme palabras de aliento cuando más lo necesité.

Muchas gracias a todos ustedes por transmitirme, inculcarme y hacer más grande el amor que tengo por la ciencia.

También quisiera agradecer a la Universidad Autónoma de Aguascalientes y al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo para la realización de mis estudios de Doctorado (Beca N° 766920).

Dedicatorias

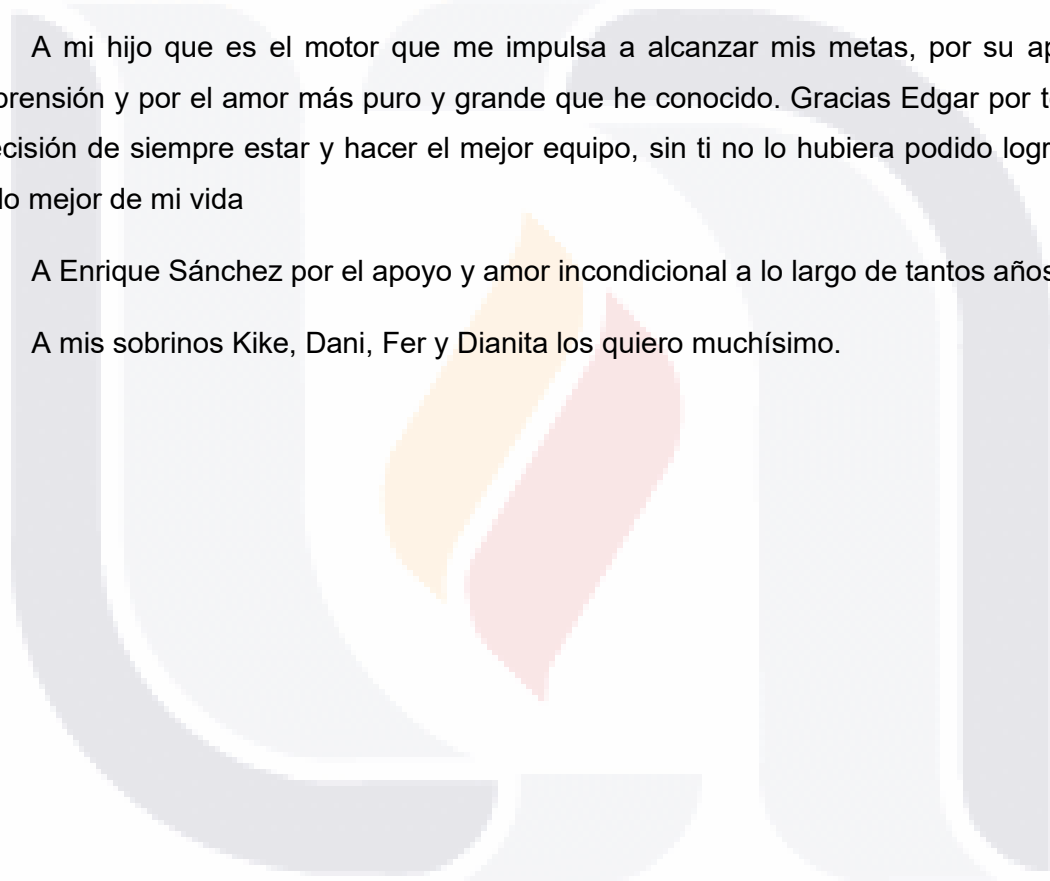
A mis padres Sergio y Esther por nunca soltarme a pesar de las dificultades, por guiarme, por sus invaluable consejos, por el amor tan inmenso que siempre me han demostrado, por su paciencia y apoyo incondicional. Gracias por hacer de mí la persona que soy ahora.

A mis hermanas Sisi, Mariana y Susy que son mis compañeras de vida, mis mejores amigas, por siempre estar conmigo, apoyarme y aconsejarme en todo momento, por sus palabras de aliento, y cariño incondicional.

A mi hijo que es el motor que me impulsa a alcanzar mis metas, por su apoyo, comprensión y por el amor más puro y grande que he conocido. Gracias Edgar por tomar la decisión de siempre estar y hacer el mejor equipo, sin ti no lo hubiera podido lograr.... Son lo mejor de mi vida

A Enrique Sánchez por el apoyo y amor incondicional a lo largo de tantos años.

A mis sobrinos Kike, Dani, Fer y Dianita los quiero muchísimo.



Índice General

1. Introducción 12

2. Antecedentes..... 14

 2.1. Cactáceas 14

 2.1.1. *Mammillaria bombycina* como modelo de estudio..... 16

 2.2. Transcriptoma..... 17

 2.3. Relevancia de la expresión génica plantas 19

 2.4. RNA-seq..... 22

 2.5. Descripción general de las plataformas de secuenciación para RNA-seq..... 24

 2.6. Pirosecuenciación..... 25

 2.7. Secuenciación en tiempo real de molécula única o SMET-seq 26

 2.8. Secuenciación por síntesis 26

 2.9. Análisis Bioinformático de las secuencias obtenidas de la plataforma Illumina .. 27

 2.10. Ensamble *de novo* con el programa Trinity 30

 2.11. Anotación funcional del Transcriptoma *de novo* de *M. bombycina* con el programa Trinotate..... 32

 2.12. Genes involucrados en algunas rutas reguladoras y funcionales en la especie de *M. bombycina* bajo condiciones *in vitro* 33

 2.12.1. Glioxalosas 34

 2.12.1.1. Importancia y biosíntesis de metilglioxal 39

 2.12.2. Factores de Transcripción 41

 2.12.2.1. Factor de Transcripción DREB 41

 2.12.2.2. Factor de Transcripción CDPK 42

 2.12.3. Subfamilias de α -expansinas y β -expansinas 42

3. Justificación 45

4. Hipótesis..... 46

5. Objetivos..... 47

5.1.	Objetivos generales.....	47
5.2.	Objetivos específicos.....	47
6.	Materiales y métodos.....	48
6.1.	Obtención del Material Vegetal.....	48
6.2.	Extracción de RNA.....	48
6.3.	RNA-seq.....	49
6.4.	Análisis FastQC, Trimming y filtrado de las secuencias obtenidas.....	49
6.5.	Ensamble del Transcriptoma <i>de novo</i> con el programa Trinity.....	50
6.6.	Anotación funcional del Transcriptoma <i>de novo</i> con Trinotate.....	50
6.7.	Genes relacionados con el estrés abiótico.....	50
6.7.1.	Análisis bioinformático en Glioxalasa I, II y DJ-1.....	50
6.7.2.	Análisis bioinformático en factores de transcripción DREB y CDPK.....	51
6.7.3.	Análisis Bioinformático de genes de α -expansinas y β -expansinas en <i>M. bombycina</i>	52
6.8.	Validación de la expresión de los genes relacionados con estrés osmótico de <i>M. bombycina in vitro</i> , basal y bajo estrés.....	52
6.9.	Análisis estadístico de la expresión de los genes relacionados con estrés abiótico de <i>M. bombycina in vitro</i> , basal y bajo estrés.....	53
6.10.	Esquema general de la metodología de la etapa 2: Validación de algunos genes expresados como respuesta ante estrés osmótico.....	54
7.	Resultados.....	55
7.1.	Propagación de <i>M. bombycina</i>	55
7.2.	Extracción de RNA de <i>M. bombycina</i>	55
7.3.	Información de la secuenciación.....	56
7.4.	Análisis FastQC, Trimming y filtrado de las secuencias obtenidas.....	57
7.5.	Ensamble de novo de <i>M. bombycina</i> con el programa Trinity.....	59
7.6.	Anotación funcional del ensamble del transcriptoma con el programa Trinotate.....	60
7.7.	Análisis de los genes Glioxalasa.....	65

7.8.	Caracterización y análisis bioinformático de proteínas putativas que corresponden a Glioxalastas de tipo I, II y III (GlyI, GlyII y GlyIII) en <i>M. bombycina</i>	68
7.8.1.	Análisis de la arquitectura de los dominios de las proteínas putativas de MbGLYI, MbGLYII y MbGLYIII.....	68
7.8.2	Dendrograma de las proteínas putativas MbGLYI, MbGLYII y MbDJ1.....	74
7.8.3	Caracterización fisicoquímica de las proteínas putativas de Gioxalasa de tipo I, II y III en la especie <i>M. bombycina</i>	76
7.8.4	Modelado hipotéticas de las proteínas putativas de Gioxalastas	76
7.8.5	Análisis de expresión de los genes relacionados con estrés abiótico	77
7.9	Caracterización y Análisis Bioinformático de proteínas putativas que corresponden a Factores de Transcripción DREB en <i>M. bombycina</i>	78
7.9.1	Análisis de la arquitectura de los dominios de las proteínas putativas de MbDREB1 y MbCDPK1	78
7.9.2	Caracterización fisicoquímica de las proteínas putativas de MbDREB1 y MbDREB2.....	80
7.9.3	Modelado 3D hipotético de las proteínas putativas de factores de transcripción DREB y CDPK	81
7.9.4	Análisis de la expresión de los genes de factores de transcripción DREB y CDPK bajo estrés por frío y ABA en <i>M. bombycina in vitro</i>	81
7.10	Caracterización y análisis bioinformático de proteínas putativas que corresponden a las subfamilias α -expansinas y β -expansinas en <i>M. bombycina</i>	82
7.10.1	Alineamiento múltiple y análisis de la arquitectura de los dominios y motivos de las proteínas putativas de α -expansinas (MbEXPA) y β -expansinas (EXPB).	82
7.10.2	Caracterización fisicoquímica de las proteínas putativas de las subfamilias α -expansinas y β -expansinas.....	90
7.10.3	Modelado 3D hipotético de las proteínas putativas de las subfamilias α -expansinas y β -expansinas.....	91
8	Discusión.....	93
9	Conclusiones.....	102
10	Bibliografía.....	103

11 Anexo 116

 Medio de cultivo Murashige y Skoog (MS): 116

 A1 Estadística básica en transcriptoma..... 117

ANEXO C. Análisis ContingX 118

12 Artículo 121

 De Novo Transcriptome of *Mammillaria bombycina* (Cactaceae) under In Vitro
 Conditions and Identification of Glyoxalase Genes 121



Índice de Tablas

Tabla 1. Inducción con Congelamiento..... 54

Tabla 2. Inducción con ABA 54

Tabla 3. Información obtenida de la secuenciación realizada en el equipo Next Seq 500, en el que se utilizó una configuración de 2X75 ciclos. 56

Tabla 4. Tabla comparativa de los resultados obtenidos del recorte de secuencias con los programas FastX Toolkit, FastP y CLC Genomics. 58

Tabla 5. Tabla comparativa de los porcentajes retenidos de los resultados obtenidos del recorte de secuencias con los programas FastX Toolkit, FastP y CLC Genomics. 58

Tabla 6. Resumen estadístico de los resultados obtenidos del ensamble con el programa Trinity..... 59

Tabla 7. Resumen estadístico de las anotaciones del transcriptoma de *M. bombycina* con el programa Trinotate..... 61

Tabla 8. Tabla de información complementaria de los transcritos obtenidos de los genes de la familia de las Glioxalastas..... 66

Tabla 9. Clasificación de las proteínas putativas de MbGLYI por el tamaño de sus dominios y presencia de motivos. 70

Tabla 10. Clasificación de las proteínas putativas de MbGLYII por el tamaño y características de sus dominios y motivos. 72

Tabla 11. Clasificación de las proteínas putativas de MbDJ-1 por el tamaño y características de sus dominios y motivos..... 73

Tabla 12. Caracterización fisicoquímica y localización celular de las proteínas MbDREB1 y MbDREB2..... 80

Tabla 13. Arquitectura de dominios de las α -expansinas y β -expansinas de *M. bombycina*. 85

Tabla 14. Información detallada de motivos encontrados. 89

Tabla 15. Análisis fisicoquímico y localización celular de los transcritos de las subfamilias de α -expansinas y β -expansinas en *M. bombycina*..... 91

Tabla 16. Preparación de las soluciones concentradas para el medio MS. 116

Índice de Figuras

Figura 1. Respuesta de orgánulos celulares ante el estrés. Las señales generadas por el estrés de todos los orgánulos en conjunto para mantener la homeostasis celular, genera una adaptación de la planta ante varios tipos de estrés (Zhu, 2016)..... 19

Figura 2. Respuesta génica bajo estrés en una planta in vitro. La respuesta generada a partir de cualquier cambio o estímulo que se presente en el ambiente controlado (Xiong et al. 2001)..... 20

Figura 3. Transcripción del mRNA durante la transcripción del ADN. La información genética desde el DNA al RNA se sintetiza cambiando las bases de ribonucleótidos (A, U, G y C), catalizada por la RNA polimerasa a partir de la secuencia del DNA, dando como resultado un mRNA o transcrito (Proudfoot et al. 2002). 21

Figura 4. Diagrama en el que se muestra un ejemplo de la posible metodología a seguir para realizar el ensamble de un Transcriptoma de novo..... 30

Figura 5. Vía de acción de la familia de las Glioxalosas. La vía de las glioxalosas se divide en dos etapas que involucra a las proteínas glyoxalasa I (GLYI) y glyoxalasa II (GLYII), la acción de estas en conjunto cataliza la conversión de metilglioxal (MG) en D-lactato utilizando glutatión (GSH) como cofactor (Creado con BioRender.com). 36

Figura 6. Arquitectura de dominios de las Glioxalosas tipo I en plantas. Tipos de dominios que pueden presentar GlyI dependiendo de la metaloenzima que tengan presente (Creado con BioRender.com) 36

Figura 7. Estructura del dominio de Glioxalasa tipo II. Dominios presentes en la estructura génica de las glioxalosas tipo II, en donde se observa un dominio Metalo-beta-lactamasa (verde) y un dominio HAGH-C (Azul) (Creado con BioRender.com). 37

Figura 8. Arquitectura de dominios presentes en las Glioxalosas tipo III. En donde se observan los dominios DJ1/Pfp1 en tándem característicos de esta proteína (Creado con BioRender.com)..... 38

Figura 9. Formación del metilglioxal a partir de varias rutas metabólicas. El MG se produce principalmente como producto del metabolismo de los carbohidratos, proteínas y los lípidos y se degrada para formar D-lactato mediante la acción de las enzimas GlyI y GlyII, utilizando como cofactor el GSH. Sin embargo, GlyIII, no utilizando glutatión (GSH) como cofactor (Creado con BioRender.com) 40

Figura 10. Cultivo in vitro de la especie *M. bombycina*. Cultivo in vitro de *M. bombycina* en medio MS + 1 mg de BA..... 55

Figura 11. Electroforesis Y Cuantificación del RNA total de *M. bombycina*. A) Gel de Electroforesis: Verificación de la integridad de las subunidades ribosomales 60s y 25s del RNA total de *M. bombycina*: carriles: 1: Ladder, 2-9: MB1 – MB8 (1ul), B) Electroferograma RIN >4 en cada una de las muestras (Mb1 – MB8)..... 56

Figura 12. Análisis FastQC de lecturas obtenidas de la secuenciación de *M. bombycina*. A) Gráfica que indica la puntuación de calidad de cada una de las lecturas. B) Electroferograma que indica la calidad de cada una de las bases por lectura. 57

Figura 13. Clasificación de datos eggNOG. En el que se obtuvo la clasificación funcionalidad de las proteínas ortólogas del ensamble de novo de *M. bombycina*..... 62

Figura 14. Clasificación de los términos GO del transcriptoma de *M. bombycina*. Términos GO asignados al componente celular, función molecular y proceso biológico..... 64

Figura 15. Distribución de hits en la base de datos Nr del transcriptoma de *M. bombycina*. Distribución de hits de las secuencias obtenidas del transcriptoma de *M. bombycina* con secuencias ya reportadas de otras especies ya secuenciadas. 65

Figura 16. Análisis KEGG de los genes Glyoxalasa del transcriptoma de novo de *M. bombycina*. Identificación de los genes pertenecientes a la familia de las glioxalosas involucradas en la descomposición del metilglioxal en D-Lactato, los cuales están marcados en rojo (4.4.1.5, 3.1.2.6 y 42.1.130). 68

Figura 17. Alineamiento múltiple de secuencias de dominios GLYI. *En los recuadros rojos están marcadas las regiones específicas en la secuencia que indican la presencia del ion Zn^{2+} .* En los recuadros azules se encuentran marcados los recuadros que indican los motivos conservados en las secuencias de MbGLYI. 70

Figura 18. Alineamientos de secuencia de proteínas parciales GLYII. En los recuadros se muestran los dominios y motivos conservados característicos de las GLYII alineadas con secuencias de proteínas de longitud completa de otras plantas. 72

Figura 19. Alineamiento de secuencia de proteínas DJ-1. En los recuadros se indican los motivos característicos de las proteínas DJ-1, alineadas con secuencias de proteínas de longitud completa de otras plantas..... 73

Figura 20. Arquitectura de los Dominios conservados de MbGLYI, MbGLYII y MbDJ-1 de *M. bombycina*. 74

Figura 21. Análisis de conglomerados de MbGLYI, MbGLYII y MbDJ1. Se observan tres grupos; en CI (naranja) se pueden observar las secuencias de proteínas que pertenecen a MbGLYI, que a su vez se dividen en dependientes de Ni^{2+} y Zn^{2+} . El grupo CII (amarillo) y el CIII (verde)..... 75

Figura 22. Estructuras 3D de las proteínas putativas de MbGLYI, MbGLYII y MbDJ-1 de *M. bombycina*. Generadas por modelado de la homología de las estructuras terciarias de las proteínas putativas de GLYI, GLYII y GLYIII de *A. thaliana* con el programa CPHmodels 3.2. 77

Figura 23. Análisis de expresión de los genes MbGLYI, MbGLYII y MbDJ-1. La expresión del gen MbGLYI-1 a las 24 hr se vió aumentada en un 185% en comparación al control. Mientras que las los genes MbGLYI-4, MbGLYI-6 y MbGLYI-9 no se presentó una diferencia de expresión entre tratamientos ni con el control. El gen MbGLYII-1 tuvo una expresión mayor en los tratamientos *ex vitro* y Ds 24h sin embargo el gen MbGLYII-3 tuvo una nula expresión en los tratamientos respecto al control. El gen MbGLYDJ1-1 no tiene una diferencia significativa en la expresión respecto al control. 78

Figura 24. Alineamiento múltiple de secuencias de proteínas putativas MbDREB. Comparación de secuencias del factor de transcripción DREB, con secuencias de proteínas de longitud completa de otras plantas..... 79

Figura 25. Alineamiento múltiple de secuencias de proteínas putativas MbCDPK. Comparación de secuencias del factor de transcripción CDPK con secuencias de proteínas de longitud completa de otras plantas..... 79

Figura 26. Arquitectura de Dominios conservados de MbDREB1, MbDREB2 y MbCDPK1. 80

Figura 27. Estructuras 3D de las proteínas putativas de factores de transcripción DREB y CDPK. Donde: a) MbDREB1, B) MbDREB2 y C) MbCDPK1 de *M. bombycina*, generadas a partir de la homología de las estructuras terciarias de *A. thaliana* con el programa CPHmodels 3..... 81

Figura 28. Análisis de expresión de los genes MbDREB1 y MbCDPK1 en *M. bombycina*. bajo tratamientos de estrés por frío y ABA..... 82

Figura 29. Alineamiento múltiple de α -expansinas (EXPA). Se identificaron con un asterisco (*), se identificaron Triptófanos conservados (W); con un rombo (♦), las Cisteínas (C) encargadas de formar los puentes disulfuro..... 83

Figura 30. Alineamiento múltiple de β -expansinas (EXPB). Se identificaron con un asterisco (*), se identificaron Triptófanos conservados (W); con un rombo (♦), las Cisteínas (C) encargadas de formar los puentes disulfuro. 84

Figura 31. Motivos encontrados entre las subfamilias de α -expansinas y β -expansinas de *M. bombycina*. En el recuadro se observan las secuencias de cada uno de los motivos presentes en las subfamilias α - expansinas y β - expansinas. 89

Figura 32. Estructura 3D de las proteínas putativas de las subfamilias de α -expansinas y β -expansinas. A) MbEXPA8.1, alineada al templado 1N10, en color naranja se observa marcado el dominio 1, en azul el dominio 2, en rosa se destaca el péptido señal y en verde se aprecia la presencia de puentes disulfuro. B) MbEXPB1.1 alineada al templado 2HCZ, de color: Naranja, dominio 1 DPBB; Azul, dominio 2 Alérgeno del Pólen; Verde, parejas de puentes disulfuro. 92

Figura 33. Estadística básica. comparación de la secuenciación: A; pertenece a la muestra MB1_S11_R1_001 y B; pertenece a la muestra MB1_S11_R2_001..... 117



Resumen

Mammillaria bombycina (*M. bombycina*) es una cactácea que se distribuye en la región central de México, en los últimos años ha sido objeto de estudio debido a su fácil adaptación a las condiciones del cultivo *in vitro*, y re inserción a su hábitat natural, así como potencial de planta modelo para uso biotecnológico. En este trabajo se realizó el transcriptoma *de novo* de *M. bombycina* bajo condiciones *in vitro*. Las muestras de RNA, fueron secuenciadas por medio de RNA-seq a través del método Illumina, obteniendo un promedio de 24 millones de lecturas limpias. A partir de la anotación de las secuencias se generaron 84,975 transcritos, de los cuales el 55 % de ellos son unigenes. Entre estos, se identificó la presencia de 13 isoformas de genes que pertenecen a Glioxalasa I, II y III. La expresión qRT-PCR de MbGLYI-1, MbGLYII-2 y MbDJ-1 aumentaron su expresión bajo condiciones *in vitro* y deshidratación a 6 y 24 h. También se identificaron dos isoformas del factor de transcripción DREB y una isoforma CDPK, las cuales fueron caracterizadas y validadas bajo la expresión génica de MbDREB1 y MbCDPK1, los cuales aumentaron considerablemente su expresión bajo congelamiento (1 y 2 h) y con Ácido abscísico (ABA) a 15 min y 1 h. De igual manera se identificaron quince isoformas de genes que codifican a α -expansinas y cinco para β -expansinas, las cuales se analizaron y caracterizaron de manera bioinformática, prediciendo que cada una de ellas está de forma activa en la planta. El ensamble *de novo* del transcriptoma de *M. bombycina* queda como modelo de estudio para trabajos futuros en cactáceas, así como también la posibilidad de obtener herramientas génicas de interés en distintas áreas de estudio.

Abstract

Mammillaria bombycina (*M. bombycina*) is a cactus that is distributed in the central region of Mexico. In recent years it has been the object of study due to its easy adaptation to *in vitro* culture conditions, and reintegration into its natural habitat, as well as a potential model plant for biotechnological use. In this work, the de novo transcriptome of *M. bombycina* was performed under *in vitro* conditions. The RNA samples were sequenced by RNA-seq using the Illumina method, obtaining an average of 24 million clean reads. From the annotation of the sequences, 84,975 transcripts were generated, of which 55% of them are unigenes. Among these, the presence of 13 gene isoforms belonging to Glyoxalase I, II and III was identified. The qRT-PCR expression of MbGLYI-1, MbGLYII-2 and MbDJ-1 increases its expression under *in vitro* conditions and dehydration at 6 and 24 h. Two isoforms of the transcription factor DREB and one CDPK isoform were also identified, which were characterized and validated under the gene expression of MbDREB1 and MbCDPK1, which greatly increased their expression under freezing (1 and 2 h) and with abscisic acid. (ABA) at 15 min and 1 hour. Likewise, fifteen isoforms of genes encoding α -expansins and five for β -expansins were identified, which were analyzed and characterized bioinformatically, predicting that each of them is active in the plant. The de novo set of the transcriptome of *M. bombycina* remains as a study model for future work in cacti, as well as the possibility of obtaining genetic tools of interest in different study areas.

1. Introducción

En la actualidad los organismos vegetales están expuestos a constantes cambios en el ambiente, los cuales la mayoría de ellos son desfavorables o estresantes y comprometen su crecimiento y desarrollo; la mayoría de las veces estos cambios incluyen al estrés abiótico, por lo que buscar soluciones biotecnológicas que ayuden a tolerar o contrarrestar estos efectos en las plantas se ha convertido en un tema primordial en la ciencia. Anteriormente la mayoría de los estudios realizados con relevancia científica se hacían en especies modelo como *Arabidopsis thaliana*, *Zea mays* y *Glycine max*, entre otras ya que, se consideraban especies modelo debido a que se creía que daban solución a problemas de distinta índole solo por el hecho de tener características deseadas en el ámbito económico o bien por su alta reproducción y fácil manejo a nivel de laboratorio. August Krogh, (1929), mencionó que “para un gran número de problemas, habrá una especie de elección, o unas pocas especies, a los cuales convenga estudiar”.

Sin embargo, en años recientes se ha observado que esto no es así, dado que los problemas que aquejan a nuestro entorno superan las necesidades económicas. Esto se refiere a que muchas especies modelo no dan respuestas a diferentes problemáticas, debido a que no mantienen su homeostasis bajo condiciones de aridez, no resisten altas temperaturas ni al estrés hídrico, entre otros. Debido a estas limitantes, el uso de organismos no modelo como son las especies extremófilas son de gran relevancia, ya que dan un panorama general a situaciones en las cuales los organismos tienen capacidades genéticas y metabólicas distintas.

Las cactáceas son uno de los organismos que a través de los años han presentado un sistema desarrollado con un sin fin de respuestas y adaptaciones que les permiten sobrevivir bajo distintos tipos de ambientes, esto gracias a los sensores de estrés y el papel que juegan los orgánulos de estos organismos en la detección en respuesta a distintos estímulos. Sin embargo, aún no se ha explotado el enorme potencial biotecnológico que estos organismos nos pueden brindar a nivel génico.

Un ejemplo de lo anterior es el uso de *M. bombycina* *Quehl* es una planta perteneciente a la familia Cactaceae. Se localiza y es nativa del estado de Aguascalientes y a pesar de que se le considera una planta no modelo, en la Universidad Autónoma de Aguascalientes se han desarrollado varios proyectos en los cuales se ha adaptado a un sistema *in vitro*, observado su fácil y rápida reproducción.

Así como también se han identificado y observado la actividad de genes responsables de la expresión de la proteína LEA tipo 2, entre otros, por lo que el uso de la Secuenciación de Nueva Generación (SNG) en este organismo, para la identificación de genes relacionados con algunas rutas reguladoras y funcionales en la especie bajo condiciones *in vitro* es necesario para comprender como se está comportando a nivel de expresión génica.

Cabe resaltar que en la actualidad hay genes y grupos de genes que no se han identificado en *cactáceas*, en específico en *M. bombycina* cómo son la familia génica de las Glioxalastas, la expresión de factores de transcripción tales cómo DREB y CDPK, genes que codifican a subfamilias de las expansinas que tienen una gran relevancia en la deshidratación y la rehidratación de las plantas.

Por estas razones en este trabajo se realizó el transcriptoma de novo de *M. bombycina*, así como las anotaciones de este bajo condiciones *in vitro*, permitiendo conocer de manera integral la regulación génica coordinada, su función y el trabajo de algunos grupos de genes en conjunto bajo esta misma condición particular.

Se analizaron y validaron algunos genes expresados bajo condiciones de estrés abiótico, por lo que se observó la expresión aumentada de grupos de genes relacionados a la protección y homeostasis celular en plantas como son la familia de genes que codifican a glioxalastas de tipo I, II y DJ-1, factores de transcripción DREB y CDPK, así como también algunos grupos de genes que codifican para α -expansinas y β -expansinas. Esto como base para la búsqueda de genes potenciales y diferentes herramientas moleculares ayudando al enriquecimiento de la información ya existente, teniendo así, un panorama más amplio de cómo se está comportando la planta bajo estas circunstancias, además de tener una evidencia de que *M. bombycina* es ideal como modelo de estudio en *cactáceas*

2. Antecedentes

2.1. Cactáceas

Las cactáceas pertenecen al reino Plantae, división Magnoliophyta, clase Magnoliopsida, orden Caryophyllales y familia Cactaceae, son un grupo de aproximadamente 1,450 especies, las cuales se dividen en aproximadamente cuatro subfamilias *Periskioideae*, *Opuntioideae*, *Maihuenioideae* y *Cactoideae*; ésta última está constituida por un gran número de géneros y especies, la cual se encuentra a lo largo del continente americano (Anderson, 2001).

Las cactáceas poseen adaptaciones en su morfología a partir de su necesidad de sobrevivir a condiciones de aridez, un ejemplo de esta evolución es la de la extrema succulencia en la mayoría de sus organismos, resalta de una manera extenuante la presencia de estos individuos en las regiones secas del mundo.

Son fáciles de identificar a simple vista, a pesar de su extraordinaria diversidad de formas de crecimiento; diferentes especies de cactus han sido estudiadas como modelos en plantas anatómicas y estudios fisiológicos, sin embargo, son pocos los estudios realizados a nivel molecular, por lo que abren una brecha importante para entender cómo llevan a cabo todos los procesos metabólicos necesarios para sobrevivir a ambientes extremos (Shishkova et al., 2008, Hernández- Hernández et al. 2011).

Uno de los géneros representantes de la subfamilia *Cactoideae* es el de *Mammillaria*, ya que hasta la fecha lo componen 166 especies, destacando que alrededor del 90% de estas se encuentran en nuestro país (Hernández et al. 2011)

Un ejemplo característico y poco estudiado de la subfamilia antes mencionada, es *Mammillaria bombycina* Quehl (*M. bombycina*), la cual, se encuentra en los estados de Jalisco y Aguascalientes, crece en el matorral subtropical en el municipio de Calvillo (en el suroeste de la entidad) y en selva baja caducifolia en el mismo municipio (Meza-Rangel et al. 2014), así como también en el Área Natural Protegida (ANP) Monumento Natural Cerro del Muerto se encuentra ubicado en los municipios de Aguascalientes y Jesús María (Rivera et al. 2018). Esta especie se encuentra sujeta a protección especial, debido a que el sobre pastoreo, el cambio de uso de suelo que hay en la región y sobre todo la sobre explotación por su gran valor ornamental han mermado de forma considerable a esta especie (NOM-059, 2010, Lustre et al. 2012).

M. bombycina es una planta en forma de bulbo, con los ápices ligeramente hundidos cubiertos por pelos blancos, mide aproximadamente 20 cm de alto y 6 cm de diámetro, tiene axilas con lana blanca abundante y cerdas, en el ápice esta lana cubre los tubérculos, las areolas son redondeadas al principio con poca lana y después son alargadas y desnudas, tienen espinas blancas cortas y espinas curvas marrones mucho más largas, rígidas y delgadas (CONABIO, 2008).

Las cactáceas son plantas que tienen un Metabolismo Ácido de las Crasuláceas, al cual también se le conoce como metabolismo CAM (por sus siglas en inglés), mecanismo que ha surgido de la necesidad de disminuir la pérdida de energía durante la fotorrespiración, el cual se caracteriza por tener dos etapas:

A) La que se lleva a cabo durante la noche, en la que efectúa la formación del aceptor primario del CO_2 , fosfoenol-piruvato (PEP) a partir de carbohidratos no estructurales en las células fotosintéticas, también se genera la fijación del CO_2 por la enzima PEP carboxilasa (PEPC) en el citosol y síntesis del ácido málico, así como el almacenaje del ácido málico (como ion malato) en la vacuola central de las células fotosintéticas.

Por esta razón la apertura de los estomas se da durante la noche, de esta manera se minimiza la pérdida de agua por transpiración, puesto que la temperatura de los tejidos es más baja y por tanto los gradientes de concentración de vapor de agua desde la hoja al aire son considerablemente menores que los valores en el día. Este mecanismo de concentración de dióxido de carbono permite disminuir la probabilidad de que entre un O_2 en el sitio activo de la Ribulosa-1,5 bifosfato carboxilasa oxigenasa (Rubisco) por lo que la eficiencia fotosintética es mayor, este tipo de plantas deben hacer frente al calor extremo y a la sequía y mantener una toma neta de CO_2 durante todo el año.

B) Durante el día: se lleva a cabo la liberación del malato de la vacuola hacia el citosol, después descarboxilación del malato en el citosol se comienza a liberar de CO_2 y formación de compuestos de tres carbonos (piruvato o PEP); se activa la asimilación del CO_2 liberado en los cloroplastos por la enzima Rubisco, seguida por el ciclo de Calvin-Benson y la regeneración de carbohidratos de almacén o gluconeogénesis.

La liberación de malato de la vacuola durante el día debe estar estrechamente orientado al metabolismo para que coincida con la tasa de descarboxilación de

malato en el citoplasma y el flujo de carbono en la gluconeogénesis, sin embargo, las especies con metabolismo tipo CAM se ha estudiado relativamente poco a nivel molecular, sobre todo en la expresión génica que se lleva a cabo durante estos procesos metabólicos (Rodríguez et al. 2016).

2.1.1. *Mammillaria bombycina* como modelo de estudio

En la actualidad son muy pocas las especies a las que se les considera como modelo de estudio, ya que, la mayoría de ellas tienen una gran importancia económica, son de fácil manejo en el laboratorio, su crecimiento y desarrollo es relativamente rápido, entre otros. Sin embargo, las características genómicas de estas especies distan de las características genómicas ideales de un organismo modelo (Russell et al. 2017).

El uso de especies no modelo para su estudio en el área de la ciencia, es de gran importancia, debido a que muchas de estas tienen un gran potencial biotecnológico, un ejemplo de esto, son los trabajos que se han realizado por la Universidad Autónoma de Aguascalientes en la especie *M. bombycina*, entre los que se encuentran los estudios realizados por meza- Rangel y colaboradores en el año 2014, en donde identificaron su ubicación geográfica en el estado de Aguascalientes, así como la importancia económica y ecológica de la misma, el relevante estudio molecular realizado por Hernández- Camacho y colaboradores (2016), en donde identificaron proteínas tipo LEA y vieron su expresión bajo distintos tratamientos de estrés abiótico su propagación *in vitro*, siendo este uno de los más importantes, ya que, *M. bombycina* es una especie que se encuentra sujeta a protección especial según la legislación ambiental federal mexicana debido a la sobreexplotación y saqueo (NOM-059, 2010). Por lo que tener a este organismo bajo condiciones *in vitro* permite aprovechar todos los beneficios que nos brinda esta técnica como es la conservación de germoplasma, la propagación masiva para el uso de las plántulas en experimentos donde se utilice un gran número de plantas, producción de metabolitos secundarios, mejoramiento genético y todo lo anterior bajo condiciones controladas, de esta manera, las variables de los experimentos realizados en estos organismos son muy reducidas (Hernández et al. 2016).

Por esta razón el uso de *M. bombycina* es ideal para realizar el análisis del Transcriptoma bajo condiciones *in vitro* y así conocer algunos de los procesos y rutas metabólicas que este organismo utiliza para mantener su crecimiento y desarrollo bajo estas circunstancias,

además de enriquecer la casi nula información que se tiene de organismos en cultivo *in vitro*.

2.2. Transcriptoma

Cuando se habla del transcriptoma, se hace referencia al análisis de todos los transcritos (mARN) presentes en una célula, tejido u órgano, en un tiempo y situación fisiológica específica en la que se encuentra el organismo, teniendo en consideración que a cada cambio de estímulo o condición la expresión de los transcritos puede variar, esto debido a que los genes se transcriben de manera individual, por lo que si el organismo en cuestión recibe estímulos ya sean positivos o negativos, van a encender o apagar los genes necesarios para mantener la homeostasis del mismo, iniciando o bloqueando la transcripción de cada uno de ellos (Rodríguez et al. 2018).

De este modo, la cantidad de moléculas transcritas es proporcional a la respuesta a una condición determinada en la cual un gen tiene una participación importante. De igual manera hay genes que permanecen apagados en la célula, sin embargo, bajo ciertos estímulos o circunstancias particulares o atípicas estos genes recobran su actividad iniciándose entonces así la transcripción, de esta manera las respuestas a las preguntas de dónde, cómo y cuándo es generado un transcrito bajo una condición dada es fundamental para el entendimiento de la actividad biológica de un gen (Bhowal et al. 2020).

Hoy en día hoy se desconoce la función de la totalidad de genes involucrados en una situación específica, radicando aquí la importancia del estudio amplio y sustancial del transcriptoma del organismo de interés, permitiendo conocer de manera integral la regulación génica coordinada, su función y el trabajo de los grupos de genes en conjunto bajo una misma condición particular, o bien, observando de manera individual un solo gen de interés conociendo los patrones empalme, modificaciones transcripcionales, entre otros (Marguerat y Bahler, 2010).

La realización del transcriptoma completo consta de 4 etapas para que las secuencias obtenidas de la secuenciación masiva tomen relevancia y puedan traducirse en información legible. En la primera etapa, se debe de definir de manera adecuada el organismo del cual se va a obtener el material genético (en este caso mARN). La segunda etapa, es la división de la muestra en bibliotecas o librerías de cDNA de un tamaño determinado, en la cual se le añadirán un pequeño índice (barcoding) que permite la secuenciación de varias muestras

TESIS TESIS TESIS TESIS TESIS

y un adaptador que permite que los fragmentos a secuenciar puedan adherirse al soporte del secuenciador que se va a utilizar. La tercera etapa es la secuenciación propia de la muestra marcada y la cuarta y última etapa es el análisis bioinformático post secuenciación de las secuencias obtenidas para poder realizar la interpretación de los resultados (Hernández et al. 2019).

En la cuarta etapa es importante destacar que a partir de la tecnología de RNA-seq se puede aplicar para especies modelo o especies muy estudiadas, en las cuales se tiene información genómica previa, como puede ser un genoma de referencia, el cual, al obtener las lecturas del secuenciador, primeramente, se realiza un mapeo, empatando las secuencias obtenidas con las ya existentes (Shi et al. 2010, Voineagu et al. 2011). Sin embargo, cuando se obtienen las lecturas secuenciadas de un organismo no modelo o bien, de un organismo del cual no se obtiene ninguna información genómica previa, se le considera que se realizará un análisis transcriptómico *de novo*, en el cual, al obtener las lecturas previamente secuenciadas, primeramente se tiene que realizar un ensamble de las secuencias cortas en transcripciones que van a permitir la reconstrucción completa, la identificación y clasificación de todos los genes expresados bajo una condición en particular, visualizar los niveles de transcripción de cada uno de ellos y separar isoformas (Grabherr et al. 2011; Zhao et al. 2011).

Como un ejemplo de lo anterior, se encuentra el transcriptoma realizado por Ibarra y Zamudio (2015) en el organismo *Lophophora williamsii*, en el que observaron el nivel de expresión de varios genes involucrados en la producción de mezcalina, involucra las anotaciones del transcriptoma y la validación de algunos genes de interés al cambio de una circunstancia específica.

De igual manera el trabajo realizado por Rodríguez et al (2018) en el que centraron su investigación en la expresión diferencial del Transcriptoma en la especie *Pachycereus pringlei*, donde se analizó la regulación génica en el crecimiento determinado de la raíz primaria como consecuencia del agotamiento del meristemo apical (RAM), por lo que se llevó a cabo la construcción del transcriptoma *de novo* en tres etapas del desarrollo de esta especie, haciendo una comparación de cada una de las diferencias de expresión transcripcional del crecimiento primario del ápice de la raíz para cada una de las etapas mencionadas.

2.3. Relevancia de la expresión génica plantas

Las plantas a través del tiempo han presentado un sistema desarrollado con un sin fin de respuestas y adaptaciones que les permiten sobrevivir bajo distintas circunstancias, un ejemplo de esto son los sensores de estrés y el papel que juega la función de los orgánulos en la detección en respuesta a distintos estímulos (Lüttge, 2004). Al cambiar la condición en la que se encuentra el organismo de una forma desfavorable, las plantas presentan 2 señales, en las cuales responden al estrés primario mientras que otros, responden principalmente con señales secundarias. Las señales primarias se componen principalmente del estrés osmótico, la segunda señal se compone de varios y complejos efectos, como son, estrés oxidativo, daño a los componentes celulares como son los lípidos de membrana, desnaturalización de proteínas y ácidos nucleicos, disfunción metabólica, etc, (Zhu JK, 2002). Sin embargo, los mecanismos genómicos y moleculares no han sido bien estudiados.

Los cambios que se generan en la planta producidos por cualquier tipo de estrés en el plegamiento de proteínas se asocian a las señales generadas en los orgánulos en conjunto como lo son el retículo endoplásmico, cloroplasto, mitocondria, peroxisoma, núcleo y pared celular, esto para regular la expresión de genes sensibles al estrés y restaurar la actividad de aquellos que mantienen la homeostasis celular, como se muestra en la Figura 1.

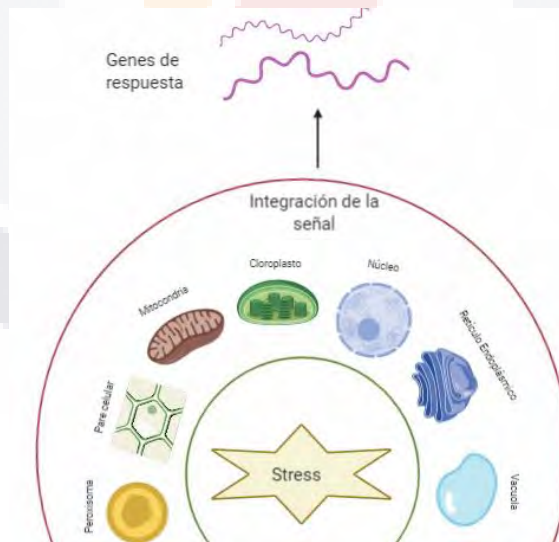


Figura 1. Respuesta de orgánulos celulares ante el estrés. Las señales generadas por el estrés de todos los orgánulos en conjunto para mantener la homeostasis celular, genera una adaptación de la planta ante varios tipos de estrés (Zhu, 2016).

La señal inicial provocada por el estrés abiótico es percibida por receptores que se encuentran en la membrana celular, que con la activación inician o suprimen una cascada de transducción de señales de las células, mediante eventos de fosforilación y desfosforilación por medio de proteínas cinasas y fosfatasas (Zhu, 2016). Luego, la señal se retransmitiría a varias ubicaciones subcelulares como el núcleo en el que se activan factores de transcripción que inducen la expresión de grupos de genes específicos que codifican proteínas que están involucradas en la respuesta al estrés en diversos procesos para ayudar a las plantas a sobrevivir bajo condiciones desfavorables Figura 2. (Xiong et al. 2001; Zhu, 2016).

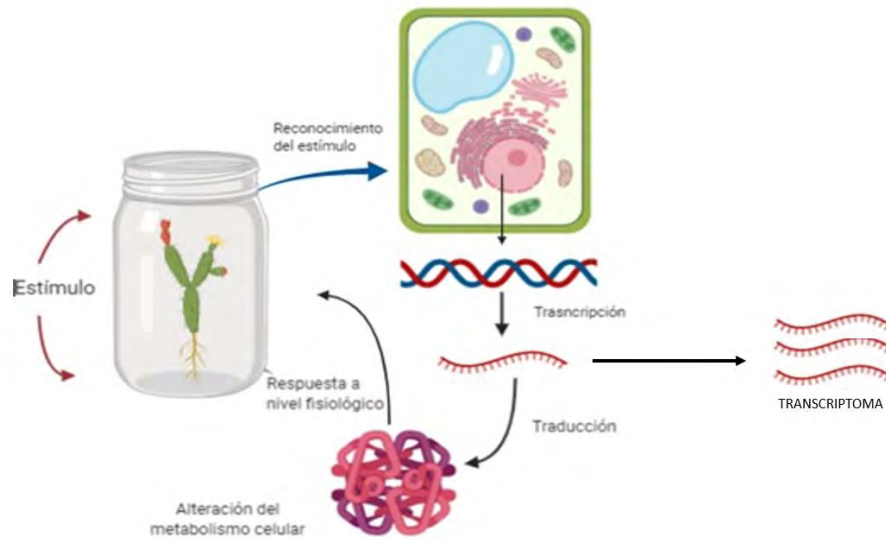


Figura 2. Respuesta génica bajo estrés en una planta in vitro. La respuesta generada a partir de cualquier cambio o estímulo que se presente en el ambiente controlado (Xiong et al. 2001).

Como se mencionó anteriormente, los estímulos intra o extracelulares activan la regulación de la expresión de genes en la que se van a llevar a cabo dos etapas: la transcripción y traducción del ARN (Figura 3).

La transcripción es el primer paso de la expresión génica, tiene como fin hacer una copia de la cadena molde del ADN de un gen para producir un ARN mensajero (mARN) con la ayuda de las ARN polimerasas, se lleva a cabo en tres pasos: iniciación, elongación y

terminación, dando como resultado un pre-mARN, y después iniciar el procesamiento de éste, para llegar al mARN maduro, también llamado transcrito, el cuál en el inicio de la cadena va a tener una CAP 5' y en el final una cadena de poli A, con el fin de darle estabilidad a la molécula de ARNm (Liu et al. 2014). Aunado a esto se efectúa el corte y empalme de la secuencia, quitando todos los intrones (la parte no codificante) y empalmado los exones (secuencias codificantes). Cabe resaltar que la transcripción de cada gen se lleva a cabo de manera independiente para cada uno de ellos; la cantidad de moléculas de mARN producidos depende de la función y de la relevancia de este para cada respuesta celular (Proudfoot et al. 2002).

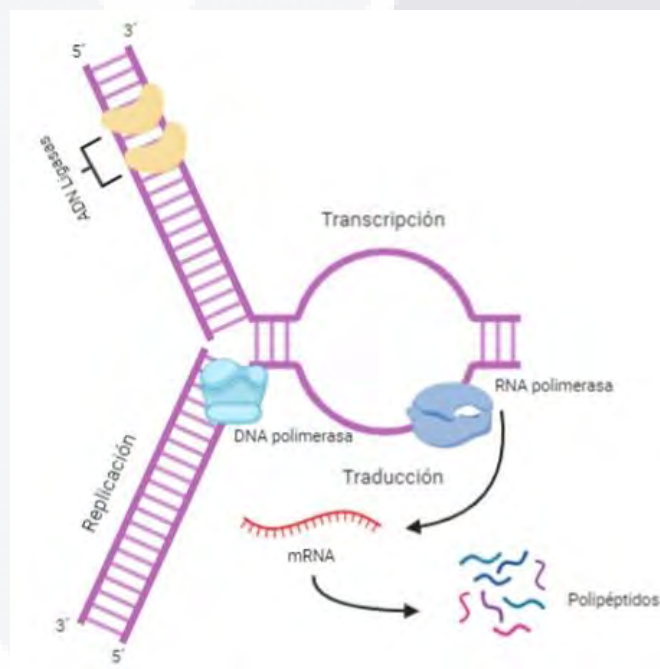


Figura 3. Transcripción del mARN durante la transcripción del ARN. La información genética desde el DNA al RNA se sintetiza cambiando las bases de ribonucleótidos (A, U, G y C), catalizada por la RNA polimerasa a partir de la secuencia del DNA, dando como resultado un mRNA o transcrito (Proudfoot et al. 2002).

La importancia de descifrar la expresión génica de una planta bajo una cierta condición radica en comprender los mecanismos por los cuales regula y mantiene su crecimiento y desarrollo. Estos mecanismos involucran la identificación de grupos de genes, así como su modulación, el número de copias expresado de cada gen, la transcripción, además de que los mecanismos que controlan dicha expresión no se han estudiado con detalle, por lo que, no se tiene la oportunidad de conocer cómo trabajan los genes en conjunto; como

consecuencia, por muchos años, se han utilizado un gran número de técnicas moleculares que permiten la secuenciación del genoma o una parte del mismo y su monitoreo a través del tiempo (Liu et al. 2014).

Una de las primeras técnicas utilizadas para la detección de moléculas de ARN es el Northern blot, la cual hace posible la comprensión de la función génica del organismo en cuestión, en la que se genera un patrón para la regulación génica coordinada, permitiendo conocer la función y trabajo de varios genes en conjunto bajo una condición específica, la base de esta técnica es la electroforesis en gel, en la que se separan por tamaño las moléculas de ARN, después, los fragmentos de ARN que corrieron en el gel se transfieren a una membrana, la cual, se expone a una sonda de ADN marcada con una etiqueta química o radioactiva, por lo que si la sonda se une a la membrana, entonces se conocía que la secuencia de ARN complementaria estaría presente en la muestra (He et al. 2013).

Otra de las técnicas utilizadas son los microarreglos, la cual permite que el análisis en paralelo de la expresión cADN o ADN, explorando el transcriptoma y las diferencias genómicas para distintos organismos para hacer más sencillo el análisis de su metabolismo bajo ciertas circunstancias, medir o cuantificar la expresión génica, identificar mutaciones en genes específicos, etc. El principio se basa en la hibridación de dos cadenas de material genético, emparejándose entre sí formando pares de bases de nucleótidos complementarios, estas secuencias híbridas serán marcadas con fluorescencia en la cual, la intensidad de señal emitida depende de la cantidad de muestra objetivo-hibridada que se unen a las sondas fluorescentes presentes en ese punto (Bhowal et al. 2018).

En la actualidad, existen diferentes tecnologías de secuenciación de nueva generación (NGS por sus siglas en inglés) que han logrado un gran avance en la ciencia en los campos de la genómica y transcriptómica, ya que estas, permiten generar información con una gran eficacia, altos rendimientos y a muy bajo costo, así como también crear nuevos métodos para el entendimiento detallado de la secuenciación, permitiendo la comprensión más a fondo los procesos de expresión génica un ejemplo de estas técnicas es el RNA-seq (Bhowal et al. 2020).

2.4. RNA-seq

La implementación de la tecnología de secuenciación directa de ARN, se le conoce como RNA-seq, la cual permite la caracterización casi completa de los mARN que se transcriben bajo un momento dado, se pueden obtener millones de lecturas cortas en paralelo (según la tecnología que se utiliza), una de las características más relevantes de esta técnica es que no se necesita un Genoma de referencia para obtener información. (Torres et al. 2021).

El RNA-seq es un implemento transcriptómico de alto rendimiento que se caracterizan por utilizar una combinación de nuevas estrategias utilizando la secuenciación directa del ARNm, en los que intervienen distintos factores de los que depende el éxito de esta, como lo son: la preparación de muestras, secuenciación, además de nuevos métodos de alineamiento y ensamblaje, cabe resaltar que la secuenciación masiva genera una gran cantidad de datos, reduciendo costos y el tiempo del análisis (Ward et al. 2012).

El principio de esta técnica de secuenciación genética cifrada en el ADN se debe a que la información que contiene la célula en sus genes se expresa a través de mecanismos de transcripción y traducción, a partir del cual se generan moléculas de ARNm y proteínas, los mecanismos celulares como la replicación, diferenciación, división celular, etc. Dando como resultado los caracteres macroscópicos tales como los rasgos fenotípicos, morfológicos, funcionales y de respuesta a estímulos, por lo que son producto de la expresión diferencial de genes, esto debido a la activación o represión de estos como respuesta a un estímulo intra o extracelular (Proudfoot et al. 2002).

De este modo la regulación de la transcripción va a depender de la unión de activadores o represores en los elementos del promotor ubicados en la región 5' de la secuencia codificante, ya que estos van a determinar la síntesis de ARNm que se va a producir durante la transcripción, la cual va a estar constituida por los GTFII (factores generales de transcripción) y la ARN polimerasa tipo II (Ward et al. 2012).

Debido a lo anterior, se siguen una serie de pasos entre los cuales el de mayor importancia es la extracción del ARNm. Cabe destacar, que para proceder a realizar una secuenciación de buena calidad es crucial que la integridad de ARNm extraído sea la adecuada asegurándose que no exista ARNr, ya que en presencia de este la construcción de la biblioteca de cDNA para la secuenciación se vería afectada, de este modo, las muestras deben de ser analizadas mediante un electroferograma Agilent Bioanalyzer (Agilent Technologies) para ver la distribución de ARNm de alto peso molecular. Del mismo modo

se utiliza el software de Agilent, que compara varias métricas de integridad de ARN y proporciona un número de integridad de ARN (RIN), las muestras con RIN superiores a ocho generalmente se consideran de alta calidad, además que la concentración mínima requerida de cada muestra debe de superar los 25 ng por cada 20 µl.

Existen dos métodos con los cuales se puede enriquecer la muestra de mARN: 1) PolyA-capture, en la que la captura de ARN poliadenilado se lleva a cabo por la selección positiva de los ARNm, siendo esta técnica una de las más utilizadas; 2) ARNr-depletion que se refiere a la eliminación de rARN; una vez que se tenga enriquecido el mARN se construyen las bibliotecas de cDNA, las técnicas utilizadas en este proceso dependen de los intereses que tenga el investigador; dependiendo del éxito de la extracción y enriquecimiento del mARN, es el éxito que se obtendrá al realizar las construcciones de las bibliotecas de RNA, en la cual la muestra se va a dividir en varias partes, a las que el tratamiento posterior dependerá del tipo de plataforma de secuenciación que se va a utilizar (Rodríguez y Shishkova, 2018).

2.5. Descripción general de las plataformas de secuenciación para RNA-seq

Existen varias plataformas en las que se puede realizar la secuenciación como son Roche/450, Solexa/Illumina, SOLid/Life Technologies, la elección de esta dependerá de los resultados que el investigador desee obtener, por ejemplo, se puede optar por secuencias cortas o largas, a las cuales se les llaman lecturas, además de que se puede secuenciar de dos tipos de formato diferente: las single-end o lecturas simples, cuando la secuenciación se realiza sólo en alguno de los extremos de cada fragmento y las paired-ends o lecturas apareadas, cuando a partir del mismo fragmento se secuencian los dos extremos (Soto y López, 2012). De igual forma el investigador puede elegir entre dos vertientes para el análisis del transcriptoma utilizando tecnologías de secuenciación de nueva generación, la elección dependerá nuevamente de los resultados que se deseen obtener, si se requiere que los resultados tengan un genoma de referencia del organismo del cual se obtuvo la muestra para que primero se puedan alinear las lecturas con las secuencias ya reportadas y después se ensambla, una de las ventajas de esta vía es que además de detectar eventos de *splicing* alternativo del mARN; por el contrario, si no se tiene ningún genoma de referencia, se le conoce como ensamble *de novo*, de este modo primero se ensambla y luego se alinean las secuencias, entre otros (Jochen, 2013).

En la actualidad las tres principales tecnologías de secuenciación de bibliotecas de cDNA son: por pirosecuenciación, la secuenciación en tiempo real de molécula única o SMET-seq y secuenciación por síntesis, las cuales se describen a continuación:

2.6. Pirosecuenciación

Esta tecnología desarrollada en el año 2005 introducido a la industria por 454 Life Sciences de Roche, se basa en la piroseceunciación. De esta técnica se obtienen longitud de secuencias de alrededor de 600pb, con un rendimiento de corrida de 0.45 GB, emplea a la PCR en emulsión para la amplificación de los fragmentos. La técnica radica en la detección de un pirofosfato, dado que cada uno de ellos tiene adaptadores en los extremos de cDNA, se van a unir a unas microperlas, las cuales tienen los oligonucleótidos unidos a la superficie debido a que son secuencias complementarias a los adaptadores, por lo que se obtienen un gran número de moléculas idénticas que emiten altas intensidades de cada señal de cada lectura (Ergan et al. 2012).

Al término del enriquecimiento de las micro perlas, cada una de ellas se van a depositar en una placa PTP (PicoTiterPlate) en las que se va a llevar a cabo la desnaturalización en la que intervienen enzimas como la ATP sulfurilasa, ADN polimerasa, luciferasa y apirasa junto con los sustratos luciferina, adenosin-5-fosfosulafo (APS) y deoxinucleótido particular (dNTP).

Este último se agrega con el fin de que cuando la AND polimerasa incorpore el dNTP a la nueva cadena que se está generando, así, libere un pirofosfato, el cual es convertido a adenosin trifosfato (ATP). Esta reaccionará con la luciferasa y formará oxiluciferina por lo que se produce luz, la cual va a ser equivalente a las cantidades de ATP que se produjeron anteriormente (cantidad de nucleótidos incorporados) y la señal será detectada con una cámara CDD (charge coupled device) y será analizada por el programa de la misma para emitir la secuencia. Sin embargo, tiene algunas desventajas, como son la generación de homodímeros, inserciones y deleciones (Morozova et al. 2009, Ergan et al. 2012, Garzón et al. 2012, Soto Sedano et al. 2012).

2.7. Secuenciación en tiempo real de molécula única o SMET-seq

El fundamento de este método fue desarrollado por la empresa Pacific Biosciences, genera una longitud de lecturas de 964 pb. En esta tecnología no requiere la realización de PCR. Sus principales aplicaciones son la secuenciación de transcriptomas y la re-secuenciación de genomas, esto permite tener secuencias más largas que otras tecnologías. Sin embargo, los precios de esta técnica siguen siendo muy altos, otra de las características es que no genera clusters de secuenciación y las secuencias obtenidas tienen un formato BAM.

El principio de esta tecnología radica en desarrollar la secuenciación de molécula única de una molécula de ADN en una estructura que contiene una única enzima de ADN polimerasa, se realiza sobre un chip que está formado de nano pozos, y que utiliza la guía de onda de modo cero (ZMW).

Dentro de estos nano pozos se encuentra la DNA polimerasa activa, a la cual se le añade el ADN molde de cadena sencilla, en donde la ZMW permite que la luz penetre y exista un espacio en el cual va a activar a la ADN polimerasa. La señal que se emite de la incorporación de cada nucleótido a este espacio por medio de la ADN polimerasa es detectada durante la síntesis de ADN, esto es posible debido a que cada una de las cuatro bases nitrogenadas están marcadas con una molécula de naturaleza diferente de fluorescencia que se une a cada grupo fosfato del nucleótido cuando se inserta en la cadena de ADN. Así, el fluoróforo-fosfato es liberado como resultado de la formación del enlace fosfodiéster durante la síntesis de la cadena de ADN, lo que permite la identificación de cada una de ellas durante la secuenciación por síntesis (Morozova et al. 2009; Chaisson et al. 2015).

2.8. Secuenciación por síntesis

Esta plataforma fue desarrollada por Solexa y comercializada por la empresa Illumina, esta tecnología desarrolla la secuenciación por síntesis utilizando como método de amplificación la PCR en puente. Tiene la característica de generar secuencias cortas de aproximadamente 75 – 100 pb con un rendimiento de corrida de 18 a 35 GB (Hernández et al. 2019).

El principio de esta técnica radica en la secuenciación de librerías de cDNA, las cuales tienen adaptadores (específicos para el equipo Illumina que se va a utilizar) en los extremos, de tal forma que se unen a oligonucleótidos ligados a una celda de flujo, estos primers pueden estar en sentido o anti sentido creando puentes que facilitan la amplificación de las secuencias generadas a las que se le llaman amplicones, de modo que estos se mantendrán adheridos a la celda de flujo hasta que comience la desnaturalización de los mismos para permitir la amplificación de la siguiente cadena de manera exponencial hasta formar millones de cadenas llamadas clusters. Para que la señal sea detectada, la secuenciación se lleva a cabo mediante la adición de 3'-O-azidometil-dNTPs, que son nucleótidos que tienen añadido un fluoróforo específico para unirse a cada base nitrogenada y una modificación en la posición 3' que tiene la función de detener la síntesis cuando se adiciona un nucleótido a la cadena, cada vez que esto sucede, se excitan los fluoróforos por la acción de un láser, que al mismo tiempo se obtienen cuatro imágenes, cada una en la longitud a la cual emiten la luz los fluoróforos excitados, de esta manera se pueden identificar los nucleótidos que se añaden a cada cluster en particular. Al terminar de capturarse todas las imágenes, son removidos el nucleótido bloqueador y el fluoróforo para un nuevo ciclo para una nueva síntesis, teniendo en cuenta que cada vez que se lleva a cabo un ciclo baja su calidad, en otras palabras. La calidad de la lectura de bases se disminuye de manera inversamente proporcional al número de ciclos de secuenciación; a pesar de que esta plataforma es una de las más usadas por los investigadores, tiene ciertas desventajas como son: baja calidad en multiplexado de muestras, sustituciones, lecturas cortas, entre otros; las secuencias obtenidas tienen el formato. fastqc, (Morozova et al. 2009; Soto-Sedano et al. 2012, Jochen, 2013, Rodríguez et al. 2018).

En este proyecto se utilizó esta tecnología debido a las múltiples ventajas, como son la alta cobertura de secuenciación, la eficacia y rapidez y los bajos costos, además de que al ser una plataforma altamente utilizada en la actualidad, el tipo de adaptadores que se utilizan para la creación de las bibliotecas de cDNA secuenciado son compatibles con diferentes programas de uso libre como FastQC, FastX tool kit, Trimomatic, Trynity, Trimotate, entre otros soportados en los sistemas UNIX/LINUX.

2.9. Análisis Bioinformático de las secuencias obtenidas de la plataforma Illumina

La enorme cantidad de información obtenida de las tecnologías de secuenciación masiva de datos requiere de herramientas bioinformáticas para su anotación, identificación y cuantificación de los genes presentes en el análisis transcriptómico de un organismo, haciendo de este un trabajo multidisciplinario debido a que aplica conocimientos en matemáticas, biología, estadística, programación (BASH, PHYTON, R, etc.), ya que la mayoría de los programas están basados en sistemas operativos como UNIX (Linux, UBUNTU, etc.) son los más apropiados para el manejo de los datos, además de que son de código libre. Sin embargo, ninguno de estos programas (FastQC, Trimmomatic, FastX-toolkit, Fastp, Trinity, Velvet, Trinotate, etc.) tienen una interfaz gráfica, por lo que el uso de la terminal y el conocimiento del lenguaje Bash como compendio de comandos es indispensable para tener éxito al hacer uso de estos programas bioinformáticos. Todo lo anterior es con el fin de obtener resultados de alta calidad Y por el hecho de conocer los datos obtenidos y hacer un buen uso de estas herramientas. Conjuntamente a los requerimientos del *software*, también existen requerimientos esenciales para el *hardware* debido a que para la simple instalación de ciertos programas. Como ejemplo, están los ensambladores la mayoría de ellos requiere de 16 GB de memoria Ram, 4 núcleos y 1 TB de almacenamiento. Uno de los repositorios más utilizados es Biolinux (<http://environmentalomics.org/bio-linux/>), basado en el sistema operativo Linux que hace al investigador biológico el análisis de secuencias masivas de datos más amigable, ya que, tiene lenguajes y programas bioinformáticos que facilitan el manejo y comprensión de gran cantidad de datos (Hernández et al. 2019).

De este modo como primer paso para dar tratamiento a las lecturas que se obtuvieron del secuenciador deben de pasar por ciertos límites de calidad, debido a que aún tienen secuencias que no son informativas y que al ensamblarlas pueden causar problemas de error de contaminación, estas secuencias basura pueden ser bases ambiguas, remanentes de los adaptadores que no se pudieron remover durante la secuenciación, lecturas completas pueden estar sobrerrepresentadas como por ejemplo la duplicación de PCR, colas polyG, colas polyA, entre otros (Chen et al. 2018).

La herramienta bioinformática FastQC es la más utilizada para el control de calidad en los datos de la secuencia sin procesar obtenidos de las plataformas de secuenciación de alto rendimiento. Sin embargo, sólo se puede procesar información con los formatos BAM, SAM o FastQ. Las ventajas son muchas, por ejemplo, suministra una visión amplia de las secuencias para poder identificar en que áreas podría haber problemas de secuenciación,

resume los gráficos y las tablas para su rápido análisis, entre los que se encuentran: la calidad media para cada base, la calidad de la secuenciación dependiendo del estado de la celda de flujo (Flow cell), la calidad que presentan cada una de las secuencias, calidad de contenido de secuencia base, porcentaje de la relación G-C, Contenido de Bases N, entre otros, además de que el formato en el que genera la información es HTML, (Anderson et al. 2010).

El principio que usa el programa FastQC para evaluar las secuencias provenientes del secuenciador, es a través del Nivel de calidad Phred (Phred quality score o Q), el cual se define como la propiedad relacionada de forma logarítmica con las probabilidades de error de las bases, $Q = -10 \log_{10} P$, (en donde P es la probabilidad de que la base secuenciada sea la base en cuestión). Sin embargo, los parámetros de calidad cambiarán dependiendo de la plataforma en la que fueron secuenciados. En este estudio, sólo tomarán relevancia los rangos de calidad descritos para la plataforma de secuenciación Illumina. Ya que es el formato en el que se obtienen las secuencias de esta plataforma es FastQ, por lo que el formato codifica un rango de calidad de Phred es de 0 a 62 usando ASCII (Código Estándar Americano para Intercambio de Información) de 64 a 126 y en datos de lectura sin procesar de 0 a 40; al término del análisis FastQC y al haber identificado cual es la parte de la secuencia que debe de ser removida, se procede a hacer el trimming (recorte), para ello se pueden utilizar varias herramientas bioinformáticas como FastX-toolkit, Trimmomatic,

Fastp, entre otros, como se observa en la Figura 4 (Anderson et al. 2010, Zhang et al. 2017, Hernández et al. 2019).

Una vez que se les ha dado el tratamiento adecuado a las secuencias, se procede a realizar el ensamble del transcriptoma, para llevarlo a cabo, existen 2 estrategias computacionales, la decisión de usar alguna de ellas depende de las características de las secuencias obtenidas y de si se tienen un genoma de referencia o no. La primera estrategia, se realiza un mapeo con el genoma de referencia antes de ensamblar el transcriptoma, entre los programas más utilizados se encuentran Scripture y Cufflinks (Trapnell et al. 2010, Guttman et al. 2010). Mientras que, en la segunda, se ensambla el transcriptoma y después se realiza el mapeo, la mayoría de los programas que se basan en este protocolo son algoritmos gráficos de Bruijn (existen sus excepciones), por ejemplo, ABySS, SOAPdenovo, Velvet, Trinity, entre otros, (Birol et al. 2009, Li et al. 2009, Grabherr et al. 2011).

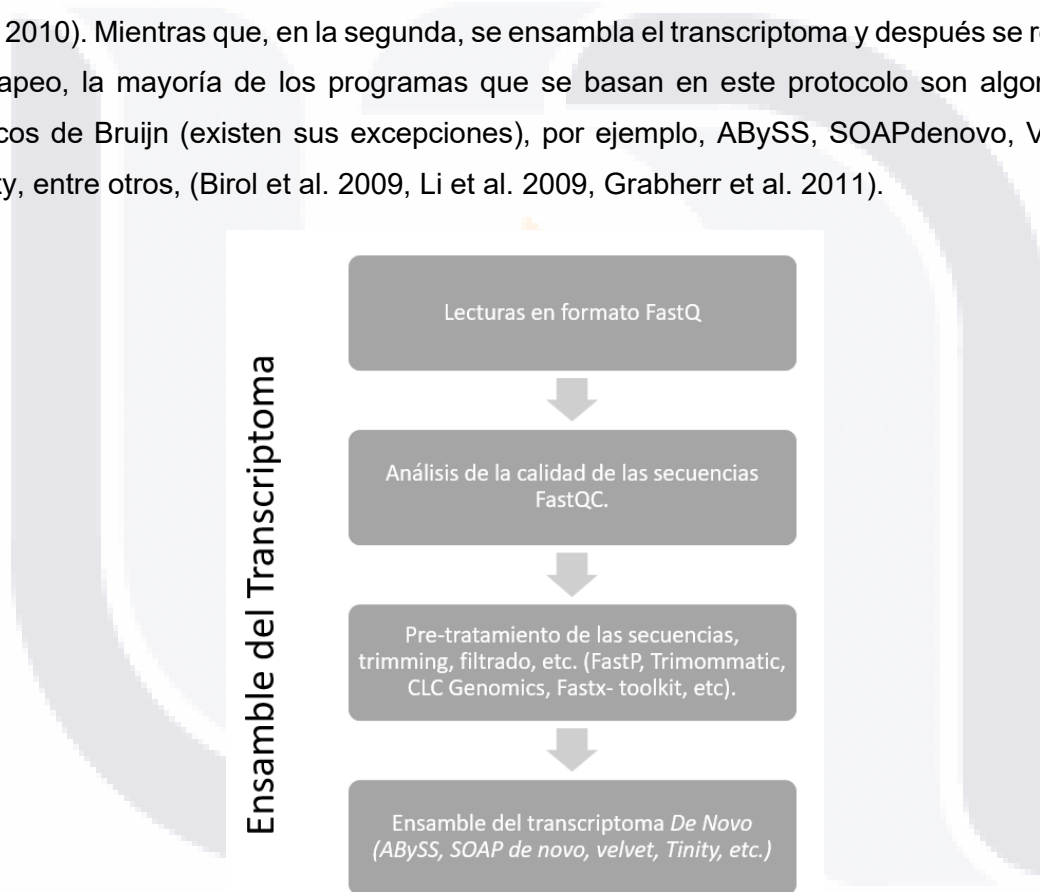


Figura 4. Diagrama en el que se muestra un ejemplo de la posible metodología a seguir para realizar el ensamble de un Transcriptoma de novo.

2.10. Ensamble *de novo* con el programa Trinity

En este proyecto para el ensamble de las secuencias a partir de los datos de RNA-seq sin un genoma de referencia, se utilizó el algoritmo Trinity. Es uno de los más utilizados

debido a su fácil manejo y a los excelentes resultados de ensamblaje. Se basa en realizar gráficos de Bruijn (es un gráfico dirigido), que está formado por un conjunto de vértices conectados por bordes, donde los bordes tienen una dirección asociada a ellos, este conjunto representa superposiciones entre secuencias de símbolos (Bruijn et al. 1946). A grandes rasgos, este pipe line fragmenta las lecturas en k-mers o pequeñas secuencias, los cuales, a partir de un k-mer primordial comienza el ensamble de pequeñas secuencias llamadas contigs hasta que la secuencia ya no puede ser secuenciada (Strickler et al. 2012). Hölzer y Marz (2019), compararon varios programas ensambladores en cuanto a la velocidad, calidad, manejo, entre otros puntos, y consideran al algoritmo Trinity como una de las mejores opciones para realizar un ensamble *de novo*. De igual forma Rodríguez y Shishkova (2018) mencionan las cualidades de Trinity, dejándolo ver como uno de los recursos computacionales más destacados.

El programa ensamblador Trinity basa su principio en la función de 3 software, Inchworm, Chrysalis y Butterfly:

a) Inchworm: ensambla todos los contigs de manera eficiente, primeramente, construye un catálogo de todos los k-mer posibles a partir de las lecturas obtenidas del secuenciador. Posteriormente elimina los posibles errores que puedan existir, seguido de esto se selecciona el k-mer que más se repite en el catálogo para comenzar el ensamblaje del contig sin tomar en cuenta los singleton (1 sola copia) y a los k-mer de baja complejidad.

Después se amplía el k-mer inicial hacia cada dirección posible para así encontrar el k-mer más amplio (con un $k-1$ superpuesto con el contig terminal), una vez que este k-mer ha sido utilizado para la extensión queda eliminado del diccionario. Por lo que extiende esta secuencia en todas las direcciones posibles hasta que no se pueda ampliar más. Al final de esto, comienza la ampliación con el siguiente k-mer más copioso, repitiéndose así hasta que sean nulos los k-mer en el diccionario. Todos estos contigs preparan una representación completa del gráfico de Bruijn basado en la superposición de secuencias, con cada k-mer único en el conjunto, y las subsecuencias $k - 1$ que definen implícitamente los bordes en el gráfico (Grabherr et al. 2011, Rodriguez et al. 2018, GHölzer et al. 2019).

b) Chrysalis: superpone los contigs generados por Inchworm en conjuntos que están conectados y de este modo construye gráficos de Bruijn para cada uno de los componentes (cada uno de ellos pertenece a una colección de contigs, algunas se derivan de formas de

empalme alternativas o de genes parálogos), este módulo primeramente agrupa los contigs en componentes de tal forma que cada uno de ellos estén conectados entre sí, después va a construir un gráfico de Buirjn por cada componente ($k-1$ para los nodos y yk para identificar a los bordes que conectan los nodos) y asigna cada borde del gráfico de Buirjn con los k -mer en la lectura original, para así, fijar una lectura a cada componente con el cual empata la mayor cantidad de k -mer (Grabherr et al. 2011, Rodriguez et al. 2018, GHölzer et al. 2019).

c) Butterfly: Construye transcripciones lineales para cada gráfico de Buirjn (generados con las lecturas originales con los extremos emparejados), primero simplifica los gráficos de Buirjn fusionando nodos consecutivos, de modo que formaran secuencias más largas, para después llevar a cabo el procedimiento llamado puntuación de ruta plausible, en donde este módulo identifica lecturas reales y pares de lectura (Grabherr et al. 2011, Rodriguez et al. 2018, GHölzer et al. 2019).

2.11. Anotación funcional del Transcriptoma *de novo* de *M. bombycina* con el programa Trinotate

Trinotate es un compendio de herramientas para realizar la anotación funcional de transcriptomas ensamblados *de novo*, de organismos modelo pero que no tienen un genoma de referencia o bien de organismos no modelo. Trinotate utiliza varias fuentes para la anotación funcional, incluida la búsqueda de homología con datos de secuencia ya establecidos como es el BLAST o SwissProt, así como permitir la identificación del dominio de proteínas con herramientas como HMMER y PFAM, ubicación del péptido señal de proteínas y predicción del dominio transmembrana con los programas signalP o tmHMM, y una de las herramientas más relevantes es GO ya que permite determinar los procesos, funciones y localizaciones que están sobre-/sub-representadas en un grupo de genes. La metodología más simple consiste en el procedimiento conocido como análisis de enriquecimiento, calculado como el porcentaje de genes de una categoría GO determinada que esta sobre-/sub-representada en un conjunto de genes, en relación con un conjunto de genes de referencia, desarrollando así vocabularios biológicos, aplicables a todas las especies, con el propósito de anotar los productos génicos de forma consistente en las diferentes bases de datos (Barrell et al., 2009). Todos los datos de anotación funcional que resulten del análisis del transcriptoma se integran en una base de datos SQLite que permite la búsqueda rápida y eficiente de términos con cualidades específicas relacionadas.

Un ejemplo de lo anterior es el trabajo de Sayadi (2016), en el que realizó el transcriptoma *de novo* del escarabajo de la semilla *Callosobruchus maculatus* en tres de sus estadios (larva, pupa y adulto). En donde utilizaron para la anotación de cada uno de ellos la herramienta bioinformática Blast2GO y el pipeline Trinotate, en donde obtuvieron en total, 33.216 genes putativos usando Blastx contra la base de datos Nr y 13.382 asignados a 34.100 términos de Ontología Genética (GO). De igual forma clasificaron 5.475 genes putativos en grupos ortólogos (COG) y se predijeron 116 mapas de rutas metabólicas en función de la anotación, mostrando así que el análisis transcriptómico es un recurso invaluable para mostrar un panorama amplio sobre cómo se está comportando la especie en tres de las etapas de su ciclo de vida.

2.12. Genes involucrados en algunas rutas reguladoras y funcionales en la especie de *M. bombycina* bajo condiciones *in vitro*

La respuesta de las plantas a diferentes tipos de estrés o de estímulos generalmente están relacionados con el aumento o la disminución de la expresión de genes específicos, dependiendo de la naturaleza, duración y severidad del estrés (Bartles et al. 2003).

Como se había mencionado anteriormente el primer paso a la respuesta de cualquier estímulo percibido por las plantas inicia con la transducción de señales, específicamente en el sensor del estrés, ya que, puede detectar cambios ambientales y transmitir, específicamente, la señal inicial del estrés a los organelos celulares específicos para traducir este estímulo, sin embargo, son muchos los componentes que se han identificado, por lo que se desconoce cómo es que interactúan todas las moléculas entre sí dentro de la red de señalización, a partir de que se percibe el estímulo comienzan a generarse moléculas de señalización como es el Ca^{2+} , inositoltrifosfato y especies reactivas de oxígeno (ERO), seguido de esto los segundos mensajeros se activan, desencadenando una cascada de señales que fosforilan los factores de transcripción, haciendo que estos regulen la expresión de un grupo de genes involucrados en la aclimatación a cualquier estímulo o estrés (Hirayama et al. 2010).

Por lo que en el transcriptoma de *M. bombycina* se podría encontrar genes pertenecientes a la familia de las proteínas LEA (Late Embriogenesis Abundant Proteins), ya que su expresión está relacionada con la protección de la integridad celular y al mantenimiento de la homeostasis iónica. Hernández-Camacho y colaboradores (2016), identificaron el gen

TESIS TESIS TESIS TESIS TESIS

DH1N que codifica para una proteína *LEA* de tipo 2 en *M. bombycina* bajo distintos tratamientos de estrés, observando que la actividad de este aumentaba conforme aumentaban las condiciones desfavorables. Sin embargo, la familia de estas proteínas está conformada de aproximadamente por 7 tipos, por lo que su identificación en esta especie sería de gran relevancia para enriquecer la información ya existente.

Otro tipo de genes que se espera encontrar son los que codifican para las enzimas clave en la biosíntesis de osmolitos como son la prolina, las poliaminas y los compuestos cuaternarios de amonio como es la glicina betaína, sacarosa, los polioles, los azúcares, alcoholes y otros oligosacáridos, así como otros aminoácidos durante condiciones de estrés ya que son presentan una función relevante en el ajuste osmótico de la célula, protegiéndolas de las especies reactivas de oxígeno (Rezaei et al. 2012)

También se podrían encontrar moléculas antioxidantes y enzimas que interactúan con ERO, entre los antioxidantes más importantes que regulan la homeostasis de ERO son el ascorbato (ASC) y el glutatión (GSH), además de taninos, flavonoides, α -tocoferol, carotenoides y precursores de la lignina y las enzimas que interactúan con ERO las más destacadas son la superoxido dismutasa (SOD) siendo esta la única enzima en las plantas que dismuta el O_2 las catalasas (CAT), varios tipos de peroxidasas, entre otros (Martí et al. 2009, Gill et al. 2010).

De igual manera, bajo condiciones de estrés abiótico se podrían identificar genes responsables de la detoxificación celular como son los genes que codifican a enzimas glioxalasa, así como también, genes que regulan la actividad de expresión de grupos de genes que son clave para la protección celular como son los factores de transcripción, de igual manera, grupos de genes que están involucrados en la deshidratación y rehidratación de la planta.

2.12.1. Glioxalzas

El sistema enzimático conformado por las glioxalzas es una vía ubicua a la cual se le considera esencial para la homeostasis celular, este sistema se compone de un grupo de enzimas que cataliza la conversión del metilglioxal a D-lactato (Kaur et al, 2016). Las dos enzimas más caracterizadas de este sistema son dos metaloenzimas, la glioxalasa I (GLYI) (lactoilglutación liasa 4.4.1.5) y la glioxalasa II (GLYII) (hidroxiacilglutación hidrolasa 3.1.2.6) (Ghosh et al. 2015). En conjunto, estas dos enzimas catalizan la isomerización del α -cetoaldehído metilglioxal (MG) a D-lactato utilizando glutatión reducido (GSH) como

cofactor, los sistemas vivos también poseen una actividad de glioxalasa III (GLYIII) que convierte directamente el MG a D -lactato sin necesidad de reducir el glutatión en un único paso sin la intervención de cofactor e independiente de iones metálicos para su actividad (Figura 5) (Ghosh et al. 2016).

Las células de las plantas al estar bajo condiciones de estrés aumentan el metabolismo primario, por lo que se traduce en la regulación de las enzimas involucradas en la glucólisis y en el ciclo de los ácidos tricarboxílicos (Umeda et al. 1994), dando como resultado el aumento de velocidad de la glucólisis, creando la pérdida de regulación de esta vía. Por lo que se produce un aumento del flujo de las triosas fosfatos al igual que los niveles de MG, Siendo así que la vía de las glioxalasas sería crucial para sobrellevar las condiciones de estrés, ya que contribuiría a la reducción de los niveles de MG (Yadav et al. 2005).

Otra de las propiedades que se le atribuyen a la vía de las glioxalasas además de la detoxificación del MG, es la regeneración de GSH ya que es un mecanismo adicional para explicar el rol de este sistema en plantas durante la respuesta a estrés (Mustafiz et al. 2011). El MG puede reaccionar de forma no enzimática con el glutatión reducido, formando el hemitioacetal (HTA), de esta manera, el GSH es titulado del medio celular e inutilizado.

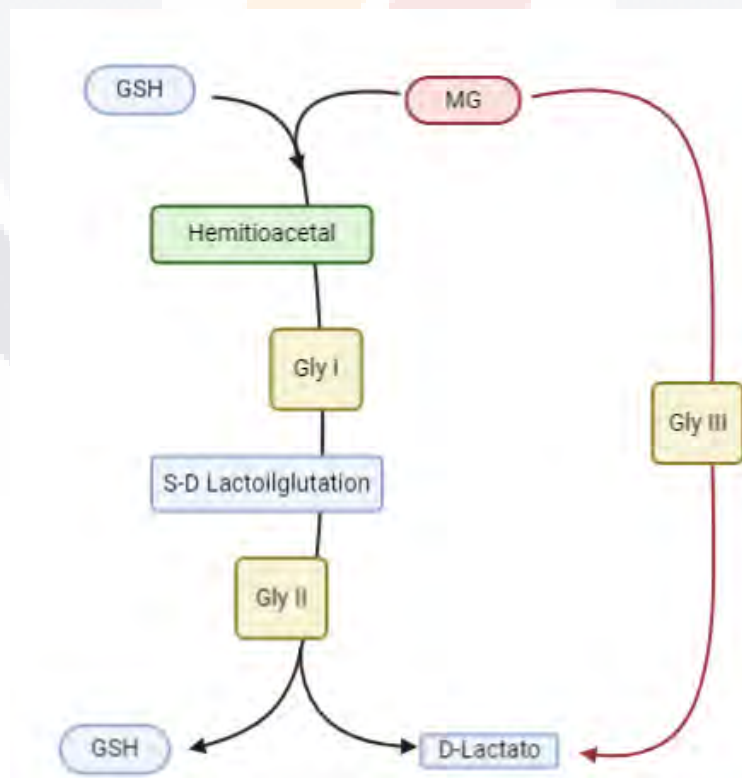


Figura 5. Vía de acción de la familia de las Glioxalosas. La vía de las glioxalosas se divide en dos etapas que involucra a las proteínas glyoxalasa I (GLYI) y glyoxalasa II (GLYII), la acción de estas en conjunto cataliza la conversión de metilglioxal (MG) en D-lactato utilizando glutatión (GSH) como cofactor (Creado con BioRender.com).

Por lo tanto, altos niveles de metilglioxal que se acumulan durante el estrés, además de ser tóxicos, podrían disminuir la concentración celular del glutatión reducido, en contraste con este mecanismo, la GLXIII es independiente de GSH, por lo cual la regeneración de dicho compuesto no sería su función más importante (Ghosh et al. 2016).

Como se había mencionado con anterioridad las enzimas GlyI y GlyII son metaloenzimas dentro de las cuales, GLYI requiere de iones metálicos divalentes, principalmente Ni^{2+} o Zn^{2+} para su activación, esto de acuerdo con Jain et al 2016, sin embargo, anteriormente se creía que el requerimiento de iones metálicos de la enzima GLYI estaba vinculado a su origen, siendo que GLYI de origen procariótico dependía de Ni^{2+} y el GLYI de origen eucariótico dependían de Zn^{2+} (Sukdeo et al. 2007). Jain, M. y colaboradores (2016) mencionaron que la clasificación anterior no es válida como se muestra en la Figura 6.

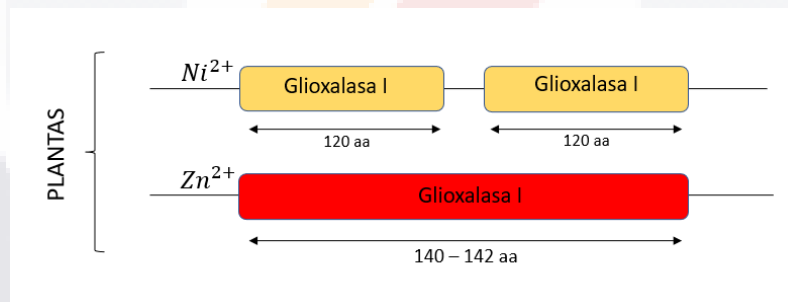


Figura 6. Arquitectura de dominios de las Glioxalosas tipo I en plantas. Tipos de dominios que pueden presentar GlyI dependiendo de la metaloenzima que tengan presente (Creado con BioRender.com)

La enzima GLYII también requiere de iones divalentes para su actividad. Las proteínas GLYII de procariotas poseen un centro de hierro y zinc (Schilling et al. 2003). De igual forma se ha encontrado que la enzima GLYII del arroz (OsGLYII-2) (Ghosh et al.2014) y

Arabidopsis (AtGLYII-2) (Limphong et al. 2010) poseen un centro binuclear Zn / Fe. Kwon y colaboradores (2013) mencionan que la actividad de la enzima GLYIII muestra una naturaleza independiente de iones metálicos como lo son el Cu^{2+} , Fe^{2+} y Zn^{2+} , de hecho, Subedi y colaboradores en el 2011 hicieron un estudio de la actividad de GLYIII de la proteína Hsp31 de *E. coli* y se percataron que los iones metálicos tienen efectos inhibitorios sobre esta, los autores propusieron que los efectos de los iones metálicos en la actividad de GLYIII podrían deberse a un cambio en el estado de oxidación de la enzima.

Las plantas poseen 3 tipos de dominios para GlyI, el cual va a depender de la metaloenzima a la cual depende su activación. La enzima GlyI que es dependiente de la metaloenzima Ni^{2+} tiene un dominio monomérico, análisis bioquímicos de las propiedades de OsGLY1-11.2 sugieren que este dominio tiene 2 sitios activos, de los cuales solo uno es funcional (Mustafiz et al. 2014). El segundo tipo de dominio de la enzima Gly1 es el que depende de la metaloenzima Zn^{2+} , es homodimérico y ya ha sido reportado en AtGLY1-2 en *Arabidopsis* y BjGLY1 de *B. juncea* (Deswal et al. 1998). El tercer tipo de dominio presente el Gly1 es una mezcla de los dos dominios anteriores, es decir, estas proteínas poseen un solo dominio como el dependiente de Zn^{2+} , sin embargo, el tamaño de este dominio está entre 115-125 aminoácidos el cual es característico del dominio dependiente de Ni^{2+} , a pesar de que ya está identificada aún no se conoce su actividad.

La secuencia de aminoácidos de las proteínas de GlyII de diferentes especies de plantas, revelan que se compone de un dominio metalo- β -lactamasa, y otro dominio que corresponde a un hidroxiglutation hidrolasa C-terminal (HAGH-C), el tamaño de este depende de la especie, sin embargo, tienen la misma arquitectura como se muestra en la Figura 7 (Gosh et al. 2016).

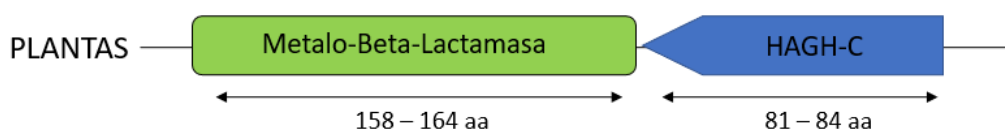


Figura 7. Estructura del dominio de Glioxalasa tipo II. Dominios presentes en la estructura génica de las glioxalosas tipo II, en donde se observa un dominio Metalo-beta-lactamasa (verde) y un dominio HAGH-C (Azul) (Creado con BioRender.com).

Estudios en la proteína DJ.1 en arroz y en *Arabidopsis* revelan que esta proteína está formada de dos dominios en tándem DJ-1/Pfp1, estos podrían haberse creado debido a eventos de duplicación de genes como se muestra en la Figura 8 (Gosh et al. 2016).

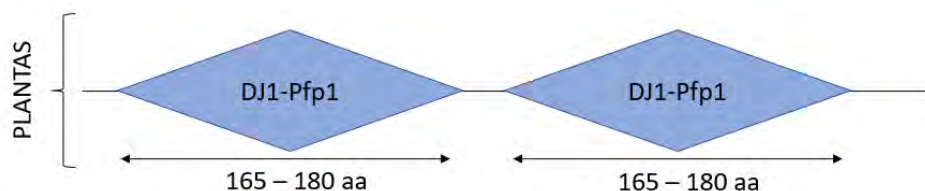


Figura 8. Arquitectura de dominios presentes en las Glioxalosas tipo III. En donde se observan los dominios DJ1/Pfp1 en tándem característicos de esta proteína (Creado con BioRender.com).

En los organismos vegetales, el papel que juega la vía de las glioxalosas se ha relacionado a varias funciones celulares. Sin embargo, los mecanismos de tolerancia al estrés se considera su papel más destacado (Kaur et al. 2014), ya que se sabe que los niveles de MG en las plantas aumentan en condiciones de estrés abiótico, y la vía de la glioxalosas, al desintoxicar la MG, proporciona tolerancia al estrés en las plantas (Yadav et al. 2005). Varios investigadores han demostrado que las plantas que sobre expresan los genes de la vía de las glioxalosas les confieren una mayor tolerancia contra la sequía, salinidad, presencia de metales pesados, etc., esto con base en plantas transformadas con ambos genes GLYI y GLYII, ya que funcionan mejor que las plantas transformadas con GLYI o GLYII (Singla-Pareek et al. 2003). Ghosh y colaboradores en 2014 trabajaron con plantas de arroz, los genes de glioxalosas caracterizados de estos organismos OsGLYI-11.2 y OsGLYII-2 también demostraron desempeñar un papel favorable en la tolerancia al estrés.

Lin y colaboradores (2010) trabajaron con hojas de tabaco que sobre expresan el gen TaGLYI del trigo y mostraron una mayor actividad de este gen por lo que llegaron a la conclusión de que le confiere tolerancia al estrés por $ZnCl_2$ en comparación con las hojas control. Por otra parte, en estudios realizados en *Arabidopsis*, en los cuales la disminución de la actividad o bien la pérdida del gen AtGLYI-2, que codifica una enzima GLYI estimulada bajo el estrés con Zn^{2+} , le confiere sensibilidad a MG y NaCl a las plantas, utilizando dosis muy pequeñas de MG le causan un retraso grave al crecimiento y desarrollo de *Arabidopsis*

TESIS TESIS TESIS TESIS TESIS

(Kaur et al. 2015). Yang, G. y colaboradores (2016) identificaron la proteína GLYI en la especie *B. napus* a la que denominaron BnGLYI3, demostraron que esta participa en la termotolerancia de las semillas, ya que sus niveles de proteínas aumentaron significativamente en respuesta al estrés por calor y frío en la levadura.

2.12.1.1. Importancia y biosíntesis de metilglioaxal

En las plantas el metilglioaxal (MG,) CH_3COCHO se produce de forma inevitable a partir de varios intermediarios metabólicos involucrados en los principales procesos celulares, como son la fotosíntesis, la glucólisis, el metabolismo de las proteínas y la peroxidación de lípidos, especialmente en condiciones estresantes (Thornalley et al. 2003). La ruta principal de producción de MG implica la descomposición no enzimática de los isómeros triosa fosfato dihidroxiacetona fosfato (DHAP) y gliceraldehído-3-fosfato (GAP) que se producen principalmente a partir de la vía de la glucólisis en condiciones tanto normales como de estrés. Sin embargo, además de las fuentes glucolíticas, la autooxidación de cuerpos cetónicos y azúcares, la reacción de Maillard entre azúcares reductores y aminoácidos, y la peroxidación de lípidos también contribuyen a la formación no enzimática de MG (Marasinghe et al. 2005). Recientemente, Kaur y colaboradores (2014) comentan que los niveles de MG aumentaban en las células vegetales debido al estrés, la generación de ROS incrementaba directamente debido a la presencia de MG o indirectamente debido a la formación de AGEs (Advanced Glycation End-Products). Yadav y colaboradores (2005) argumentan que, bajo condiciones de un metabolismo normal, la planta mantiene niveles bajos de MG (30 – 75 μM).

Sin embargo, cuando en la planta se presenta un estrés abiótico y su metabolismo se compromete, se incrementa la concentración de MG. Así mismo lo mencionan Nahar y colaboradores (2015), quienes realizaron experimentos con *Mung bean* (*Vigna radiata L.*), en los cuales indujeron estrés por sequía y observaron que el nivel de MG aumentaba en relación con la producción de ROS. El incremento de MG incrementó del mismo modo en especies como calabaza (*Cucurbita máxima L*) y tabaco (*Nicotina tabacum L.*), bajo estrés

por salinidad, entre otros experimentos del mismo tipo, por lo que esto indica que las plantas bajo estrés abiótico comúnmente incrementan los niveles de MG.

El MG es un α , β - dicarbonilo, el cual puede tener dos funciones dentro de la célula, como agente de glicación y como xenotóxico. Se compone de dos grupos funcionales: un grupo aldehído y un grupo cetónico (Leoncini, 1979). El grupo dicarbonilo que forma parte de la molécula del MG funciona como un potente agente de glicación, ya que modifica los grupos amino de las proteínas, ADN o ARN formando proteínas y nucleótidos AGEs (Rabbani et al., 2014), estos AGEs son modificaciones moleculares disfuncionales, los cuales disocian las estructuras de las proteínas y su actividad biológica, así como también distorsiona la integridad genómica (Thornalley et al. 2003).

Los sitios específicos en los que se puede producir el MG son el citosol, cloroplasto y mitocondria. Sin embargo, el sitio en el que se produce el MG depende del tipo de célula, tejido u órgano, además del estado fisiológico y ambiente en el que se encuentre la planta (Kaur et al. 2015).

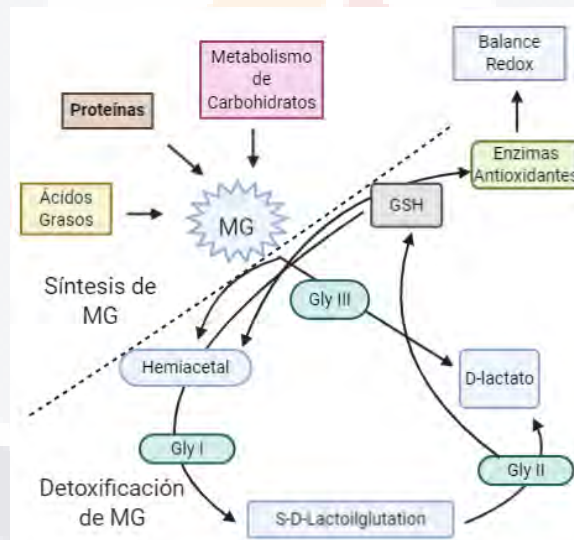


Figura 9. Formación del metilglioxal a partir de varias rutas metabólicas. El MG se produce principalmente como producto del metabolismo de los carbohidratos, proteínas y los lípidos y se degrada para formar D-lactato mediante la acción de las enzimas Gly I y GlyII, utilizando como cofactor el GSH. Sin embargo, GlyIII, no utilizando glutatión (GSH) como cofactor (Creado con BioRender.com)

2.12.2. Factores de Transcripción

Cuando la planta se encuentra bajo estrés, no depende de una sola vía reguladora de la homeostasis celular, si no que enciende varias rutas a la vez, ya que, gracias a los receptores de estas señales de estrés, se crea una necesidad fisiológica en la planta, generando una respuesta transcripcional desencadenando una respuesta que involucra la síntesis de una gran cantidad de proteínas quinasas (Sowdhamini, 2014). Dando como resultado el encendido y apagado de algunos receptores de estrés, la activación de algunos factores de transcripción, entre otros (Yamaguchi et al. 2006).

Los factores de transcripción (TF, por sus siglas en inglés) están involucrados y que tienen un papel crucial en varias rutas reguladoras de la homeostasis en las plantas. Estos tienen la función de regular la transducción de señales en la planta, modificando el nivel de expresión de genes que codifican a grupos de proteínas relacionadas con la respuesta al estrés abiótico (Malhotra et al. 2014).

Hay varios tipos de familias de TF que están presentes en el transcriptoma de *M. bombycina*, entre los que encontramos bZIP, MYC, MYB, NAC, WRKY y DREB a los que se les considera como los principales reguladores génicos de plantas en presencia de cualquier tipo de estrés abiótico (Malhotra et al. 2014).

2.12.2.1. Factor de Transcripción DREB

Los genes conocidos como factores de transcripción de unión a elementos sensibles a la deshidratación (DREB), son conocidos como uno de los genes reguladores transcripcionales que son responsables a la tolerancia del estrés abiótico (Golldack et al. 2011). En su estructura génica se encuentra a la familia TF APETALA2/Factor de unión sensible a etileno (AP2/ERF), en el que se encuentra un único dominio AP2 el cual está involucrado en la codificación de proteínas involucradas principalmente en las respuestas ante la presencia de ABA (Ácido abscísico), en la respuesta en presencia de etileno, estrés biótico y abiótico (Agarwal et al. 2010).

El TF DREB se ha reportado en distintos trabajos en varias especies de plantas como *A. thaliana*, en el cual se han identificado varias isoformas del gen entre los que se encuentran DREB1, DREB1A, DREB2A (Rehman et al., 2015). En *Oryza sativa* se han realizado análisis de expresión en presencia de varios tipos de estrés en los genes OsDREB2s, OsDREB2A, OsDREB2B, OsDREB2C, OsDREB2E y OsABI4 (Matsukura et al. 2010).

2.12.2.2. Factor de Transcripción CDPK

El factor de transcripción de transcripción CDPK, es miembro de la familia de cinasas de serina/treonina dependientes de Calcio (CDPK), y de acuerdo con recientes evidencias, se sugiere que las CDPK están relacionadas a la respuesta a factores abióticos. De igual manera son reguladoras importantes de señalizaciones de rutas mediadas por hormonas. (Nie et al. 2015)

De acuerdo con el estudio de Saijo (2000), menciona que la sobreexpresión de una CDPK confiere tolerancia de temperaturas bajas y sequía en plantas de arroz. Además, de acuerdo con Nie y Wang (2015), se han encontrado evidencias de la relación de diferentes subfamilias de CDPK que, entre otras cosas, tienen funciones contra el estrés por sequía, la señalización de la luz y respuesta al estrés por temperaturas bajas, sin embargo, la más importante y la que interesa a nuestra investigación debido al enfoque a la respuesta a factores abióticos, es la encontrada que las CDPK1 y CDPK activadas constitutivamente, alteran la expresión de genes de respuesta a estrés.

2.12.3. Subfamilias de α -expansinas y β -expansinas

Las expansinas son proteínas no hidrolíticas que están relacionadas directamente con la pared celular al que se le atribuyen diversas funciones fisiológicas como es la maduración de frutos, el crecimiento y desarrollo, el aflojamiento y el agrandamiento celular de una manera dependiente del pH y están involucradas en el crecimiento de las plantas y las respuestas al estrés abiótico según estudios relevantes, además de que también tienen un papel relevante en algunos procesos biológicos en la planta como son el factor de virulencia, alérgeno del polen y potenciador de la actividad de celulasas en la planta (Cosgrove et al. 2017).

En la planta, las expansinas estaban involucradas en la relajación de la pared celular, que libera la tensión mecánica del polímero de la pared celular al romper el enlace no covalente que une el microfilamento y hace que algunos polisacáridos se deslicen bajo presión de turgencia, propiciando así aflojamiento de la pared y la extensión celulares. Posteriormente, las expansinas se encontraron ampliamente en tejidos de plantas en crecimiento y frutos maduros. También se encuentran proteínas relacionadas en organismos no vegetales como bacterias, hongos, moluscos y nematodos que pueden ayudar a digerir la biomasa vegetal (Kang-Ming et al. 2020).

Las expansinas durante la maduración de frutos ocurren algunos cambios fisiológicos, bioquímicos y estructurales, entre estos destacan la producción de azúcares solubles, síntesis de pigmentos y compuestos volátiles, e hidrólisis de polisacáridos de almacenamiento; en consecuencia, el fruto sufre un ablandamiento que determina la calidad y vida después de la cosecha, ya que la maduración del fruto requiere múltiples enzimas que actúen sinérgicamente al modificar la pared celular, actúa primero sobre la pared celular, promueve el ablandamiento y la extensión de la pared (Morales et al. 2013). Existen múltiples estudios de estas proteínas y su presencia en diversas especies vegetales. Además, ya se ha descrito su función biológica y por ende su importancia en los procesos llevados a cabo por las plantas. Como se menciona en un estudio realizado por Quiroz-Castañeda y Folch (2011), las expansinas son proteínas que están involucradas en la remodelación de la pared celular vegetal, ellas inducen la extensibilidad y el relajamiento de la tensión en las paredes celulares de las plantas. Estudios indican que las expansinas estarían participando en distintos eventos que van desde la germinación hasta la maduración de frutos, así como en la polinización y la respuesta de crecimiento en condiciones de inundación.

En cuanto a las características biológicas de las expansinas estas proteínas constan de dos dominios y están anteceditas por un péptido señal, esto con base en el análisis filogenético de sus secuencias, actualmente se agrupan cuatro familias de expansinas: α - expansinas (EXPA), β -expansinas (EXPB), proteínas tipo expansinas "A" (EXLA) y tipo "B" (EXLB) (Sampedro et al. 2005).

Una característica de las expansinas es que comparten solamente un 20-40 % de identidad entre las diferentes familias de estas proteínas. El dominio 1 de las expansinas mantiene un alto grado de conservación y además tiene homología distante con la familia 45 de las

glicosil-hidrolasas (GH45), familia que comprende principalmente endoglucanasas fúngicas. Según investigadores el motivo estructural clásico de las endoglucanasas de la familia 45 (un barril- β de 6 cadenas presentando un surco de unión al substrato en el dominio catalítico) debería estar presente en las expansinas (Tovar, 2017).

Recientemente, se han descrito proteínas denominadas expansinas que podrían jugar un papel importante en la hidrólisis de celulosa y materiales lignocelulósicos. La búsqueda de nuevas proteínas con características que mejoren procesos biotecnológicos como la producción de bioetanol de segunda generación es un área que continúa en crecimiento. (Tovar, 2017)

La identificación de estas proteínas en el transcriptoma *de novo* de *M. bombycina* es de gran relevancia, ya que no hay mucha información acerca de la función de éstas en cactáceas y cuál es el papel que tienen en funciones específicas, sin embargo, se puede inferir que las expansinas caracterizadas a lo largo de este estudio compartirán propiedades de esta familia, mismas que ya se han reportado en otras especies.

3. Justificación

Las cactáceas son plantas que se caracterizan por tener modificaciones en sus estructuras vegetativas como areolas, tallos suculentos y hojas reducidas a espinas, algunas de ellas presentan una gran diversidad de flores muy llamativas en tamaños, formas y colores, siendo un gran atractivo como plantas ornamentales y debido a su gran capacidad adaptativa a diversos ambientes, hay un gran interés por descifrar a nivel génico cómo es que mantienen su homeostasis. Un ejemplo de esto es que, en los últimos años, se han realizado varios proyectos en la cactácea *M. bombycina*, en el estado de Aguascalientes, colocándola como una posible especie que se podría utilizar como modelo de estudio en cactáceas. Sin embargo, no se conoce de forma integral la expresión génica coordinada que se está llevando a cabo en esta planta bajo condiciones *in vitro*, siendo así, una de las principales necesidades para la obtención de herramientas moleculares para su uso en futuras investigaciones,

Debido a lo anterior, en el presente trabajo se planteó la secuenciación del transcriptoma *de novo* de *M. bombycina* mediante el uso de la Secuenciación de Nueva Generación (SNG), para la identificación de los genes que se expresan bajo condiciones *in vitro* e identificar genes relacionados con algunas rutas reguladoras y funcionales y en un futuro sean genes potencialmente biotecnológicos para la modificación de plantas de interés agrícola confiriendo resistencia contra diferentes factores bióticos y abióticos, además de tener una evidencia relevante de que *M. bombycina* es ideal como modelo de estudio en cactáceas.

4. Hipótesis

A partir del Transcriptoma de *M. bombycina*, se podrán identificar transcritos relacionados con algunas rutas reguladoras y funcionales en la especie bajo condiciones *in vitro*, permitiendo entender los posibles procesos biológicos y moleculares que regulan esta característica, como son los genes Gly.



5. Objetivos

5.1. Objetivos generales

-Generar el transcriptoma *de novo* de *M. bombycina* bajo condiciones *in vitro* e identificar y análisis de expresión de los genes *Gly* bajo condiciones de estrés.

5.2. Objetivos específicos

-Generar el Transcriptoma *de novo* de la especie *M. bombycina* bajo condiciones *in vitro*.

-Anotar funcionalmente los transcritos de la especie *M. bombycina* bajo condiciones *in vitro*.

- Identificar la posible presencia de los genes *Gly I*, *Gly II* y *Gly III*, en el Transcriptoma de *M. bombycina*.

-Validar la expresión de algunos genes que estén involucrados en la regulación de la homeostasis de *M. bombycina* bajo tratamientos por desecación a distintos tiempos de exposición.

6. Materiales y métodos

6.1. Obtención del Material Vegetal

Las plántulas de *M. Bombycina* se obtuvieron del Banco de Germoplasma *in vitro* de la Unidad de Biotecnología Vegetal de la Universidad Autónoma de Aguascalientes. El material vegetal se propagó en el medio de cultivo Murashige and Skoog (MS) (Murashige y Skoog, 1962) preparado según las especificaciones señaladas en el anexo B.1; para lograr la generación de brotes al medio MS se le agregó 1 μ l de BA.

6.2. Extracción de RNA

El RNA total de *M. bombycina* se extrajo por triplicado con el reactivo TRIzol (Invitrogen). Las muestras colectadas se homogenizaron con ayuda de nitrógeno líquido y un mortero con pistilo. Una vez que las muestras alcanzaron la consistencia de un polvo muy fino, se colocó un volumen aproximado de 500 μ l de material vegetal pulverizado en un microtubo de 1.5 ml. Enseguida se agregaron 1000 μ l de TRIzol (Invitrogen). Las mezclas se homogenizaron rápidamente en un vortex y se dejaron reposar a temperatura ambiente durante 2 min. Al término, se agregaron 200 μ l de cloroformo a cada tubo y se mezclaron manualmente por agitación por 20s. Los tubos se incubaron a temperatura ambiente por 5 min. y se centrifugaron a 4°C en una microcentrífuga refrigerada a 16.2g (13,000 rpm) durante 15 min.

Después de la centrifugación, los tubos se colocaron en hielo y se removió cuidadosamente la fase superior (acuosa) de la mezcla, colocándola en un tubo limpio de 1.5 ml. Al sobrenadante se agregaron 1000 μ l de isopropanol, se mezcló manualmente por inversión y se incubó a temperatura ambiente por 30 min. Enseguida se centrifugó a 16.2 g (13,000 rpm) a 4°C, por 20 min. El sobrenadante se separó por decantación, y la pastilla de RNA formada en el fondo del tubo se enjuagó con 1000 μ l de solución de etanol al 80%, se centrifugó para eliminar el líquido de lavado y se dejó reposar. Cada pastilla se resuspendió en 30 μ l de agua DEPC, se cuantificó en nanodrop y se almacenó a -80°C.

6.3. RNA-seq

Las librerías de cDNA normalizadas fueron secuenciadas por el método de Síntesis en un equipo Illumina Next-Seq 500, en el Instituto de Biotecnología de la UNAM (Universidad Autónoma de México), en Cuernavaca, Morelos. Mediante esta técnica se generaron archivos tipo. fastq, los cuales contienen la información de pequeñas lecturas.

6.4. Análisis FastQC, Trimming y filtrado de las secuencias obtenidas

Se utilizó la herramienta FastQC, obtenida de [Babraham Institute, 2000] (<http://www.bioinformatics.babraham.ac.>), para evaluar la calidad de datos crudos de secuenciación de alto rendimiento. Al término del análisis con el software FastQC de las secuencias obtenidas del SNG, se realizó un Trimming y filtrado con cuatro herramientas bioinformáticas, uno de ellos es FastX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), el cual, brinda varias herramientas con las cuales a través de una línea de comandos podremos eliminar lecturas de baja calidad e incluso eliminar la contaminación de algún adaptador que no se haya eliminado en la secuenciación a través de la verificación de la calidad con FastQC.

También se utilizó el programa Trimmomatic <http://www.usadellab.org/cms/?page=trimmomatic>, este es una herramienta que realiza múltiples procesos, se basa en una línea de comandos los cuales pueden eliminar adaptadores y recortar secuencias Illumina (fastq),

Otro de los programas utilizados es Fastp <http://opengene.org/fastp/fastp>, es un compendio de herramientas que incluye las funciones de FASTQC, Cutadapt, Trimmomatic y AfterQC, a pesar de que realiza funciones de distinta naturaleza, la velocidad con la que procesa la información es mucho mayor a cualquiera de sus pares. Una de las ventajas de esta herramienta es el preprocesamiento de identificador molecular único (UMI), recorte de cola polyG por lectura y división de salida. *Fastp* también proporciona informes de control de calidad para datos pre y post filtrados dentro de una sola página HTML, por lo que se omite el realizar un análisis nuevo en FastQC (Shifu et al. 2018).

De igual forma se utilizó el programa CLC Genomics, que a diferencia de los programas anteriores, está basado en una interfaz gráfica (<https://digitalinsights.qiagen.com>), lo que hace muy sencillo su manejo, en el cual, se pueden realizar el Recorte de calidad basado

en puntajes de calidad, recorte de ambigüedad para recortar, por ejemplo, tramos de Ns, adaptador de recorte, Recorte de base para eliminar un número específico de bases en el extremo de 3 'o 5' de las lecturas y el recorte de longitud para eliminar lecturas más cortas o más largas que un umbral especificado.

6.5. Ensamble del Transcriptoma *de novo* con el programa Trinity

Para evaluar la calidad de datos crudos de secuenciación de alto rendimiento se utilizó la herramienta FastQC, (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>), después se llevó a cabo un Trimming y filtrado de las lecturas con el programa Fastp (<http://opengene.org/fastp/fastp>), seguido nuevamente de un análisis FastQC a las lecturas resultantes. La normalización digital y el ensamblaje del transcriptoma *de novo* se llevaron a cabo utilizando el software Trinity 2.0.5 con un tamaño k-mer predeterminado de 25 en el que se concatenaron los datos de RNA-Seq de todas las muestras para generar un único conjunto Trinity (Grabherr et al. 2011).

6.6. Anotación funcional del Transcriptoma *de novo* con Trinotate

Para anotar los genes putativos ensamblados con el programa Trinity se realizó un blastx en varias bases de datos, entre ellas, NCBI (no redundante) (Nr) (<ftp://ftp.ncbi.nih.gov/blast/db/>), la base de datos Swissprot-Uniprot, Enciclopedia de genes y genomas de Kyoto (KEGG), con un valor de corte de E establecido en 10^{-5} . Los marcos de lectura abierta de genes (ORF) se predijeron utilizando Transdecoder v.2.0.1 con un valor de corte E de 10^{-5} (<http://transdecoder.sourceforge.net/>). Para complementar la anotación funcional se utilizó el pipeline Trinotate (<https://trinotate.github.io/>), a través de varios softwares como son Hmmer v.3.1b1, Tmhmm v.2, SignalP v.4.1, GOseq y egglog v.3.0.

6.7. Genes relacionados con el estrés abiótico

6.7.1. Análisis bioinformático en Glioxalasa I, II y DJ-1

La identificación de las secuencias de genes putativos de GlyI, GlyII y GlyIII fueron obtenidas a partir de las anotaciones del transcriptoma de *M. bombycina* y se confirmaron

mediante blastx en NCBI (<https://www.ncbi.nlm.nih.gov/>). La identificación de los dominios se realizó en la Base de datos Pfam (<http://pfam.xfam.org/>) y Prosite (<https://prosite.expasy.org>).

La construcción de árboles filogenéticos sin raíces se realizó con otras secuencias de aminoácidos del dominio de lactoilglutación liasa N-terminal (PF00903) de proteínas similares a GLYI: *Vitis vinífera* (VvGLY), *Oryza sativa* (OsGLY), *Arabidopsis thaliana* (AtGLY), *Glicine Max* (GmGLY) y *Medicago truncatula* (MtGLY), *Brassica juncea* y del humano utilizando el programa MEGA versión 7 (Kumar et al., 2016) con el método Neighbor-Joining y 1000 réplicas de bootstrap (Kumar et al, 2016). Se realizó el mismo procedimiento para proteínas similares a GLYII. El árbol filogenético para las secuencias putativas con el dominio DJ-1 / Pfpl (PF01965) N-terminal y C-terminal de *M. bombycina*, se realizó en conjunto de las secuencias de *Oryza sativa* (OsGLY), *Arabidopsis thaliana* (AtGLY), junto con el dominio DJ-1 / Pfpl de humano, utilizando el mismo método. El alineamiento múltiple de las secuencias se llevó a cabo utilizando ClustalW (Larkin et al., 2007) y se visualizó con Jalview (Waterhouse et al. 2009).

La localización subcelular se predijo con el programa WolfPSORT [<https://www.genscript.com/wolf-psort.html>] (Horton et al., 2007). El análisis fisicoquímico de las secuencias putativas en aminoácidos de GlyI, GlyII y GlyIII se realizó con la herramienta ProtParam (<https://web.expasy.org/protparam>) (Gasteiger et al. 2005). El diseño de la arquitectura de los dominios se realizó con DOG 2.0 (<http://dog.biocuckoo.org>) (Jian, 2009).

La estructura secundaria se realizó con el programa PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred>) (Jones, 1999). La predicción de las estructuras tridimensionales y de la estructura terciaria de las secuencias putativas de aa se realizó con el software CPHmodels 3.2 Server (<http://www.cbs.dtu.dk/services/CPHmodels/>), mientras que la estructura 3D se realizó en UCSF Chimera versión 1.14 (Pettersen et al., 2004).

6.7.2. Análisis bioinformático en factores de transcripción DREB y CDPK

La identificación de las secuencias de genes putativos de los factores de transcripción DREB y CDPK fueron obtenidas a partir de las anotaciones del transcriptoma de *M. bombycina* y se confirmaron mediante blastx en NCBI (<https://www.ncbi.nlm.nih.gov/>).

La identificación de los dominios se realizó en la Base de datos Pfam (<http://pfam.xfam.org/>) y Prosite (<https://prosite.expasy.org>).

El alineamiento múltiple de las secuencias se llevó a cabo utilizando ClustalW (Larkin et al. 2007) y se visualizó con Jalview (Waterhouse et al. 2009).

La localización subcelular se predijo con el programa WolfPSORT [<https://www.genscript.com/wolf-psort.html>] (Horton et al., 2007). El análisis fisicoquímico de las secuencias putativas en aminoácidos de MbDREB y MbCDPK se realizó con la herramienta ProtParam (<https://web.expasy.org/protparam>) (Gasteiger et al. 2005).

6.7.3. Análisis Bioinformático de genes de α -expansinas y β -expansinas en *M. bombycina*

La identificación de las secuencias de genes putativos de las subfamilias de α -expansinas y β -expansinas fueron obtenidas a partir de las anotaciones del transcriptoma de *M. bombycina* y se confirmaron mediante blastx en NCBI (<https://www.ncbi.nlm.nih.gov/>). La identificación de los dominios se realizó en la Base de datos Pfam (<http://pfam.xfam.org/>) y Prosite (<https://prosite.expasy.org>).

El alineamiento múltiple de las secuencias se llevó a cabo utilizando ClustalW (Larkin et al., 2007) y se visualizó con Jalview (Waterhouse et al. 2009).

La localización subcelular se predijo con el programa WolfPSORT [<https://www.genscript.com/wolf-psort.html>] (Horton et al., 2007). El análisis fisicoquímico de las secuencias putativas en aminoácidos de MbEXPA y MbEXPB se realizó con la herramienta ProtParam (<https://web.expasy.org/protparam>) (Gasteiger et al. 2005).

6.8. Validación de la expresión de los genes relacionados con estrés osmótico de *M. bombycina* in vitro, basal y bajo estrés.

A partir de las secuencias obtenidas en el transcriptoma para las glioxalosas, se diseñaron oligonucleótidos para GLYI, GLYII y GLYIII (Tabla 1). El diseño de cada uno de los oligonucleótidos se llevó a cabo a partir del análisis bioinformático realizado a cada una de las isoformas de las secuencias que se les consideró completas y posiblemente activas. Para GlyI se diseñaron 4 pares de cebadores para algunas de las isoformas de las secuencias encontradas bajo distintos criterios (MbGLY1-1, MbGLY1-3, MbGLY1-4 y

MbGLY1-7), En el caso de las isoformas de las secuencias de GlyII se diseñaron 2 pares de cebadores (MbGLY2-1 y MbGLY2-3) y para el gen GlyIII sólo se diseñó 1 par de oligonucleótidos MbDJ-1, debido a que sólo se encontró una isoforma para ese gen. Se utilizará como gen de referencia el 25S de Tomate (Dennis et al. 2018).

Se utilizaron plántulas de *M. Bombycina* de 4 cm de alto y 1 año de haberse propagado en medio MS. Para cada uno de los tratamientos se utilizó material vegetal por triplicado, en el caso del control se utilizó la plántula *in vitro*, para el tratamiento 1 se usaron plántulas en invernadero, para el tratamiento 2 se sometieron plántulas *in vitro* a deshidratación por 6 h (Ds 6 h) y para el tratamiento 3 se usaron plántulas *in vitro* a las que se les sometió deshidratación por 24 h.

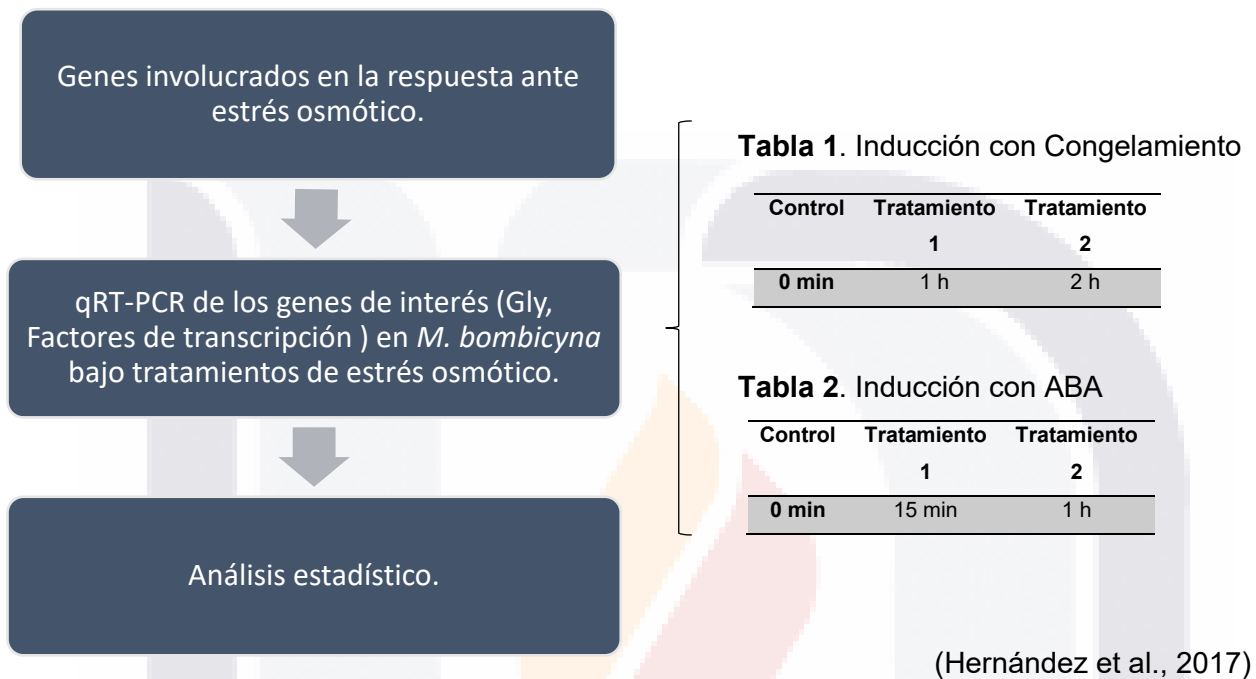
El RNA total de *M. bombycina* se extrajo por triplicado con el reactivo TRIzol (Invitrogen) siguiendo las recomendaciones del fabricante. La síntesis de cDNA se llevó a cabo con el kit Radiant 1- step siguiendo las recomendaciones del fabricante.

6.9. Análisis estadístico de la expresión de los genes relacionados con estrés abiótico de *M. bombycina in vitro*, basal y bajo estrés.

El análisis de expresión se realizó en el Centro de Investigaciones Biológicas del Noreste (CIBNOR) en La Paz, Baja California Sur en el laboratorio de Ecología Microbiana Molecular, en donde se llevó a cabo la amplificación de los transcritos con la siguiente mezcla de reacción: 5.0 µL de SYBER Green (Mezcla maestra de RT-PCR, Applied Biosystem, Carlsbad, CA, USA), 0.3 µM de cada oligonucleótido, 10 ng de cDNA, 3 µM de agua en una reacción con volumen total de 10µL. Las condiciones de amplificación fueron: 80 ciclos de 95°C (15 seg), (60°C (1min), 95°C (30seg) y 60°C (15 seg) en el equipo Applied Biosystems 7500 Fast Real-Time 2.0.

El análisis estadístico de la expresión de los genes gly I, gly II y gly III bajo condiciones *in vitro*, invernadero y por desecación 6 y 24 hrs.

6.10. Esquema general de la metodología de la etapa 2: Validación de algunos genes expresados como respuesta ante estrés osmótico.



7. Resultados

7.1. Propagación de *M. bombycina*

Se propagaron 256 explantes de *M. bombycina in vitro*, distribuidos en 32 frascos de cultivo, de los cuales el 15.62 % de los explantes de contaminaron (5 frascos). En la Figura 10, se observa algunas de las cactáceas de *M. bombycina* en medio MS.



Figura 10. Cultivo *in vitro* de la especie *M. bombycina*. Cultivo *in vitro* de *M. bombycina* en medio MS + 1 mg de BA.

7.2. Extracción de RNA de *M. bombycina*

Para verificar la integridad de extracción de RNA se llevó a cabo el análisis de las muestras de RNA total a una concentración de 4ug/20uL de *M. bombycina* por medio de un electroferograma RIN, En la Figura 11, en el inciso A) Electroforesis realizada a cada una de las muestras en los carriles 2-9 marcados por MB1-MB8, en donde se observa que cada una de las bandas que indican las subunidades ribosomales se encuentran integras, sin barridos. En el inciso B) se observa el electroferograma de cada una de las muestras, en donde se identifican los dos picos característicos de cada una de las subunidades ribosomales sin contaminantes, en el cual se corrobora que todas las muestras tienen un valor RIN > 4. Por lo que estos resultados indican que las muestras son viables, por lo que

se procedió a su secuenciación en el Instituto de Biotecnología de la UNAM, campus Cuernavaca.

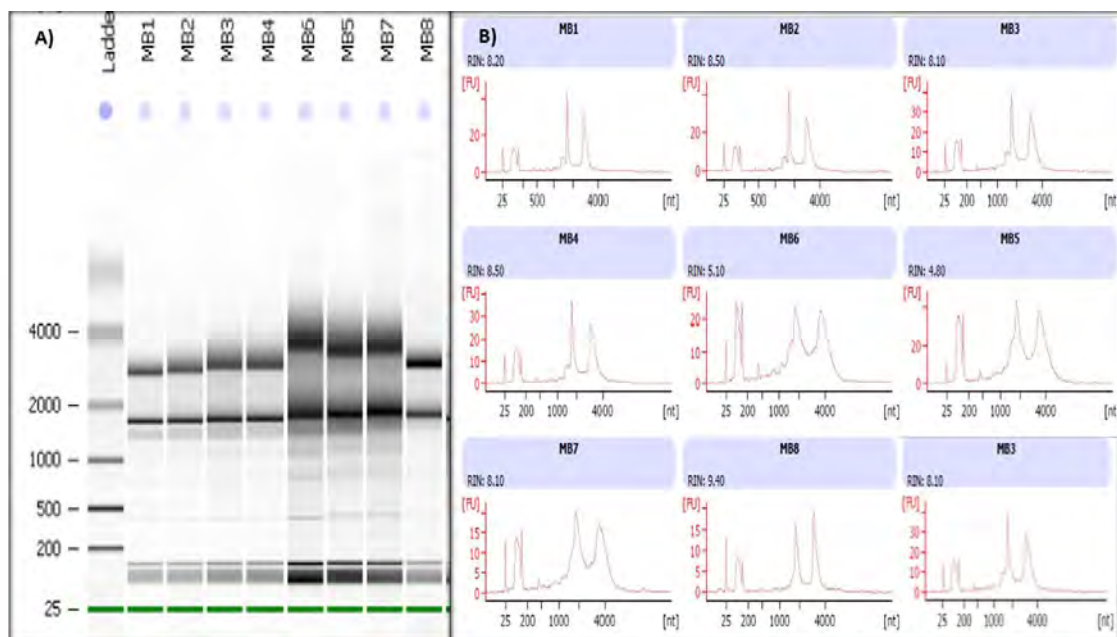


Figura 11. Electroforesis Y Cuantificación del RNA total de *M. bombycina*. A) Gel de Electroforesis: Verificación de la integridad de las subunidades ribosomales 60s y 25s del RNA total de *M. bombycina*: carriles: 1: Ladder, 2-9: MB1 – MB8 (1ul), B) Electroferograma RIN >4 en cada una de las muestras (Mb1 – MB8).

7.3. Información de la secuenciación

La secuenciación se realizó en el equipo Illumina Next Seq-500, con una configuración de 2 X 75 ciclos para cada una de las librerías de cDNA que se realizaron a partir de la muestra por triplicado de *M. bombycina* en condiciones *in vitro*, en la Tabla 3, se observa la cantidad de información obtenida de cada una de las librerías en GB (Giga Bytes), la cantidad de lecturas obtenidas, así como también el % de GC para cada una de ellas.

Tabla 3. Información obtenida de la secuenciación realizada en el equipo Next Seq 500, en el que se utilizó una configuración de 2X75 ciclos.

Muestra	Tamaño GB	Lecturas obtenidas	Longitud de las lecturas	%GC
---------	-----------	--------------------	--------------------------	-----

MB1R1	1.4	6,089,759	76	49
MB1R2	1.25	6,089,759	76	50
MB2R1	1.92	9,289,382	76	51
MB2R2	1.92	9,289,382	76	53
MB3R1	1.79	8,667,632	76	52
MB3R2	1.79	8,667,632	76	53

7.4. Análisis FastQC, Trimming y filtrado de las secuencias obtenidas

La evaluación de la calidad de datos crudos de secuenciación de alto rendimiento mediante la herramienta FastQC mostró un alto puntaje de calidad Phred en cada una de las secuencias obtenidas. Al inicio de las secuencias se observó un pequeño sesgo en las primeras 15 bases en las lecturas para cada una de las muestras, como se muestra en el ejemplo de la Figura 12. Para comprobar si este sesgo potencial se debió a que no se eliminaron las secuencias del adaptador del paso de secuenciación a bases de baja calidad, se hizo uso de varios softwares para limpiar las lecturas.

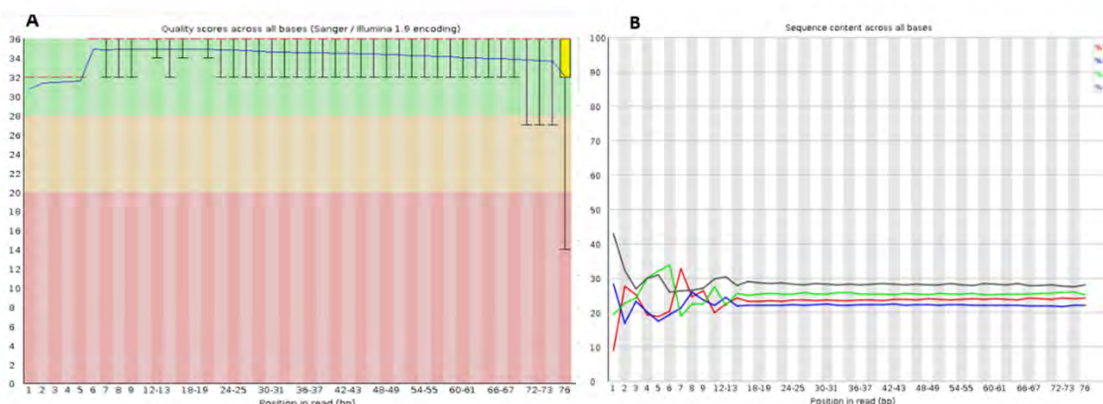


Figura 12. Análisis FastQC de lecturas obtenidas de la secuenciación de *M. bombycina*.

A) Gráfica que indica la puntuación de calidad de cada una de las lecturas. B) Electroferograma que indica la calidad de cada una de las bases por lectura.

Para esto, se realizó un Trimming y filtrado con tres herramientas bioinformáticas con el objetivo de generar un ensamble de mayor calidad (Tabla 4). Con la herramienta FastX se obtuvo 6,547,132 en promedio. Con el programa CLC Genomics (CLC) se obtuvieron en promedio un total de 6,441, 866 de lecturas.

Tabla 4. Tabla comparativa de los resultados obtenidos del recorte de secuencias con los programas FastX Toolkit, FastP y CLC Genomics.

	Fast X toolkit			FastP			CLC Genomics		
	Lecturas	Longitud	%GC	Lecturas	Longitud	%GC	Lecturas	Longitud	% G C
MB1R1	5440656	76	49	6015549	15-76	49	5334646	64	49
MB1R2	4884523	76	49	5583001	15-76	49	4790460	64	49
MB2R1	8020082	76	50	8885020	15-76	50	7844352	64	50
MB2R2	6961841	76	50	7924826	15-76	50	6824453	64	50
MB3R1	7436816	76	50	8019521	15-76	49	7436815	65	49
MB3R2	6538879	76	49	7373640	15-76	49	6420472	64	49

El programa Fastp a diferencia de las herramientas bioinformáticas para realizar recorte de secuencias, mostró varias ventajas, como son: el control de calidad, recorte de adaptadores, filtrado de calidad, poda de calidad por lectura y muchas otras operaciones con un solo escaneo de los datos FASTQ como se muestra en la Tabla 4, en donde se observa los porcentajes de secuencias retenidas con cada uno de los programas de recorte. Otro de los puntos que de manera muy particular de las secuencias de *M. bombycina* y que se pudo solucionar con gran eficacia es el recorte automático de PolyG, ya que, es un problema común observado en Illumina NextSeq y Serie NovaSeq. Esto debido a que se basan en la química de dos colores, utilizan dos luces diferentes rojo y verde para representar cuatro bases: una base con solo una señal de luz roja detectada se identifica cómo Citocina (C); una base con solo una señal de luz verde detectada se identifica a la Timina (T); una base con ambos la luz roja y verde detectada se identifica a la Adenina (A); y una base sin luz detecta a la Guanina (G).

Tabla 5. Tabla comparativa de los porcentajes retenidos de los resultados obtenidos del recorte de secuencias con los programas FastX Toolkit, FastP y CLC Genomics.

	Secuenciador	% obtenidoFast X toolkit	% obtenido FastP	% obtenido CLC Genomics
MB1	6,089,759	89%	98.7%	87.6%
MB2	9,289,382	86.33%	95.6%	84.4%
MB3	8,667,632	85.7%	92.52%	85.7%

7.5. Ensamble de novo de *M. bombycina* con el programa Trinity

El ensamble *de novo* de *M. bombycina* se llevó a cabo utilizando el ensamblador Trinity a los 3 Trimming realizados (Tabla 5). Las lecturas filtradas se concatenaron en extremos emparejados (en archivos únicos 'left.fq' y 'right.fq', respectivamente), por lo que se normalizaron digitalmente y se ensamblaron de novo con un k-mer predeterminado de 25. Al observar los resultados del ensamble con el programa ContingX, se descartaron los resultados obtenidos de las secuencias tratadas con los programas FastXtoolKit y CLC Genomics debido a que se perdía una gran cantidad de información (Anexo C), por lo que sólo se utilizaron los resultados obtenidos con el programa FastP.

Se obtuvieron un total de 78,412,335 bases ensambladas, traducidas a un total de 79,881 transcripciones generadas por Trinity (Tabla 6), de las cuales 47,406 corresponden a genes únicos, con base en las longitudes de los contigs del transcriptoma ensamblados. Se calculó la estadística de longitud Nx convencional, en la que se obtuvo que al menos el 50% de los nucleótidos del ensamble se encuentran en contigs que tienen al menos 1574 bases de longitud, con un promedio de 981 bases por contig.

En cuanto a las isoformas más largas o representativas por gen, se generaron un total de 37,675,612 bases ensambladas, de las cuales observamos que al menos el 50% de los nucleótidos se encuentran en contig que tienen al menos 1417 bases de longitud, con un promedio de 794 bases.

Tabla 6. Resumen estadístico de los resultados obtenidos del ensamble con el programa Trinity.

Counts of transcripts.	
Total trinity 'genes'	47406
Total trinity transcripts	79881
Percent GC	44.22
Stats based on ALL transcript contigs	
Contig N10	3539
Contig N20	2742
Contig N30	2227
Contig N40	1870
Contig N50	1,574
Median contig length	641
Average contig	981.61
Total assembled bases	78,412,335
Stats based on ONLY LONGEST ISOFORM per 'GENE'	
Contig N10	3337
Contig N20	2547
Contig N30	2077
Contig N40	1730
Contig N50	1417
Median contig length	418
Average contig	794.74
Total assembled bases	37,675,612

7.6. Anotación funcional del ensamble del transcriptoma con el programa Trinotate

Los marcos de lectura abiertos de genes (ORF) se predijeron utilizando Transdecoder v.2.0.1 (<http://transdecoder.sourceforge.net/>). A partir de esto todos los genes putativos ensamblados (genes putativos), se buscaron en varias bases de datos (la base de datos de proteínas NCBI (no redundante) (Nr) (<ftp://ftp.ncbi.nih.gov/blast/db/>), base de datos Swissprot-Uniprot, Enciclopedia de genes y genomas de Kyoto (KEGG), GO

(Ontología genética), EggNog e InterproScan) utilizando BlastX con un valor de corte de e^{-3} .

En la tabla 7 se observan los resultados obtenidos a partir de la base de datos Trinotate.xls. en la cual se obtuvieron un total de 84975 id para cada uno de los transcritos obtenidos, de los que se generaron 47,406 hits a los que se les asignó un id único para cada gen, de estos 44,629 se les pudo asignar un id para la proteína a la que codifican a las que a 32,539 de ellas se les pudo asignar una coordenada única. En cuanto a los resultados obtenidos por Blastx podemos ver que se generaron 43,640 hits con la base de datos con la que se comparó de los cuales 38,175 son hits para genes únicos, de igual manera se obtuvieron los resultados para el Blastp de los que se obtuvo un total de 34,192 hits de los cuales se asignaron 28,201 hits. Se obtuvieron 25,403 hits para rutas metabólicas únicas de Pfam para el total de genes asignados. También se asignaron 988 hits para proteínas ortólogas de las cuales 205 hits únicos.

Tabla 7. Resumen estadístico de las anotaciones del transcriptoma de *M. bombycina* con el programa Trinotate.

	Unique	Total
gene_id	47406	84975
transcript_id	79881	84975
prot_id	44629	44629
prot_coords	32539	44629
TmHMM	1729	44629
sprot_Top_BLASTX_hit	38175	43640
gene_ontology_BLASTX	9638	42361
Kegg	11468	39241
sprot_Top_BLASTP_hit	28201	34192
gene_ontology_BLASTP	8947	33218
Pfam	25403	30825
gene_ontology_Pfam	1547	19238
SignalP	1284	2115
eggnoG	205	988

Se utilizó TrinotateR para obtener los gráficos de la anotación, este es un paquete con funciones para manipular el producto final de Trinotate, de la cual se produjo la sig. Información:

-Clasificación de función eggNOG

Mediante la base de datos eggNOG se obtuvo la clasificación funcionalidad de las proteínas ortólogas. En este estudio, se asignaron 988 genes a 17 clasificaciones eggNOG, entre las cuales la clasificación de traducción, estructura ribosomal y biogénesis es la que obtuvo un número mayor de hits (Figura 13)

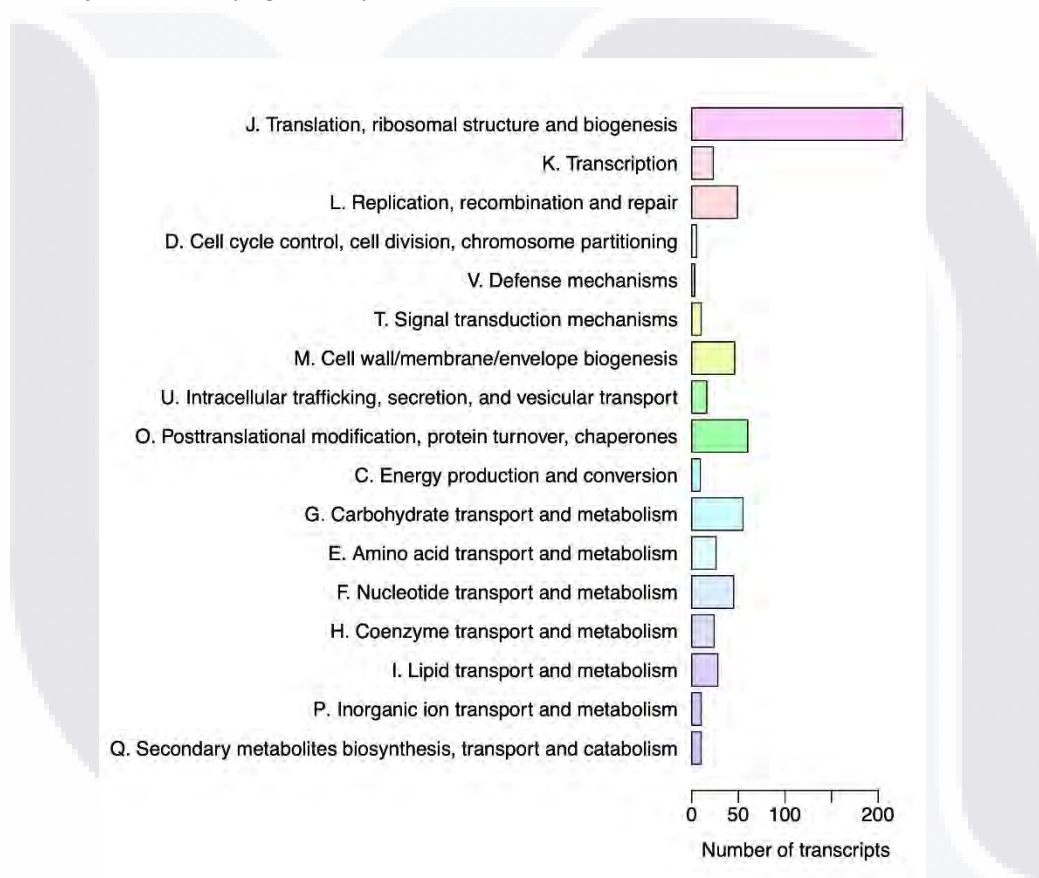
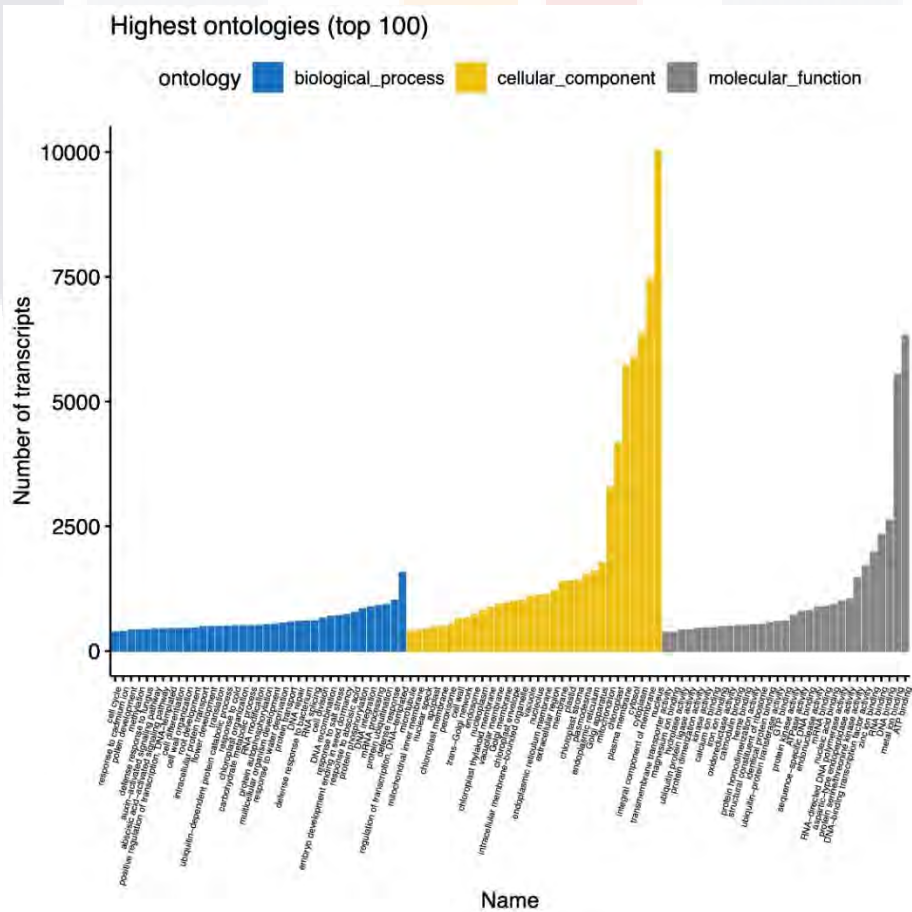


Figura 13. Clasificación de datos eggNOG. En el que se obtuvo la clasificación funcionalidad de las proteínas ortólogas del ensamble *de novo* de *M. bombycina*.

-Clasificación de términos Go

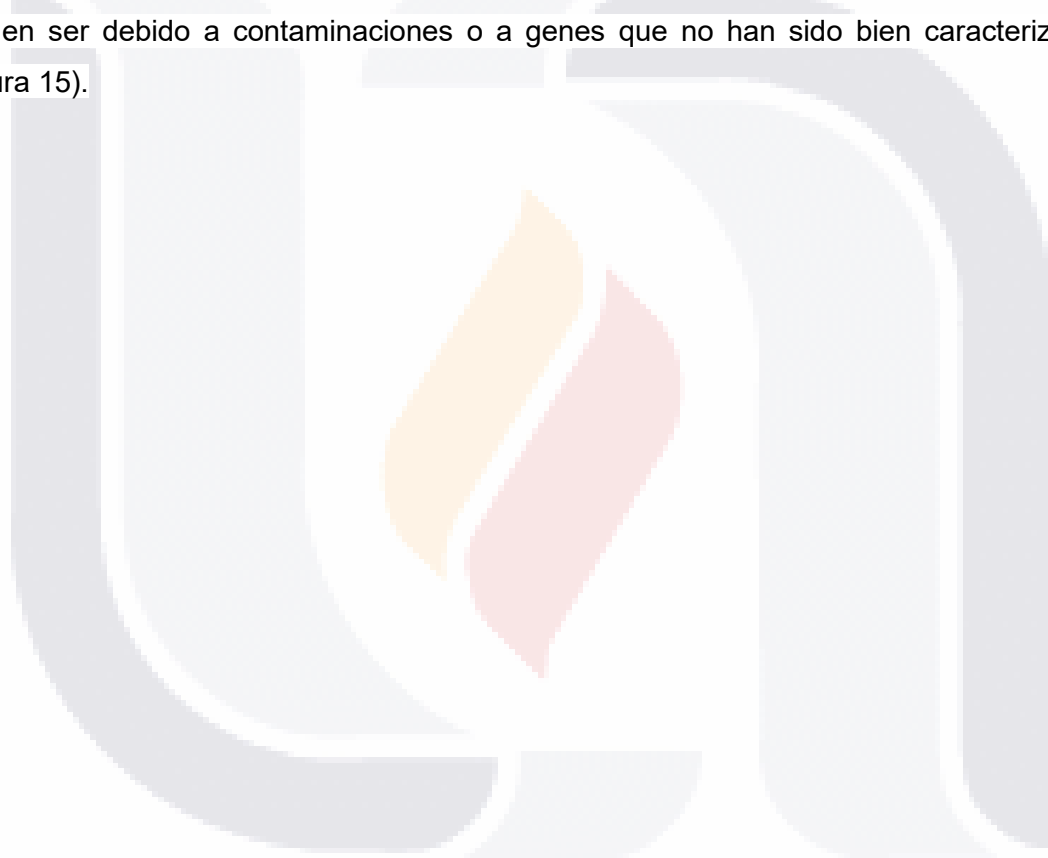
En total, 43,640 hits para términos GO se asignaron a 38175 hits para genes únicos (genes con Nr blast hits). La mayoría de los términos de GO se asignaron al componente celular, seguida de la función molecular y por último para proceso biológico. Los términos se derivaron de 99 grupos funcionales diferentes (Figura 14). Dentro de la función molecular, la mayor proporción se asignó a las categorías de unión a iones metálicos y unión a ATP; dentro de los procesos biológicos a la regulación de la transcripción del ADN y en cuanto el proceso celular a componentes de membrana y citosol. Los términos GO se infirieron utilizando la anotación de genes que tenía coincidencias de blast con proteínas pertenecientes a la base de datos UniprotKB. UniprotKB es una base de datos bien anotada, compuesta por secciones de registros anotadas manual y automáticamente. Estos



resultados implican que la anotación GO está impulsada principalmente por coincidencias de blast que se encuentran en las bases de datos de UniprotKB.

Figura 14. Clasificación de los términos GO del transcriptoma de *M bombycina*. Términos GO asignados al componente celular, función molecular y proceso biológico.

La distribución de hits en la base de datos Nr mostró la mayoría de ellos con *Arabidopsis thaliana*. El segundo taxón más afín a las anotaciones obtenidas fue *Oryza sativa* y de ahí con otras plantas que, aunque no son muy similares entre sí, se encontraban en el repositorio. Los hits restantes a los cuales no se les encontró un match en específico pueden ser debido a contaminaciones o a genes que no han sido bien caracterizados (Figura 15).



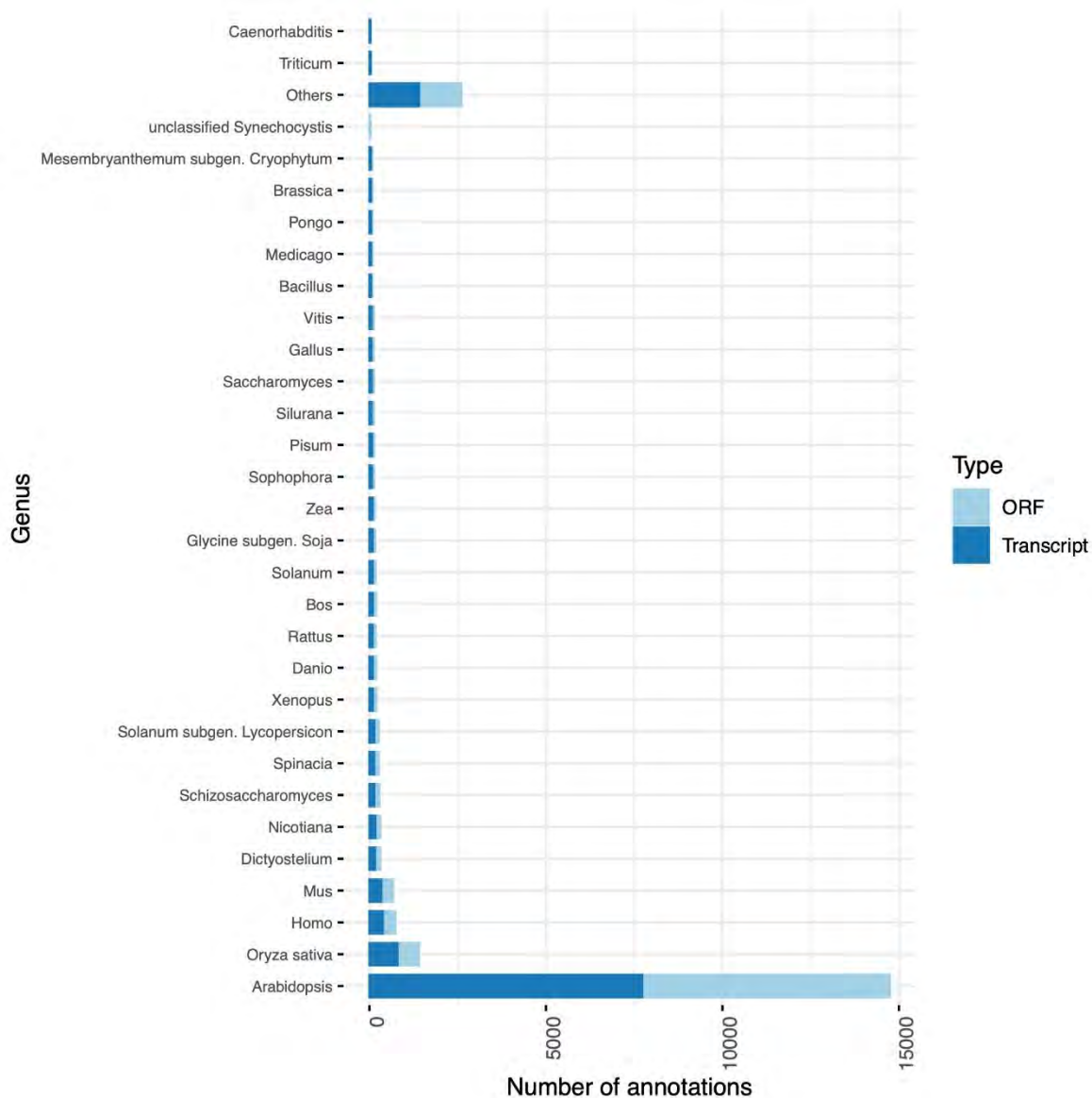


Figura 15. Distribución de hits en la base de datos Nr del transcriptoma de *M. bombycina*. Distribución de hits de las secuencias obtenidas del transcriptoma de *M. bombycina* con secuencias ya reportadas de otras especies ya secuenciadas.

7.7. Análisis de los genes Glioxalosas

A partir de los análisis obtenidos de la anotación funcional del transcriptoma *de novo* de *M. bombycina* se identificaron los genes pertenecientes a la familia de las Glioxalosas, los cuales están clasificados en el término GO de componente celular (Tabla 7).

Dando como resultado para GlyI doce isoformas, de las cuales sólo diez presentaban los dominios característicos, para GlyII se obtuvieron cinco isoformas, de las cuales cuatro presentaban los dominios característicos de las glioxalasa tipo II. Para GlyIII se identificaron tres isoformas, de las cuales sólo una secuencia presentaba los dominios característicos de las glioxalasa tipo III.

Cómo inicio del análisis, se identificaron a partir de las anotaciones del transcriptoma *de novo* *M. bombycina* la relación de genes tipo glioxalasa con las rutas metabólicas que asignaba la base de datos KEGG (Tabla 8). Como se observa observar el la Figura 16, en el que se identificaron cada una de las vías en el metabolismo del piruvato, específicamente en la descomposición del metilglioxal (MG) a D-Lactato.

Tabla 8. Tabla de información complementaria de los transcritos obtenidos de los genes de la familia de las Glioxalasa.

Gen id	Transcritos	ortólogo	KEEG
TRINITY_DN4660 0	TRINITY_DN4660_c0_g1_i1	LGUL_SCHPO^Q:9-138	spo:SPBC12C2.12c`KO:K01759
TRINITY_DN14362 62	TRINITY_DN14362_c0_g1_i1	LGUL_CICAR^Q:50-173	cam:101505694`KO:K01759
	TRINITY_DN14362_c0_g1_i2	LGUL_SOYBN^Q:4-186	gmx:547667`KO:K01759
	TRINITY_DN14362_c0_g1_i3	LGUL_SOYBN^Q:53-235	gmx:547667`KO:K01759
	TRINITY_DN14362_c0_g1_i4	LGUL_SOYBN^Q:53-235	gmx:547667`KO:K01759
	TRINITY_DN14362_c0_g1_i5	LGUL_SOYBN^Q:53-235	gmx:547667`KO:K01759
	TRINITY_DN14362_c0_g1_i6	LGUL_SOYBN^Q:4-186	gmx:547667`KO:K01759
	TRINITY_DN14362_c0_g1_i7	LGUL_SOYBN^Q:53-235	gmx:547667`KO:K01759
TRINITY_DN7846 6	TRINITY_DN7846_c0_g1_i1	LGUC_ARATH^Q:1-252	ath:AT1G67280`KO:K01759
TRINITY_DN15304 04	TRINITY_DN15304_c0_g1_i1	GLX1_ARATH^Q:109-975	ath:AT1G11840`KO:K01759
	TRINITY_DN15304_c0_g1_i2	GLX1_ARATH^Q:1-287	ath:AT1G11840`KO:K01759

	TRINITY_DN15304_c0_g2_i1	GLX1_ARATH^Q:8-289	ath:AT1G11840`KO:K01759
TRINITY_DN9859	TRINITY_DN9859_c0_g1_i2.p1	GLO2C_ARATH^Q:417-88	ath:AT2G31350`KO:K01069
TRINITY_DN14689	TRINITY_DN14689_c0_g1_i3.p1	GLO2N_ARATH^Q:144-1046	ath:AT2G31350`KO:K01069
TRINITY_DN15347	TRINITY_DN15347_c0_g1_i3.p1	GLO2C_ARATH^Q:983-210	ath:AT2G31350`KO:K01069
TRINITY_DN9859	TRINITY_DN9859_c0_g1_i1.p1	GLO2N_ARATH^Q:69-1004	ath:AT2G31350`KO:K01069
	TRINITY_DN9859_c0_g1_i2.p1	GLO2N_ARATH^Q:27-965	ath:AT2G31350`KO:K01069
TRINITY_DN10140	TRINITY_DN10140_c0_g1_i1.p1	DJ1D_ARATH^Q:1559-387	ath:AT3G02720`KO:K18881
TRINITY_DN13681	TRINITY_DN13681_c0_g1_i.p1	DJ1B_ARATH^Q:159-1406	ath:AT1G53280`KO:K03152
TRINITY_DN16556	TRINITY_DN16556_c2_g1_i1.p1	DJ1C_ARATH^Q:1695-541	ath:AT4G34020

En los recuadros rojos se identifican a los genes pertenecientes a la familia de las glioxalidas, con los números de identificación 4.4.1.5, 3.1.2.6 y 42.1.130, para GlyI, GlyII y GlyIII respectivamente, los cuales se obtuvieron de la anotación del transcriptoma de *M. bombycina* (Figura 16).

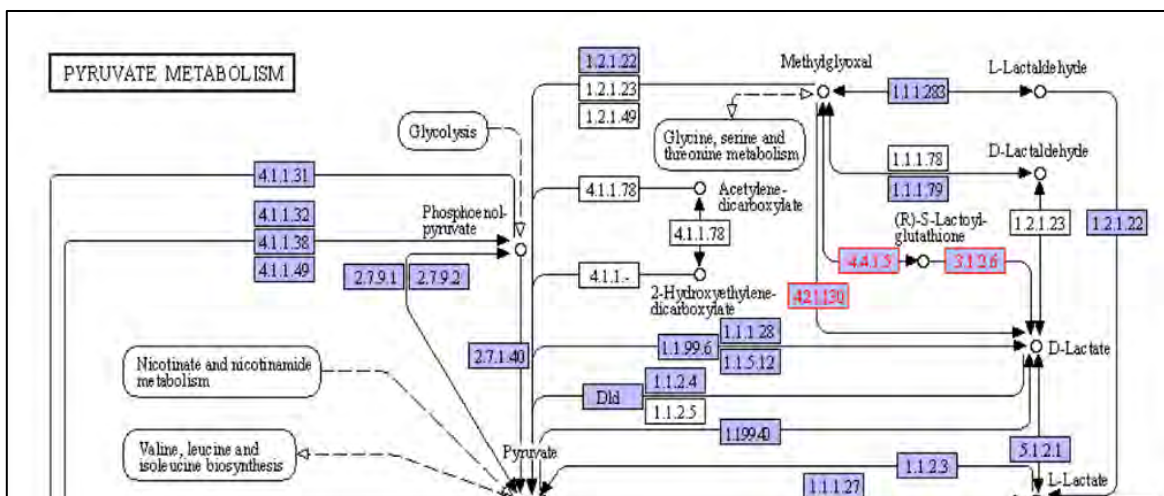


Figura 16. Análisis KEGG de los genes Glyoxalasa del transcriptoma *de novo* de *M. bombycina*. Identificación de los genes pertenecientes a la familia de las glioxalosas involucradas en la descomposición del metilglioxal en D-Lactato, los cuales están marcados en rojo (4.4.1.5, 3.1.2.6 y 4.2.1.30).

7.8. Caracterización y análisis bioinformático de proteínas putativas que corresponden a Glioxalosas de tipo I, II y III (GlyI, GlyII y GlyIII) en *M. bombycina*.

7.8.1. Análisis de la arquitectura de los dominios de las proteínas putativas de MbGLYI, MbGLYII y MbGLYIII

Todas las proteínas similares predichas para Glioxalosas putativas en *M. bombycina* se analizaron usando Pfam para identificar los dominios conservados, para el caso de MbGLYI se identificó el/los dominios (según sea el caso Zn^{2+} o Ni^{2+}) glioxalasa (PF00903), de igual forma para MbGLYII se identificaron secuencias parciales de GLYII en las cuales se identificó el dominio HAGH-C que es un sitio de unión al sustrato que generalmente se encuentra en el extremo C-terminal de las proteínas GLYII junto con el dominio catalítico metalo-beta-lactamasa y para el caso de MbGLYIII se identificaron dos dominios DJ-1 / Pfpl característicos de las glioxalosas de tipo III.

De las doce isoformas obtenidas para GlyI, únicamente diez tienen la presencia de por lo menos un dominio de glioxalasa conservado (PF00903), los análisis revelaron que todas las proteínas con características de Glioxalasa I (MbGlyI) tienen al menos un dominio GLYI, mientras que algunas de ellas tienen dos. Las proteínas MbGLYI_1, MbGLYI_2 y MbGLYI_3 cuentan con un solo dominio Glioxalasa (PF00903) de aproximadamente 147

aa. Siete de las secuencias contenían dos dominios GLYI consecutivos, de los cuales cada uno de ellos de una longitud de 124 aa aproximadamente (Tabla 7).

Las 10 secuencias MbGLYI con al menos un dominio Glioxalasa se alinearon con el programa MEGA7 (Clustal W), con las secuencias de *Arabidopsis thaliana*, *Glycine max*, *Vitis vinifera*, *Oriza sativa* y *Hommo sapiens*. En la Figura 17 se observa que las proteínas se dividen en dos categorías, aquellas que tiene las regiones específicas para la isoforma dependiente de Zn^{2+} . (MbGLYI_1, MbGLYI_2 y MbGLYI_3), que están enmarcadas en los recuadros rojos. Por otra parte, aquellas secuencias que tienen ausencia de estas secuencias (marcadas en recuadro rojo) se les considera que son las isoformas dependientes de Ni^{2+} (MbGLYI_4 - MbGLYI_10). Independientemente del metal al que se unen, cada una de las secuencias cuenta con los 4 sitios que caracterizan a las Glioxalastas (HQ/E/HQ/E), que son el sitio activo de la proteína, el sitio de unión al metal, el sitio de unión a Glutación (GSH) y la interfaz del dímero.

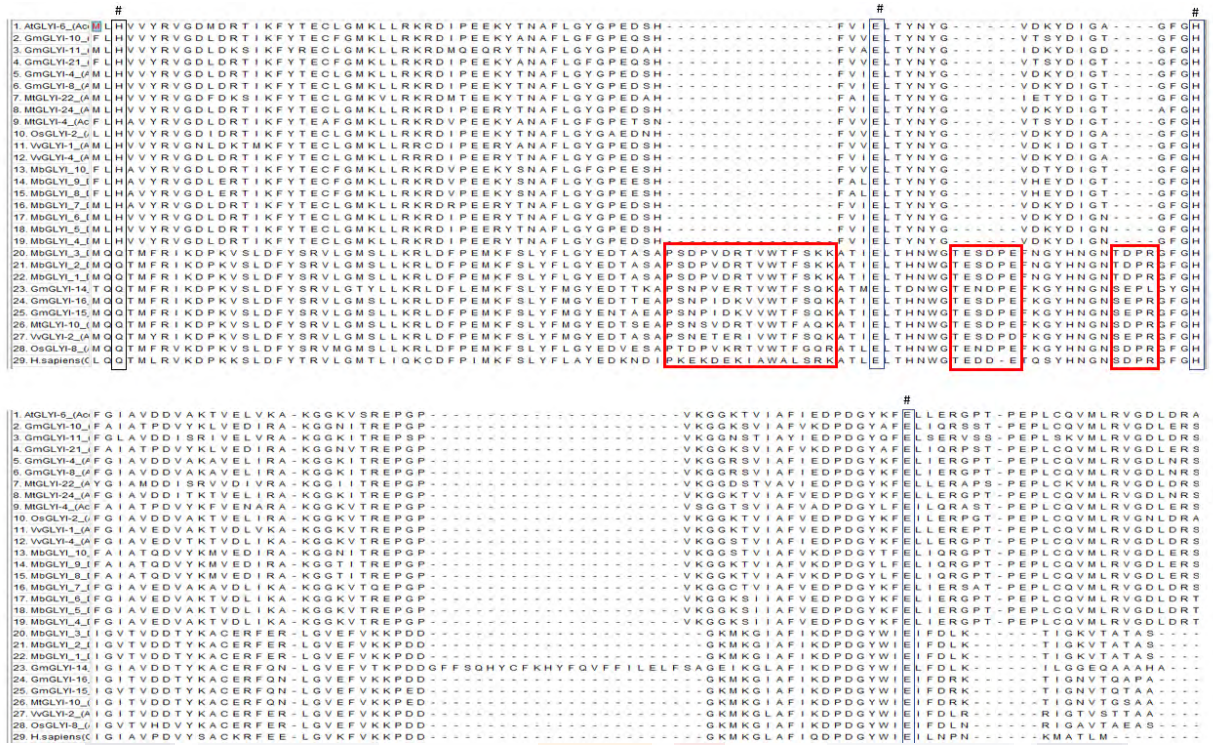


Figura 17. Alineamiento múltiple de secuencias de dominios GLYI. En los recuadros rojos están marcadas las regiones específicas en la secuencia que indican la presencia del ion Zn^{2+} . En los recuadros azules se encuentran marcados los recuadros que indican los motivos conservados en las secuencias de MbGLYI.

Las secuencias de dominios de las proteínas MbGLY1-1 MbGLY1-10, presentan los cuatro sitios de unión de metales conservados de proteínas GLYI los cuales se representan con un enmarcado negro y con el símbolo #, mientras que la región específica para la isoforma dependiente de Zn^{2+} se marcó con un recuadro rojo.

Otra de las maneras de confirmar la clasificación de las proteínas putativas de MbGLYI es por el tamaño de sus dominios, como podemos observar en la Tabla 9 las secuencias de GLYI dependientes de Zn^{2+} tienen un único dominio más largo (más de 140 aa) que los dependientes de Ni^{2+} , los cuales cuentan con 2 dominios de alrededor de 120 aa y tienen ausencia de las regiones características de la unión a Zn^{2+} en sus secuencias.

Tabla 9. Clasificación de las proteínas putativas de MbGLYI por el tamaño de sus dominios y presencia de motivos.

Proteína	Dominio (PF00903)			Sitio Activo	Sitio unión a metal	Sitio unión a GSH	Interfaz del dímero	ion metálico	Aminoácidos esenciales
	Inicio	Fin	Long.						
MbGLYI_1	77	224	147	+	+	+	+	Zn ²⁺	Presente
MbGLYI_2	77	224	147	+	+	+	+	Zn ²⁺	Presente
MbGLYI_3	28	175	147	+	+	+	+	Zn ²⁺	Presente
MbGLYI_4	101	225	124	+	+	+	+	Ni ²⁺	Presente
	231	355	124						
MbGLYI_5	1	124	123	+	+	+	+	Ni ²⁺	Presente
	130	254	124						
MbGLYI_6	1	124	123	+	+	+	+	Ni ²⁺	Presente
	130	154	124						
MbGLYI_7	94	218	124	+	+	+	+	Ni ²⁺	Presente
	224	348	124						
MbGLYI_8	23	147	124	+	+	+	+	Ni ²⁺	Presente
	153	281	128						
MbGLYI_9	23	147	124	+	+	+	+	Ni ²⁺	Presente
	153	281	128						
MbGLYI_10	23	147	124	+	+	+	+	Ni ²⁺	Presente
	153	281	128						

Para el caso de GlyII las proteínas putativas que se encontraron en el transcriptoma de *M. bombycina* únicamente eran secuencias parciales, sin embargo, contenían el dominio metalo-beta-lactamasa (PF00753). Se realizó un alineamiento múltiple (Figura 18), de las secuencias de interés junto con las secuencias de otras plantas, incluidas *A. thaliana*, *Vitis vinifera*, *G. max*, entre otras. Se identificaron los sitios de unión de metales conservados presente en la familia de las metalo-β-lactamasas, que requieren de Zn²⁺ (THXHXDH/H/D/H), un sitio activo (C/GHT) y siete sitios de unión de GSH conservados (C/K(R)/F(Y)/Y/N/R/K), a este alineamiento también se le incluyó la secuencia humana (*Homo sapiens*). Como se observa en la Tabla 10, las secuencias contienen un motivo de unión a metales altamente conservado (THXHXDH) que también está presente en la familia de las metalo-β-lactamasas, que requieren Zn (II)

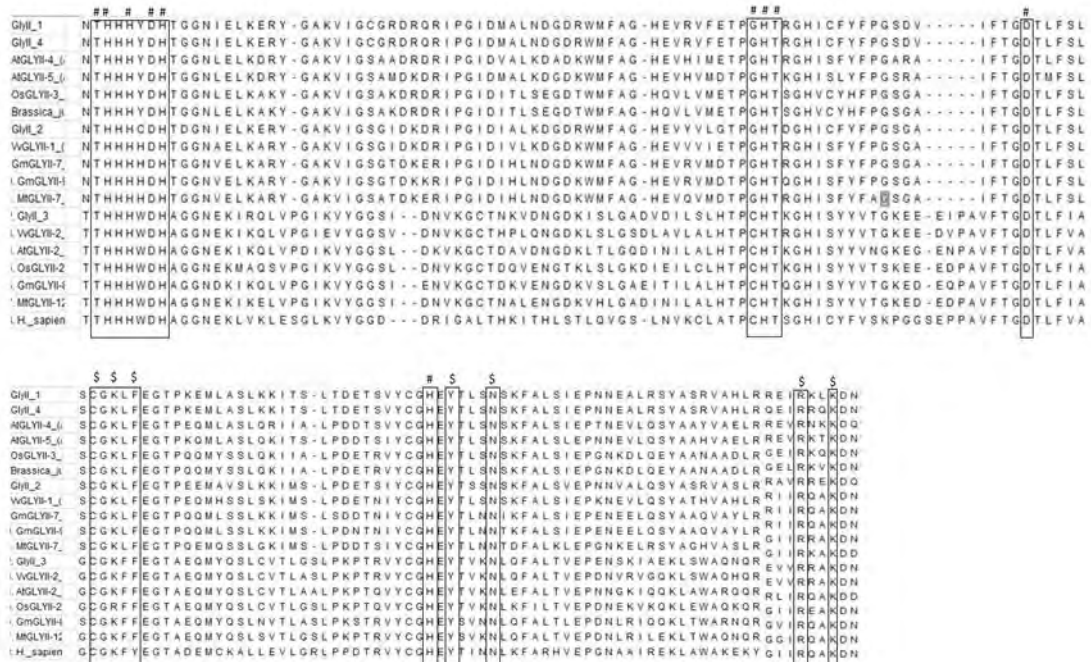


Figura 18. Alineamientos de secuencia de proteínas parciales GLYII. En los recuadros se muestran los dominios y motivos conservados característicos de las GLYII alineadas con secuencias de proteínas de longitud completa de otras plantas.

Tabla 10. Clasificación de las proteínas putativas de MbGLYII por el tamaño y características de sus dominios y motivos.

Proteína	Dominio			Dominio (PF16123)			Sitio Activo	Sitio unión a metal	Sitio unión a GSH
	Inicio	Fin	Long.	Inicio	Fin	Long.			
MbGLY II_1	91	249	197	250	334	80	+	+	+
MbGLY II_2	78	236	197	237	321	80	+	+	+
MbGLY II_3	106	263	197	264	348	80	+	+	+
MbGLY II_4	42	206	197	207	288	80	+	+	+

En el caso de la Glioxalasa de DJ-1 se identificaron tres secuencias de isoformas putativas para MbDJ-1, sin embargo, después de realizar los primeros análisis pudimos observar que

únicamente una de las secuencias contaba con los dos dominios característicos DJ-1 / Pfpl, por lo que sólo se utilizó esa secuencia para el alineamiento múltiple (Figura 19) e identificación de la triada catalítica de este dominio como podemos observar en la Tabla 11.

Tabla 11. Clasificación de las proteínas putativas de MbDJ-1 por el tamaño y características de sus dominios y motivos.

Proteína	Dominio (PF00903)			Triada catalítica del sitio Activo			Aminoácidos esenciales
	Inicio	Fin	Long.				
MbDJ-1	8	193	185	+	+	+	Presente
	201	386	185				

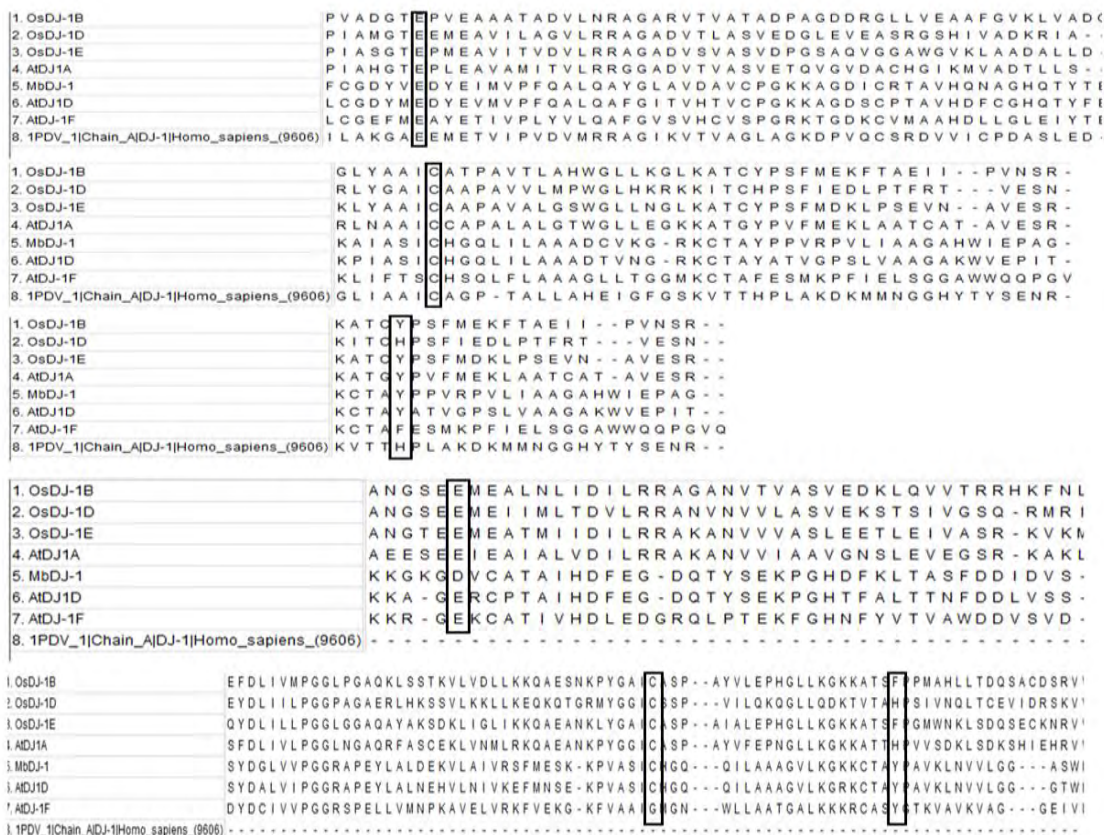


Figura 19. Alineamiento de secuencia de proteínas DJ-1. En los recuadros se indican los motivos característicos de las proteínas DJ-1, alineadas con secuencias de proteínas de longitud completa de otras plantas.

Para tener una idea más concreta de la estructura de cada una de las proteínas de interés, se diseñó la arquitectura presente en MbGLYI, MbGLYII y MbGLYIII de los dominios de cada una de las isoformas pertenecientes a GLYI, GLYII y GLYDJ-1. En la Figura 20 se observa a las proteínas pertenecientes a GLYI en color azul que las tres primeras secuencias tienen unión al ion Ni^{2+} , debido a que en su secuencia únicamente tienen presentes 1 dominio, mientras que, en las secuencias de la cuatro a la diez, contienen dos dominios característicos de las proteínas que tienen unión al ion Zn^{2+} . Para MbGLYII, cada una de las secuencias contienen los dos dominios característicos de estas proteínas, que son el metalo- β -lactamasa y el hidroxiaclilglutación hidrolasa. Mientras que en las secuencias de proteínas pertenecientes a MbDJ-1, la secuencia presenta dos dominios en tándem de DJ-1/Pfp1.

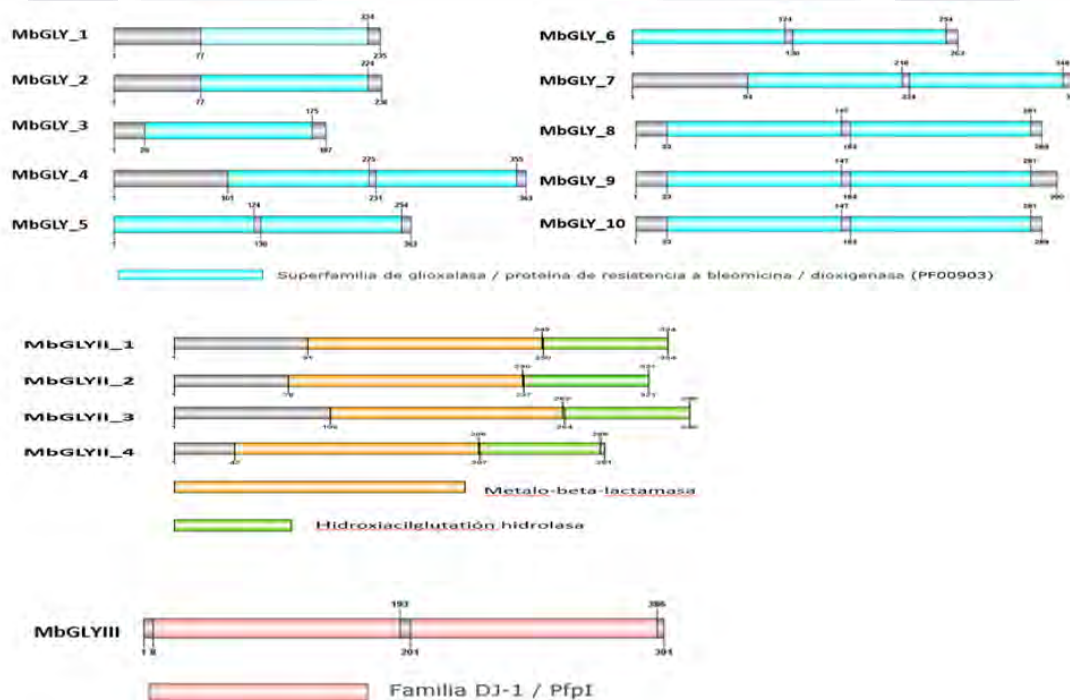


Figura 20. Arquitectura de los Dominios conservados de MbGLYI, MbGLYII y MbDJ-1 de *M. bombycina*.

7.8.2 Dendrograma de las proteínas putativas MbGLYI, MbGLYII y MbDJ1

El árbol muestra dos subfamilias, de acuerdo al tipo de metal al que se unen las proteínas (Zn^{2+} ó Ni^{2+}). muestra que MbGLYI-3, MbGLYI-4, MbGLYI-5 y MbGLYI-6 se posicionan en

7.8.3 Caracterización fisicoquímica de las proteínas putativas de Gioxalasa de tipo I, II y III en la especie *M. bombycina*

Los resultados obtenidos de la caracterización de las proteínas MbGLYI (Tabla 12), en la cual se puede observar la longitud del transcrito que oscila entre los 1023 pb y 2457 pb, la longitud del CDS que va de 708 pb a 1092 pb, el tamaño de las proteínas de 235 aa hasta los 300 aa, El Punto isoeléctrico y la localización celular.

Tabla12. Caracterización fisicoquímica y localización celular de las proteínas MbGLYI,

Nombre del gen	Transcritos	Longitud (pb)	CDS (pb)	Proteína			Localización celular
				Proteína putativa	kDa	PI	
MbGLYI-1	TRINITY_DN14362_c0_g1_i4	1161	708	235	26.35703	8.82	Chlo ^{ac} , nucl ^b
MbGLYI-2	TRINITY_DN14362_c0_g1_i7	1023	564	187	20.88262	5.43	Chlo
MbGLYI-3	TRINITY_DN14567_c0_g1_i2	1701	1092	363	40.21782	6.64	Chlo ^{ac} , mit ^b
MbGLYI-4	TRINITY_DN14567_c0_g1_i5	2457	792	263	29.53397	5.13	Nucl ^a , cyt ^b ,
MbGLYI-5	TRINITY_DN14567_c0_g2_i1	1400	1071	356	39.31778	5.80	Chlo ^{abc} ,
MbGLYI-6	TRINITY_DN15304_c0_g1_i1	1219	870	289	32.68124	5.32	Nucl ^a , Nucl ^c
MbGLYI-7	TRINITY_DN15304_c0_g1_i2	1024	903	300	33.65052	5.70	Cyto ^{ac}
MbGLYI-8	TRINITY_DN15304_c0_g2_i1	1386	870	289	32.43995	5.26	Cyto ^{ac}
MbGLYII_1	TRINITY_DN9859_c0_g1_i1	1112	NA	334	37.19777	8.72	Chlo ^a , mito ^c
MbGLYII_2	TRINITY_DN9859_c0_g1_i2	1073	NA	321	35.72691	8.45	Chlo ^{ab} , mito ^c
MbGLYII_3	TRINITY_DN14689_c0_g1_i3	1306	NA	348	38.42322	7.30	Chlo ^a , cyto ^b Nucl ^c
MbGLYII_4	TRINITY_DN15347_c0_g1_i3	1079	NA	290	32.34300	6.55	Cyto ^a , mit ^c
MbDJ-1	TRINITY_DN10140_c0_g1_i1	1617	1177	392	41974.14	5.57	Cyto ^{ac}

MbGLYII y MbDJ1.

Abreviaturas: CDS secuencia del DNA codificante, NA no aplica, kDa kilo Daltons, pI punto isoeléctrico, Chlo cloroplasto, Cyto citoplasma. La predicción de la localización fue en: a: Wolf Psort, b: Localizer y c: CELLO, el pI y el peso molecular en el programa ProtParam.

7.8.4 Modelado hipotéticas de las proteínas putativas de Gioxalosas

El modelado hipotético de las proteínas 3D de las secuencias pertenecientes a MbGLYI, MBGLYII y MbDJ-1, en el cual se puede observar en la Figura 22 las ocho secuencias

putativas para MbGLYI (MbGLYI-1 a MbGLYI-8) encontradas en el transcriptoma de *M. bombycina*, en las imágenes A, B y C, se identifican en color rojo la región de aminoácidos que le confiere a la proteína la capacidad de unión al ion Zn^{2+} . En las figuras D, E, F, G, H, I y J, la secuencia difiere bastante, ya que tienen la ausencia de las secuencias de unión a Zn^{2+} (zona roja), por lo tanto, se unen al ion Ni^{2+} , además de que están compuestos por dos subunidades que son el dominio PF00903. Para el caso de MbGlyII (MbGLYII_1 a MbGLYII_4), en las figuras K, L, M y N, se pueden observar los sitios de unión THHHYD de color naranja, GTH de color rojo y el sitio CGKLF de color amarillo característicos de esta proteína. En el inciso Ñ de la Figura 22 se muestra la proteína putativa de MbDJ-1, en donde se pueden observar las dos subunidades de los dominios en tándem DJ1-D/Pfpl, en donde cada uno de ellos se identifica la triada catalítica característica de esta proteína.

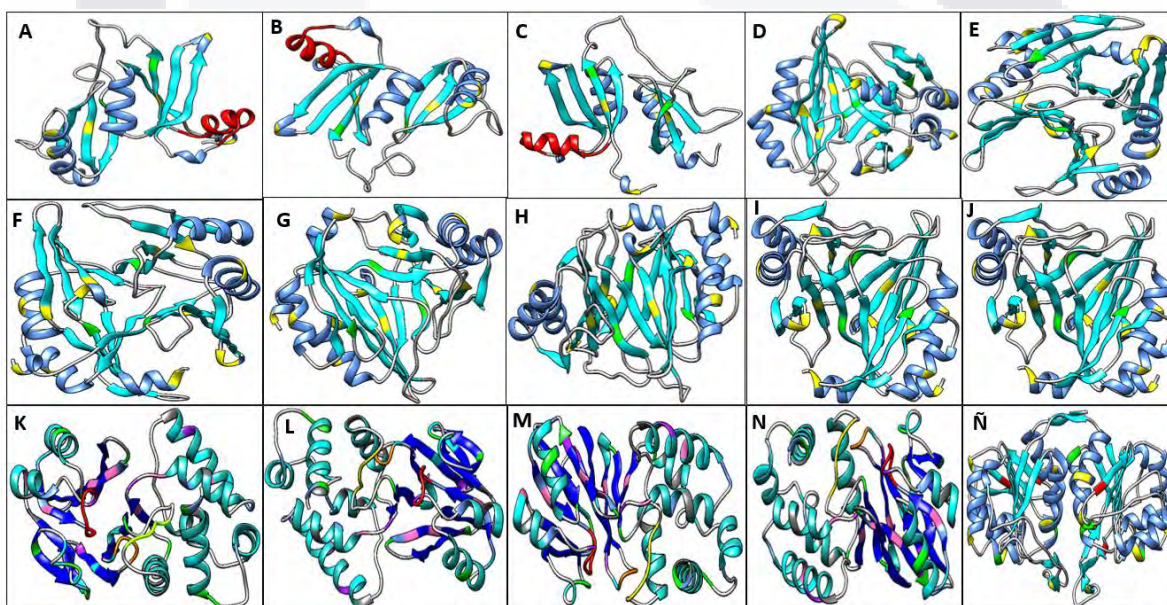


Figura 22. Estructuras 3D de las proteínas putativas de MbGLYI, MbGLYII y MbDJ-1 de *M. bombycina*. Generadas por modelado de la homología de las estructuras terciarias de las proteínas putativas de GLYI, GLYII y GLYIII de *A. thaliana* con el programa CPHmodels 3.2.

7.8.5 Análisis de expresión de los genes relacionados con estrés abiótico

La expresión del gen MbGLYI-1 en el tratamiento en el tratamiento Ds a 24 h se observa un aumento en la expresión en un 185% en comparación al control, mientras que la expresión de MbGLYI-3, MbGLYI-4 y MbGLYI-7 disminuyó en todos los tratamientos en comparación al control. En el caso de MbGLYII-1 se presentó una mayor expresión en los tratamientos ex

vitro y Ds 24 h en comparación al control. La expresión en MbGLYII-3 es significativamente más bajo em en todos los tratamientos con respecto a MbGLYII-1 y no hay una diferencia significativa con respecto al control. En el caso de MbDJ-1 la expresión se mantuvo para los tratamientos *ex vitro* y Ds 24 h respecto al control, sin embargo, en el tratamiento Ds 6 h hubo una disminución significativa de la expresión respecto al mismo.

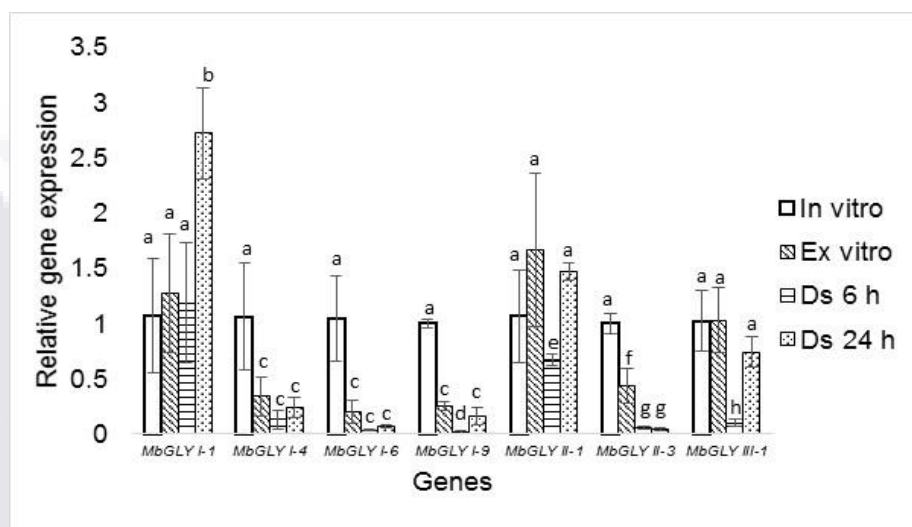


Figura 23. Análisis de expresión de los genes MbGLYI, MbGLYII y MbDJ-1. La expresión del gen MbGLYI-1 a las 24 hr se vió aumentada en un 185% en comparación al control. Mientras que las los genes MbGLYI-4, MbGLYI-6 y MbGLYI-9 no se presentó una diferencia de expresión entre tratamientos ni con el control. El gen MbGLYII-1 tuvo una expresión mayor en los tratamientos *ex vitro* y Ds 24h sin embargo el gen MbGLYII-3 tuvo una nula expresión en los tratamientos respecto al control. El gen MbGLYDJ1-1 no tiene una diferencia significativa en la expresión respecto al control.

7.9 Caracterización y Análisis Bioinformático de proteínas putativas que corresponden a Factores de Transcripción DREB en *M. bombycina*.

7.9.1 Análisis de la arquitectura de los dominios de las proteínas putativas de MbDREB1 y MbCDPK1

Todas las proteínas similares predichas para los factores de transcripción DREB y CDPK putativas en *M. bombycina* se analizaron usando Pfam para identificar los dominios conservados. Se identificaron los dominios YRG/RAHD característicos de las proteínas DREB como se muestra en la Figura 24.

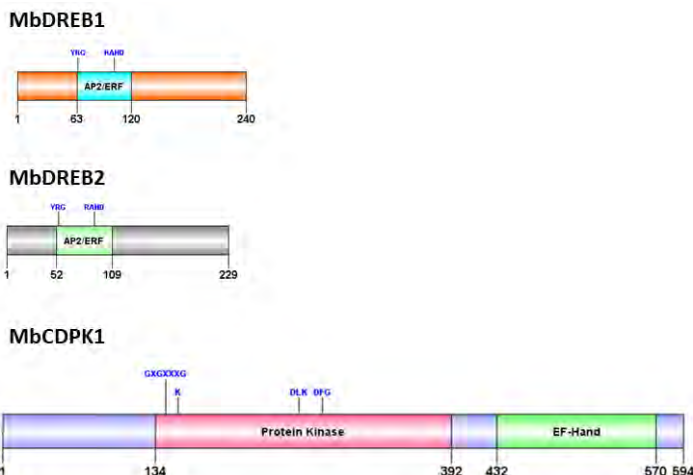


Figura 26. Arquitectura de Dominios conservados de MbDREB1, MbDREB2 y MbCDPK1.

7.9.2 Caracterización fisicoquímica de las proteínas putativas de MbDREB1 y MbDREB2

Se realizó la caracterización fisicoquímica de las proteínas MbDREB1, MbDREB2 y MbCDPK1. En la tabla 13 se puede observar la longitud del transcrito de 720, 687 y 1782 pb respectivamente, la longitud del CDS de 243, 232 y 597 pb respectivamente, el tamaño de las proteínas de 240 y 229 aa. Las 2 proteínas MbDREB1 y MbDREB2 se encuentran en el núcleo. La proteína que pertenece a MbCDPK1 se encuentra en cloroplasto.

Tabla 12. Caracterización fisicoquímica y localización celular de las proteínas MbDREB1 y MbDREB2.

Tabla 13. Lista de los genes putativos DREB y CDPK encontrados en el transcriptoma de Mb.

Nombre del gen	Transcritos	Longitud del transcrito (pb)	CDS (pb)	Proteína			Localización celular
				Longitud (aa)	kDa	pI	
MbDREB1	>TRINITY_DN13609_c0_g1_i1.p1	720	243	240	26315.18	5.39	Núcleo
MbDREB2	>TRINITY_DN13609_c0_g1_i2	687	232	229	24930.61	5.01	Núcleo
MbCDPK1	>TRINITY_DN11945_c0_g2_i1.p1	1782	597	594	65946.98	5.26	Cloroplasto

7.9.3 Modelado 3D hipotético de las proteínas putativas de factores de transcripción DREB y CDPK

El modelado de la homología de las estructuras terciarias de las proteínas putativas que corresponden a los factores de transcripción DREB y CDPK identificados en el transcriptoma de *M. bombycina*. Se obtuvieron 2 modelados 3D a partir del template de la estructura cristalina de AtERF96 (5WX9.1) de *A. thaliana* MbDREB1 (A) y MbDREB2 (B) (Figura 27). Para la proteína putativa de MbCDPK1 se obtuvo 1 modelado a partir del template 4yga.4.A de *A. thaliana* (C)



Figura 27. Estructuras 3D de las proteínas putativas de factores de transcripción DREB y CDPK. Donde: a) MbDREB1, B) MbDREB2 y C) MbCDPK1 de *M. bombycina*, generadas a partir de la homología de las estructuras terciarias de *A. thaliana* con el programa CPHmodels 3.

7.9.4 Análisis de la expresión de los genes de factores de transcripción DREB y CDPK bajo estrés por frío y ABA en *M. bombycina in vitro*.

El análisis de expresión del gen MbDREB1 tiene un aumento significativo en presencia de estrés por frío a 1 y 2 hr, ya que se observa un aumento del 180% y 420% respectivamente. Y este mismo gen en presencia de estrés por ABA a .15 y 1 hr aumenta 150% y 160% con respecto al control, teniendo en cuenta que se encuentra normalizado a cero (Figura 28).

En el caso del gen CDPK en presencia de estrés por frío a 1 y 2 hr, se observa un aumento en la expresión de 60% y 220% respectivamente, mientras que este mismo gen en presencia de ABA a .15 y 1 hr aumenta 170% y 230% respectivamente con respecto al control.

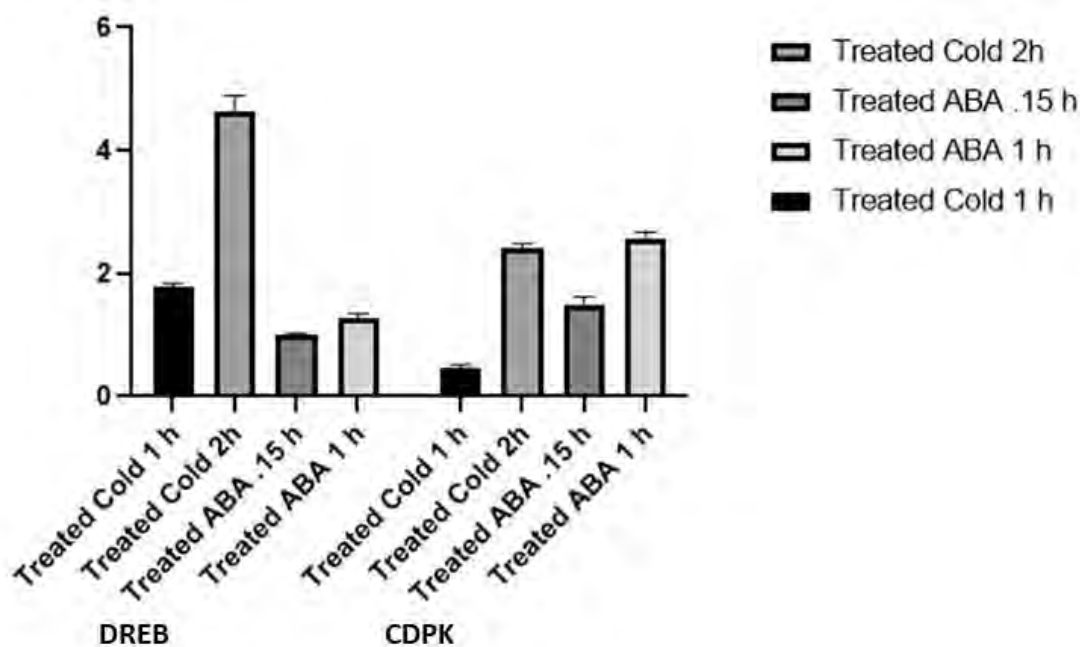


Figura 28. Análisis de expresión de los genes MbDREB1 y MbCDPK1 en *M. bombycina* bajo tratamientos de estrés por frío y ABA.

7.10 Caracterización y análisis bioinformático de proteínas putativas que corresponden a las subfamilias α -expansinas y β -expansinas en *M. bombycina*.

7.10.1 Alineamiento múltiple y análisis de la arquitectura de los dominios y motivos de las proteínas putativas de α -expansinas (MbEXPA) y β -expansinas (EXPB).

Se realizó el alineamiento múltiple de las secuencias de aminoácidos para cada subfamilia de las α -expansinas, con las secuencias homólogas de Jitomate (*Solanum lycopersicum*), Maíz (*Zea mays*) y *Arabidopsis* (*Arabidopsis thaliana*), en el cual se observaron las regiones conservadas características de estas proteínas en los que se identifican en recuadros negros los sitios conservados en cada subfamilia. Los Triptófanos

(W) conservados en EXPA: uno anterior al dominio 1 en el N-terminal (posición 46) y cuatro en el dominio 2 hacia el carboxilo terminal (posiciones 221, 228, 232 y 265). Con un rombo (♦), las cisteínas encargadas de formar los puentes disulfuro (posiciones 67, 95, 103, 112, 145 y 179). Se observaron los motivos conservados característicos de EXPA para las secuencias reportadas que son: TFWYG, GGACGYG, HFD (el sitio catalítico), WGQ y LSFpVT. Además de las regiones ATFYG, TAALS (inicio del dominio DPBB), RVPC, GGWCNPP, RVPC (término del dominio DPBB), RFTINGHS (inicio del dominio Alérgeno del pólén). Bajo la clasificación de las expansinas, el primer motivo a identificar fue el HFD, pues sólo está presente en EXPA y EXPB, por lo cual, quedan completamente descartadas las EXLA y EXLB. Además, motivos que sólo aparecen en EXPA son TAALS, GGACGYG y RFTINGHS con esos aminoácidos en ese orden (Figura 29 y Figura 30).



Figura 29. Alineamiento múltiple de α -expansinas (EXPA). Se identificaron con un asterisco (*), se identificaron Triptófanos conservados (W); con un rombo (♦), las Cisteínas (C) encargadas de formar los puentes disulfuro.

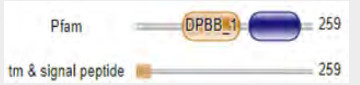
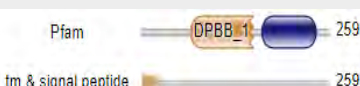
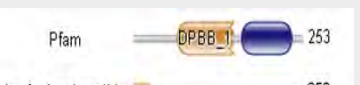












Figura 30. Alineamiento múltiple de β -expansinas (EXPB). Se identificaron con un asterisco (*), se identificaron Triptófanos conservados (W); con un rombo (◆), las Cisteínas (C) encargadas de formar los puentes disulfuro.

Para identificar a cada tipo de subfamilias de expansinas se obtuvieron los dominios conservados y se resumieron los resultados en los que se observa la arquitectura y los datos de los elementos de interés en la Tabla 14. Los rangos y longitudes de la péptido señal, de los dominios y de los puentes disulfuro; de estos últimos se omite la longitud debido a que no representan un rango, sino un enlace entre dos aminoácidos.

Tabla 13. Arquitectura de dominios de las α - expansinas y β -expansinas de *M. bombycina*.

Expansina	Arquitectura	Elemento	Posición inicial	Posición final	Longitud (aa.)
MbEXPA1.1		Péptido Señal	1	19	19
		Dominio 1	68	153	86
		Dominio 2	164	241	78
		Puente disulfuro 1	54	82	N/A
		Puente disulfuro 2	85	158	N/A
		Puente disulfuro 3	90	97	N/A
MbEXPA1.2		Péptido Señal	1	19	19
		Dominio 1	68	153	86
		Dominio 2	164	241	78
		Puente disulfuro 1	54	82	N/A
		Puente disulfuro 2	85	158	N/A
		Puente disulfuro 3	90	97	N/A
MbEXPA1.4		Péptido Señal	1	22	22
		Dominio 1	67	152	86
		Dominio 2	163	240	78
		Puente disulfuro 1	53	81	N/A
		Puente disulfuro 2	84	157	N/A
		Puente disulfuro 3	89	96	N/A
MbEXPA1.7		Péptido Señal	N/A	N/A	N/A
		Dominio 1	63	148	86
		Dominio 2	159	236	78
		Puente disulfuro 1	49	77	N/A
		Puente disulfuro 2	80	153	N/A
		Puente disulfuro 3	85	92	N/A
MbEXPA2.2		Péptido Señal	1	21	21
		Dominio 1	66	151	86
		Dominio 2	162	239	78
		Puente disulfuro 1	52	80	N/A
		Puente disulfuro 2	83	156	N/A
		Puente disulfuro 3	88	95	N/A
MbEXPA4.2		Péptido Señal	1	22	22
		Dominio 1	68	153	86
		Dominio 2	164	241	78
		Puente disulfuro 1	54	82	N/A
		Puente disulfuro 2	85	158	N/A
		Puente disulfuro 3	90	97	N/A

MbEXPA4.3		Péptido Señal	1	22	22
		Dominio 1	68	155	88
		Dominio 2	166	143	78
		Puente disulfuro 1	54	82	N/A
		Puente disulfuro 2	85	160	N/A
		Puente disulfuro 3	90	97	N/A
MbEXPA4.4		Péptido Señal	1	22	22
		Dominio 1	68	155	88
		Dominio 2	166	243	78
		Puente disulfuro 1	54	82	N/A
		Puente disulfuro 2	85	160	N/A
		Puente disulfuro 3	90	97	N/A
MbEXPA8.1		Péptido Señal	1	25	25
		Dominio 1	66	151	86
		Dominio 2	162	239	78
		Puente disulfuro 1	52	80	N/A
		Puente disulfuro 2	83	156	N/A
		Puente disulfuro 3	88	95	N/A
MbEXPA8.2		Péptido Señal	N/A	N/A	N/A
		Dominio 1	63	148	86
		Dominio 2	159	236	78
		Puente disulfuro 1	49	77	N/A
		Puente disulfuro 2	80	153	N/A
		Puente disulfuro 3	85	92	N/A
MbEXPA8.3		Péptido Señal	1	25	25
		Dominio 1	66	151	86
		Dominio 2	162	239	78
		Puente disulfuro 1	52	80	N/A
		Puente disulfuro 2	83	156	N/A
		Puente disulfuro 3	88	95	N/A
MbEXPA8.4		Péptido Señal	N/A	N/A	N/A
		Dominio 1	63	148	86
		Dominio 2	159	236	78
		Puente disulfuro 1	49	77	N/A
		Puente disulfuro 2	80	153	N/A
		Puente disulfuro 3	85	92	N/A
		Péptido Señal	1	36	36
		Dominio 1	73	159	87
		Dominio 2	170	247	78

MbEXPA13		Puente disulfuro 1	59	87	N/A
		Puente disulfuro 2	90	164	N/A
		Puente disulfuro 3	95	102	N/A
MbEXPA20		Péptido Señal	1	20	20
		Dominio 1	69	156	96
		Dominio 2	167	244	77
		Puente disulfuro 1	55	83	N/A
		Puente disulfuro 2	86	161	N/A
		Puente disulfuro 3	91	98	N/A
MbEXPA29		Péptido Señal	1	23	23
		Dominio 1	64	152	88
		Dominio 2	163	240	77
		Puente disulfuro 1	58	78	N/A
		Puente disulfuro 2	81	157	N/A
		Puente disulfuro 3	86	85	N/A
MbEXPB1.1		Péptido Señal	1	31	31
		Dominio 1	71	147	77
		Dominio 2	159	244	86
		Puente disulfuro 1	56	85	N/A
		Puente disulfuro 2	88	152	N/A
		Puente disulfuro 3	93	98	N/A
MbEXPB1.2		Péptido Señal	1	31	31
		Dominio 1	71	147	77
		Dominio 2	159	244	86
		Puente disulfuro 1	56	85	N/A
		Puente disulfuro 2	88	152	N/A
		Puente disulfuro 3	93	98	N/A
MbEXPB1.3		Péptido Señal	1	31	31
		Dominio 1	71	147	77
		Dominio 2	159	244	86
		Puente disulfuro 1	56	85	N/A
		Puente disulfuro 2	88	152	N/A
		Puente disulfuro 3	93	98	N/A
MbEXPB3.1		Péptido Señal	1	22	22
		Dominio 1	69	151	83
		Dominio 2	163	244	82
		Puente disulfuro 1	54	83	N/A
		Puente disulfuro 2	86	156	N/A

		Puente disulfuro 3	91	97	N/A
MbEXPB3.2		Péptido Señal	1	22	22
		Dominio 1	69	151	83
		Dominio 2	163	244	82
		Puente disulfuro 1	54	83	N/A
		Puente disulfuro 2	86	156	N/A
		Puente disulfuro 3	91	97	N/A

Se obtuvieron seis secuencias de los motivos conservados para cada una de las secuencias de las subfamilias de α -expansinas y 3 secuencias consenso para la subfamilia de β -expansinas en las proteínas de *Mammillaria bombycina*. En la Figura 31, se muestran los motivos únicamente de este organismo sin los homólogos.

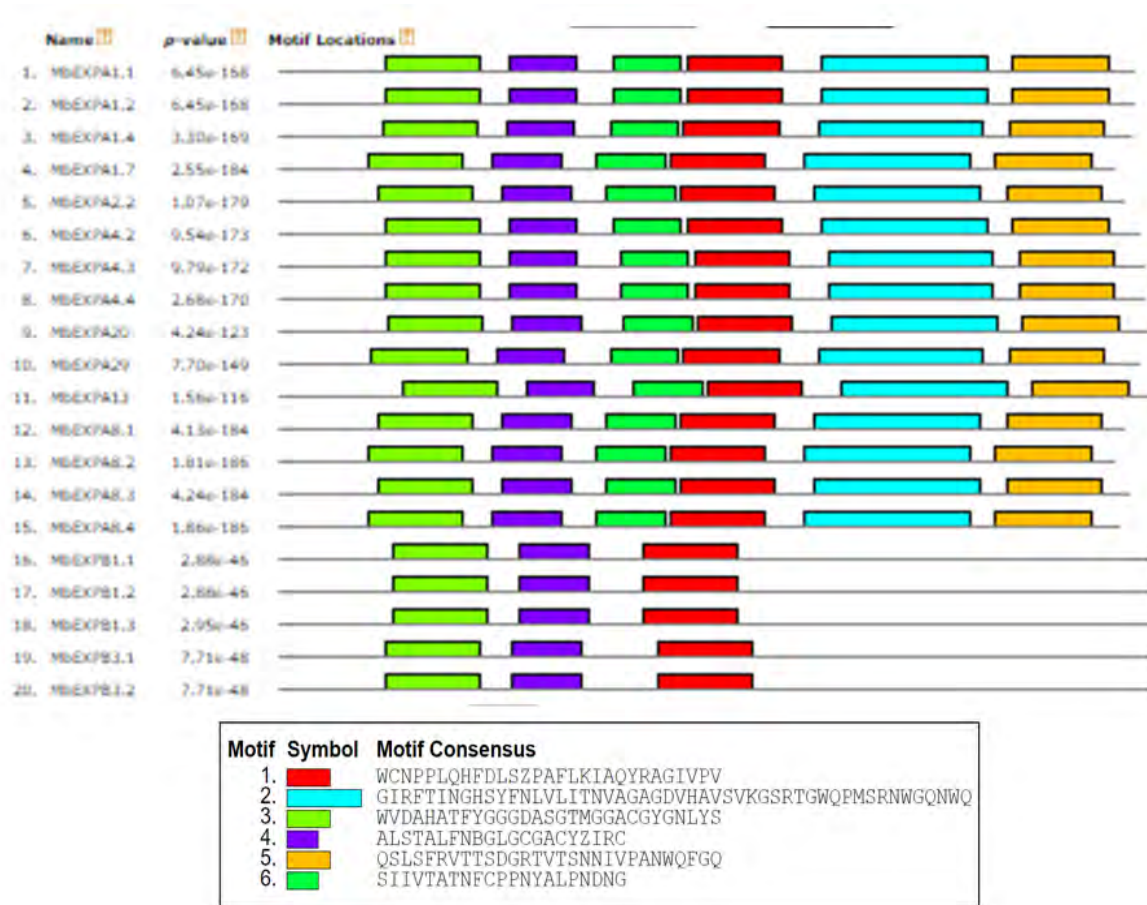


Figura 31. Motivos encontrados entre las subfamilias de α -expansinas y β -expansinas de *M. bombycina*. En el recuadro se observan las secuencias de cada uno de los motivos presentes en las subfamilias α -expansinas y β -expansinas.

Cada una de las secuencias de los motivos obtenidos se verificó a que dominio pertenece como se muestra en la Tabla 15. Donde se observa que pertenecen a alérgenos y a la secuencia del dominio DPBB, pertenecientes a las subfamilias de las α y β -expansinas.

Tabla 14. Información detallada de motivos encontrados.

Motivo	Longitud (a.a.)	Secuencia	Dominio HMMER
1	29	WCNPPLQHFDSLSPAFCLKIAQYRAGIVPV	DPBB
2	50	GIRFTINGHSYFNLVLTINVAGAGDVHAVSVKGSRTGW QPMSRNWQNWQ	Alérgeno
3	29	WVDAHATFYGGGDASGTMGGACGYGNLYS	N/A

4	21	ALSTALFNBGLGCGACYZIRC	DPBB
5	29	QSLSFrvTTSDGRTVTSNNIVPANWQFGQ	Alérgeno
6	21	SIIVTATNFCPPNYALPNDNG	DPBB

7.10.2 Caracterización fisicoquímica de las proteínas putativas de las subfamilias α -expansinas y β -expansinas.

Los resultados obtenidos de la caracterización de las proteínas MbEXPA y MbEXPB (Tabla 16), en la cual se puede observar la longitud del transcrito que oscila entre los 750 pb y 1524 pb, la longitud del CDS que va de 660 pb a 786 pb, el tamaño de las proteínas de 250 aa hasta los 262 aa, El Punto isoeléctrico y la localización celular.

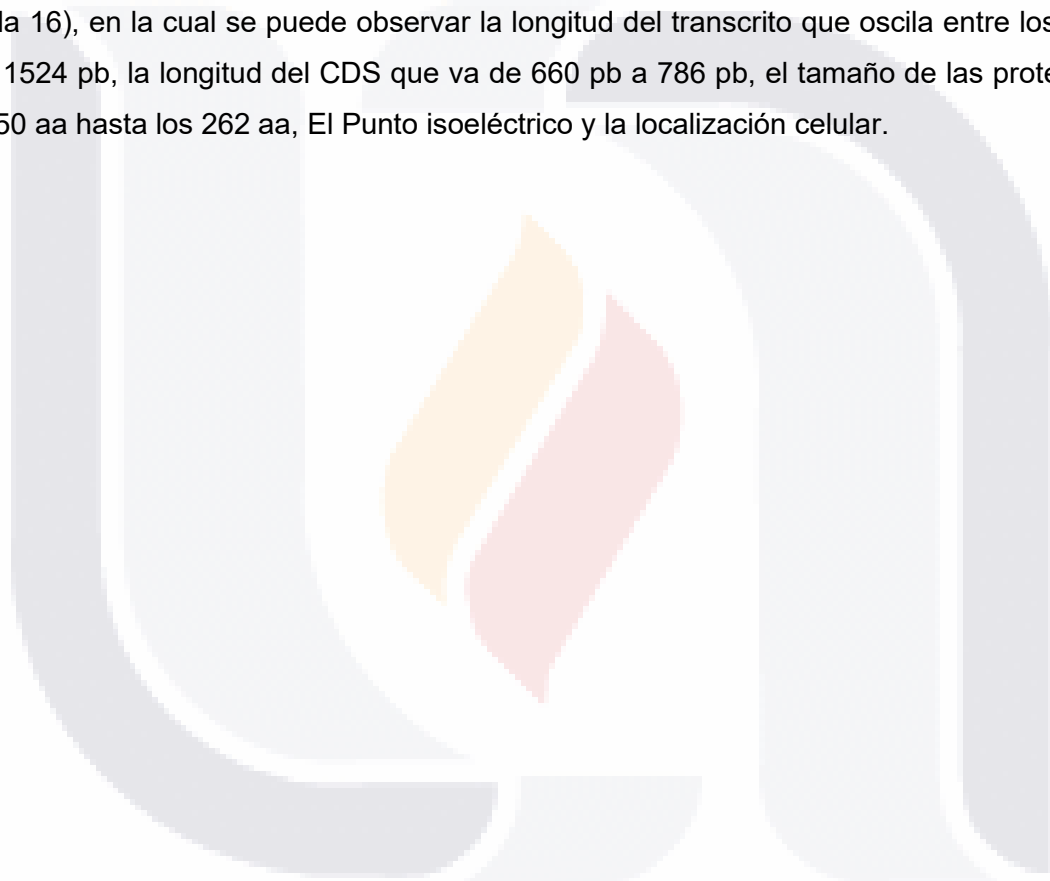


Tabla 15. Análisis fisicoquímico y localización celular de los transcritos de las subfamilias de α -expansinas y β -expansinas en *M. bombycina*.

Expansina	Transcrito	Longitud (pb)	Longitud (a.a.)	Peso molecular (kDa)	CDS (pb)	Punto isoeléctrico	Localización celular
MbEXPA1.1	>TRINITY_DN16283_c3_g1_i1	1404	256	27.04936	768	9.00	Chlo ^a Extr ^c
MbEXPA1.2	>TRINITY_DN16283_c3_g1_i2	1524	256	27.04936	768	9.00	Chlo ^a Extr ^c
MbEXPA1.4	>TRINITY_DN16283_c3_g1_i6	1387	255	26.86119	765	8.55	Extr ^{a,c}
MbEXPA1.7	>TRINITY_DN16606_c1_g2_i2	1134	250	26.92522	750	8.59	Nucl ^a Extr ^c
MbEXPA2.2	>TRINITY_DN16606_c1_g2_i4	971	253	26.68788	759	8.58	Nucl ^a Chlo ^b Extr ^c
MbEXPA4.2	>TRINITY_DN16283_c3_g1_i5	1466	257	28.07281	771	9.58	Chlo ^a Extr ^c
MbEXPA4.3	>TRINITY_DN16283_c3_g1_i7	1472	259	28.29420	777	9.34	Chlo ^a Extr ^c
MbEXPA4.4	>TRINITY_DN16283_c3_g2_i3	1007	259	28.04193	777	9.24	Extr ^{a,c}
MbEXPA8.1	>TRINITY_DN16606_c1_g1_i1	1061	253	27.25276	678	9.08	Nucl ^{a,b} Extr ^c
MbEXPA8.2	>TRINITY_DN16606_c1_g2_i3	1033	250	26.68785	678	8.07	Nucl ^{a,b} Extr ^c
MbEXPA8.3	>TRINITY_DN16606_c1_g1_i1.p1	759	253	27.25276	678	9.08	Nucl ^{a,b} Extr ^c
MbEXPA8.4	>TRINITY_DN16606_c1_g2_i3.p1	750	250	26.68785	678	8.07	Nucl ^{a,b} Extr ^c
MbEXPA13	>TRINITY_DN5999_c0_g1_i1	1030	262	28.21606	786	8.40	Chlo ^{a,b} Extr ^c
MbEXPA20	>TRINITY_DN3370_c0_g1_i2	1153	260	28281.72	780	5.03	Extr ^{a,c}
MbEXPA29	>TRINITY_DN16606_c0_g1_i2	1416	256	27652.41	768	9.10	Extr ^{a,c}
MbEXPB1.1	>TRINITY_DN9214_c0_g1_i1	982	261	28.44764	693	5.35	Extr ^{a,b} Perip ^c
MbEXPB1.2	>TRINITY_DN9214_c0_g1_i1	982	261	28.44764	693	5.35	Extr ^{a,b} Perip ^c
MbEXPB1.3	>TRINITY_DN9214_c0_g1_i1.p1	783	261	28.44764	693	5.35	Extr ^{a,b} Perip ^c
MbEXPB3.1	>TRINITY_DN17684_c1_g1_i1	1205	261	28.51954	660	8.05	Extr ^{a,b} Cytop ^c Perip ^c
MbEXPB3.2	>TRINITY_DN17684_c1_g1_i1.p1	783	261	28.51954	660	8.5	Extr ^{a,b} , Cytop ^c Perip ^c

*La predicción de la localización fue en Wolf Psort (**a), Localizer (**b) y CELLO (**c).

7.10.3 Modelado 3D hipotético de las proteínas putativas de las subfamilias α -expansinas y β -expansinas.

Estructuras 3D de las proteínas putativas de MbEXPA8.1 y MbEXPB1.1 de *M. bombycina*. Generadas por modelado de la homología de las estructuras terciarias de las proteínas putativas de la estructura cristalina de Phl p 1 (1N10, *Phleum pratense*), un importante alérgeno de polen de hierba de Timothy en el programa CPHmodels 3.2. En la Figura 32 en el inciso A) MbEXPA8.1, se observa en color naranja el dominio DPBB, en azul el dominio Alérgeno del polen, en rosa se destaca el péptido señal y en verde se

aprecia la presencia de puentes disulfuro. B) MbEXPB1.1, estructura diseñada a partir del templete 2HCZ (*Zea mays*), de color: Naranja, dominio 1 DPBB; Azul, dominio 2 Alérgeno del Pólen; Verde, parejas de puentes disulfuro

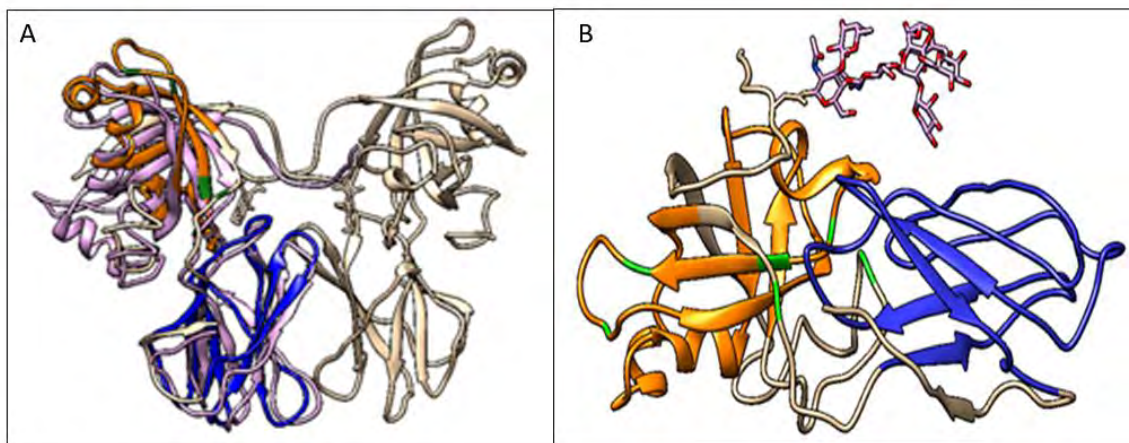


Figura 32. Estructura 3D de las proteínas putativas de las subfamilias de α -expansinas y β -expansinas. A) MbEXPA8.1, alineada al templete 1N10, en color naranja se observa marcado el dominio 1, en azul el dominio 2, en rosa se destaca el péptido señal y en verde se aprecia la presencia de puentes disulfuro. B) MbEXPB1.1 alineada al templete 2HCZ, de color: Naranja, dominio 1 DPBB; Azul, dominio 2 Alérgeno del Pólen; Verde, parejas de puentes disulfuro.

8 Discusión

En la actualidad el manejo de herramientas de secuenciación masiva como el RNA-seq son de gran importancia, debido a la basta cantidad de información que se obtiene en comparación a otros métodos como los microarreglos, además de que no se requiere de un conocimiento previo de las secuencias de RNA mensajeros, se puede detectar la transcripción desde niveles muy bajos, en un tiempo relativamente corto, a costos bajos, entre otros beneficios (Zhao et al. 2014).

Lo anterior se refleja en la gran cantidad de trabajos realizados en los últimos años por los investigadores, como lo menciona Rodríguez et al (2018), en el que analizaron el aumento del número de publicaciones que utilizan RNA-seq y que se encuentran indizadas por PubMed. La búsqueda se realizó restringiéndose a las publicaciones de estudios de expresión génica que incluyen RNA-seq en el título o el resumen, observaron que, en el año 2009, las publicaciones con las características antes mencionadas eran menos de 500, mientras que en el año 2017 rebasaban las 3500 publicaciones.

Debido a los múltiples beneficios de esta técnica, en este trabajo se realizaron las anotaciones del Transcriptoma de *M. bombycina*, debido a que no existe ningún análisis previo de secuenciación de nueva generación en esta especie. Por lo que se considera relevante analizar los transcritos que se están generando en condiciones *in vitro*, tomando como base la importancia de utilizar técnicas que permitan el estudio de organismos de los que no se tienen conocimientos previos de secuenciación (sin un genoma de referencia) a los que se les considera no modelo., que radica en que se pueden estudiar procesos biológicos que son inaccesibles cuando se estudian especies a las que se les considera modelo, como lo son *A. Thaliana*, *Z. mays*, *G. max*, entre otros. La importancia económica o bien por el fácil manejo de estas en el laboratorio, por lo que el uso de especies no modelo como son las especies extremófilas podrían ayudar a enfrentar problemas derivados del cambio climático, la destrucción del hábitat y las poblaciones humanas en expansión, entre otros, (Gladfelter, 2015).

Apartir de explantes *in vitro* de *M. bombycina*, se extrajo el material biológico (ARN) a una concentración de 4ug/20uL, obteniendo un valor RIN > 4, indicando una buena integridad además de que los electroferogramas indican que no tiene contaminantes para cada una de las muestras analizadas.

TESIS TESIS TESIS TESIS TESIS

Se secuenciaron muestras por triplicado de *M. bombycina* en condiciones *in vitro* en el equipo Illumina Next Seq-500. De aquí, se obtuvieron entre 6 y 9 millones de lecturas, con una longitud de secuencia de 76 pb, por lo que los datos obtenidos en el transcriptoma de *M. bombycina* coinciden en cantidad y calidad con otras investigaciones como es el trabajo realizado en la especie *Hylocereus polyrhizus* (Xu et al. 2019).

Sin embargo, a medida que la secuenciación por síntesis procede a la secuencia de los últimos ciclos, la fuerza de la señal de cada grupo de ADN se vuelve más débil de manera progresiva. Este problema hace que algunas bases como son T y C sean mal interpretadas como G en las colas de lectura, esto genera un error de secuenciación conocido como cola polyG. (Shifu et al. 2018). Por esta razón el resultado de las secuencias obtenidas del programa Fastp fueron los utilizados para el ensamble, representando el 95% aprox. de las lecturas obtenidas por el secuenciador fueron retenidas para su posterior ensamble (Li et al. 2015).

Son muchas las ventajas de utilizar esta plataforma de secuenciación como es que las longitudes de las lecturas obtenidas sean cortas, el alto número de lecturas generadas incrementarán la cobertura y permitirá extender la secuencia hasta poder, en algunos casos, obtener la secuencia de todo el transcrito, además de que su costo es relativamente bajo y el tiempo de secuenciación es muy corto (Soto Sedano et al., 2012).

Otra de las ventajas de secuenciar en plataformas Illumina, es que al ser una de las más utilizadas por la comunidad científica, las herramientas bioinformáticas compatibles para realizar el pretratamiento (trimming y filtrado) que se les da a las secuencias obtenidas son muchas y son de acceso libre al utilizar una plataforma UNIX (Shifu et al. 2018).

Después de realizar el ensamble con el programa Trinity se obtuvieron un total de 78,412,335 pb, más, de las que obtuvieron en otras especies como *Dimocarpus longan* (Lai et al. 2013), *Camellia sinensis* (Shi et al. 2011), *Gossypium hirsutu* (Wang et al. 2010) y *Amaranthus tricolor* (Liu et al. 2014), entre otros.

Al realizar el análisis de ContigNX se obtuvo que por lo menos el 50% de los unigenes obtenidos contaban con una longitud de 1,574 pb, estos resultados corroboran que los datos de secuenciación son de cantidad y calidad suficientes para garantizar la precisión del ensamblaje de la secuencia, de igual forma demuestran que la tecnología de

secuenciación Illumina NEXTSeq 500 puede ser una herramienta eficaz para el descubrimiento de genes en organismos no modelo (Bhowal et al. 2018).

Las anotaciones del transcriptoma *de novo* de *M. bombycina* se realizaron en condiciones *in vitro* debido a que no existían análisis previos de secuenciación de nueva generación en esta especie. Por lo tanto, *M. bombycina* se propone como planta modelo para su estudio en cactáceas.

Se obtuvieron un total de 84 975 transcripciones encontradas en bases de datos públicas, como se muestra en la Tabla 6. De los cuales se obtuvieron 47 406 unigenes (el 55 % son únicas) y la mayoría de estas provienen de *A. thaliana* (Figura 15). Se obtuvieron resultados similares a *Dimocarpus longan* (Lai et al. 2013), donde de 13 muestras, los autores generaron 64 876 258 lecturas limpias; sin embargo, solo obtuvieron 68.925 unigenes. De igual forma, en *Agave sisalana* (Sarwar et al. 2019) se obtuvieron 276,8 millones de lecturas de 9 muestras, registrándose 93.141 genes, de los cuales 67.328 son unigenes. Por lo tanto, si se hace una comparación entre las transcripciones encontradas y los unigenes encontrados a partir de ellos se observa que las anotaciones del transcriptoma de *M. bombycina* *de novo* es un análisis sólido y de buena calidad.

Al realizar el análisis de las anotaciones obtenidas, se identificó la presencia de la familia de las Glioxalosas, la cual tiene una importante participación en la desintoxicación del metilglioxal (MG). El MG es un subproducto metabólico de varios procesos celulares como son la fotosíntesis, la glucólisis, el metabolismo de las proteínas y la peroxidación de lípidos (Thornalley et al. 2003). Sin embargo, cuando la planta se encuentra bajo condiciones estresantes la cantidad de MG en las células se eleva considerablemente (Yadav et al. 2005), dañando las células ya que es un agente glicante, que causa cambios negativos irreversibles a proteínas, lípidos y ácidos nucleicos, dando como resultado modificaciones en los componentes celulares de la planta (Kaur et al. 2015).

Para contrarrestar la intoxicación que el MG pueda generar en la célula, la familia de las glioxalosas puede actuar a través de dos vías; la primera vía está compuesta por la acción de dos glioxalosas (metaloenzimas). La primera es la glioxalasa I (GLYI) y la segunda glioxalasa II (GLYII). La acción sinérgica de ambas cumple la función de convertir MG en D-lactato, y posteriormente se metaboliza a piruvato por D-lactato deshidrogenasa (D-LDH) (Bhowal et al. 2020). La segunda vía de descomposición de MG es independiente de

metaloenzimas, a la cual se le conoce como glioxalasa III (DJ-1), la cual actúa de una manera directa para convertir el MG a D-Lactato deshidrogenasa (Li et al. 2019).

Kaur et al (2014), menciona que debido a la capacidad de desintoxicación del MG por la acción de las glioxalosas sería apropiado denominar al MG y a las glioxalosas como posibles biomarcadores de tolerancia al estrés en plantas, por esta razón en los últimos años se ha visto una gran cantidad de trabajos centrados en identificar la acción de las glioxalosas en las plantas, como es *Arabidopsis thaliana*, *Vitis vinífera* (Bhowal et al. 2020), *Medicago truncatula* (Gosh et al. 2017), *Oryza sativa* (Mustafiz et al. 2011), *Glycine max* (Gosh et al. 2016), entre otros. Sin embargo, en la familia *Cactaceae* es casi nula, por lo que aquí radica la importancia de ubicar a esta familia enzimática en el transcriptoma de *M. bombycina*.

De las diferentes isoformas que se encontraron en el transcriptoma de *M. bombycina*, únicamente diez tenían por lo menos un dominio característico de las Gly de tipo 1, por lo que primeramente se analizaron cada una de ellas, observando que MbGLYI_1, MbGLYI_2 y MbGLYI_3 únicamente tenían un dominio de aproximadamente 140 aa (Figura 20). Las secuencias MbGLYI_4 – MbGLYI_10 estaban conformadas por dos dominios de aproximadamente 120 aa cada uno de ellos (Figura 20), coincidiendo con otras especies de plantas como son *Arabidopsis thaliana*, *Oryza sativa* (Mustafiz et al., 2011), *Glycine max* (Gosh et al., 2016), *Medicago truncatula* (Gosh et al. 2017) y *Vitis vinífera* (Bhowal et al. 2020).

Las secuencias del dominio GLYI de todas las proteínas putativas de MbGLYI se alinearon con las de las proteínas de otras plantas bien caracterizadas (Figura 17), dando como resultado que MbGLYI_1, MbGLYI_2 y MbGLYI_3 tienen unión al ion metálico Zn^{2+} (Figura 20) es homodimérico (cuenta con un solo dominio PF00903) y ya ha sido reportado en AtGLY1-2 en *Arabidopsis* y BjGLY1 de *B. juncea* (Deswal et al. 1998).

De acuerdo con la alineación, las proteínas MbGLYI_1, MbGLYI_2 y MbGLYI_3 dependiente de Zn^{2+} tiene un tramo adicional de 12 aa (Figura 17), que está completamente ausente en las proteínas MbGLYI_4 – MbGLYI_10 dependientes de Ni^{2+} como es el caso de las proteínas de GLYI de *O. sativa* como lo menciona Mustafiz et al (2011).

Mientras que las proteínas MbGLYI_4 – MbGLYI_10 tienen unión al ion metálico Ni^{2+} , cuentan con dos dominios PF00903, de acuerdo con análisis bioquímicos de las

propiedades de OsGLY1-11.2 sugieren que este dominio tiene 2 sitios activos, de los cuales solo uno es funcional (Mustafiz et al. 2014).

La importancia de identificar las proteínas putativas de MbGLYI es que pertenecen a Zn^{2+} o Ni^{2+} es que dependiendo de la clasificación de la afinidad de cada una de las secuencias es crucial para poder caracterizar las proteínas, debido a que se ha descubierto que la actividad enzimática de las proteínas GLYI depende de los iones metálicos (Kaur et al. 2013). Un ejemplo de esto se puede observar en el trabajo realizado por Batth et al (2020) en donde mencionan que las proteínas AtGLY12 dependientes de Zn^{2+} es la principal enzima responsable de la desintoxicación de MG y la tolerancia al estrés por salinidad.

Para identificar si cada una de las proteínas putativas de GLYI son activas, las secuencias de MbGLYI_1 MbGLYI_10 se clasificaron como se puede observar en la Tabla 9, en donde las diez secuencias encontradas se clasificaron como similares a GLYI activas, esto independientemente del ion metálico al que se unen, ya que cada una de ellas cuenta con los 4 sitios de unión conservados necesarios para la actividad de la glioxalasa: el sitio activo, el sitio de unión al metal, el sitio de unión de GSH y la interfaz del dímero, comparándolas en el alineamiento múltiple con proteínas bien caracterizadas (material suplementario) como son VvGLYI (Bhowal et al. 2020), OsGLYI (Mustafiz et al. 2011), y MtGLYI (Gosh et al. 2017).

En el caso de las proteínas putativas clasificadas como MbGLYII, se identificaron cuatro secuencias, las cuales presentaban los dominios Metallo-beta-lactamasa (PF00753) e Hidroxiacil glutatión hidrolasa C-terminal (PF16123), además de los sitios de unión característicos para que a esta proteína se le considere activa: el sitio de unión a metales (THXHXDH), un sitio activo (C/GHT) y siete sitios de unión conservados (C/K(R)/F(Y)/Y/N/R/K) como se puede observar en el alineamiento múltiple (Figura 18) en el que se comparó con las secuencias bien caracterizadas de *Arabidopsis thaliana* (Limphong, 2009), *Oryza sativa* (Mustafiz et al. 2011), *Glycine max* (Gosh et al. 2016), *Medicago truncatula* (Gosh et al. 2017) y *Vitis vinífera* (Bhowal et al. 2020).

La identificación del sitio THXHXDH tiene una relevancia alta en la actividad de la proteína debido a su participación en la unión al ion zinc como lo menciona Saxena et al (2005) en donde concluyen que la expresión de BjpglyII bajo diversas condiciones de estrés abiótico mostró que está regulada positivamente por la salinidad, el estrés por metales pesados y ABA. De igual forma Yadav et al (2007) reportan que en el análisis de la transcripción en

arroz mostró que la expresión del gen *OsglyII* se estimula en 15 minutos en respuesta a diversas tensiones abióticas, así como al tratamiento con ácido abscísico o ácido salicílico. La vía de desintoxicación del MG en donde participan conjuntamente las *GLYI* y *GLYII* se ha relacionado fuertemente con la tolerancia a estrés abiótico como lo es en el trabajo de Singla-Pareek et al (2003), en donde mencionan que la sobreexpresión de *GLYI* + *GLYII* en conjunto confiere una mejor tolerancia a la salinidad en plantas de tabaco (*Brassica juncea*).

En cuanto a las proteínas putativas *MbGLYIII*, sólo una de ellas; *MbGLYIII_1* mostró ser activa, ya que, al realizar el análisis de la arquitectura de los dominios presentó los fragmentos en tándem de la región DJ-1 / Pfpl, por lo que sólo se utilizó esa secuencia para el alineamiento múltiple (Figura 19) e identificación de la triada catalítica Asp (Glu) -His-Cys conservada en su sitio activo (Quigley et al. 2003). Es importante destacar que las proteínas que tienen el dominio DJ-1 / Pfpl se le atribuyen otras funciones fisiológicas, ya que, además de descomponer el MG en un solo paso en D-Lactato sin necesidad de reducir GSH, como son la regulación de la transcripción Clements et al., 2006), la estimulación de enzimas con relevancia en la acción antioxidante (Xu et al. 2010), entre otras.

Una vez realizado los análisis de alineamiento múltiple y las relaciones filogenéticas de la proteína putativa *MbGLYDJ-1* con otras proteínas bien caracterizadas se encontró que es muy parecida a la isoforma *AtDJ1D* de *A. thaliana* que se identificó primeramente como una chaperona molecular que es inducible por calor (Subedi et al. 2011), lo que permite predecir su función en la especie *M. bombicyna* (Figura 21).

A pesar de que la vía de la glioxalasa dependiente del glutatión, (glioxalasa I/II), se le considera que es la ruta más importante para la desintoxicación de MG (MacLean et al. 1998; Kaur et al. 2014), se debe tener en cuenta que para que este sistema funcione el nivel de GSH disponible debe estar en cierta cantidad. Sin embargo, esta disponibilidad es variable debido a las condiciones en las que se encuentra la planta, sobre todo en la accesibilidad que tiene a la glucosa en un medio mínimo (Subedi et al. 2011).

La acción de la glioxalasa III puede tener una actividad bastante buena bajo condiciones en donde la fuente de carbono sea limitada, como se puede observar en el trabajo realizado por Benov et al (2014), en donde demostraron que la actividad de la glioxalasa I y II en *E. coli*, era constitutiva independientemente de la fase de crecimiento en la que se encontraba, mientras que la glioxalasa III alcanzaba su máxima expresión durante la fase estacionaria.

ESTESIS ESTESIS ESTESIS ESTESIS ESTESIS

Esto abre una gran brecha para la investigación de la proteína DJ1-D en plantas, como una probable solución para cultivos de interés agronómico que se encuentren en condiciones de suelos deficientes.

También se caracterizaron en el transcriptoma *de novo* los factores de transcripción DREB y CDPK que tienen una gran relevancia en el mantenimiento de la homeostasis en la planta cuando se encuentra bajo estrés ambiental. Se identificaron dos isoformas para el gen MbDREB y una isoforma MbCDPK, los cuales en este trabajo se postulan como proteínas activas ya que cada uno de ellos tienen los dominios y motivos conservados característicos de estos dos genes (Figura 21)

Se ha observado en distintos trabajos la sobre expresión de estos genes en otras plantas, como lo es *Ipomoea batatas*, en presencia de deshidratación, frío, condiciones de alta salinidad y cadmio. Por ejemplo, en el gen swDREB1 en la especie *Ipomoea batatas*, se sobre expresó tanto en tallos como en raíces (Kim et al. 2007). En lo que respecta a CDPK, Se han reportado trabajos relacionados con diferentes subfamilias de CDPK que tienen funciones contra el estrés por sequía, la señalización de la luz y respuesta al estrés por temperaturas bajas. Sin embargo, la más importante y la que interesa a nuestra investigación debido al enfoque a la respuesta a factores abióticos, son CDPK1 y CDPK ya que son activadas constitutivamente y alteran la expresión de genes de respuesta a estrés (Nie et al. 2015).

En los análisis de expresión mediante tiempo real de estos genes, se observó que el gen MbDREB tiene un aumento considerable en el estrés por frío a 1 y 2 h con 180 y 240% de expresión respectivamente, esto considerando que el control está normalizado a cero (Figura 28). Sin embargo, en el estrés por ABA no se observa un aumento significativo con respecto al control. De acuerdo con Hirayama y Shinozaki (2010), el factor de transcripción DREB es una de las piezas clave frente a la respuesta al frío, ya que logra aumentar de manera considerable y a regular la homeostasis de varios grupos de genes que favorecen la estabilidad de membranas celulares frente a lesiones provocadas por el congelamiento.

En el caso del gen CDPK en presencia de estrés por frío a 1 y 2 h, se observó un aumento en la expresión de 60% y 220% respectivamente, mientras que este mismo gen en presencia de ABA a .15 y 1 h aumenta 170% y 230% respectivamente con respecto al control. Pandey (2011) menciona los genes CDPK aumentan su expresión en presencia de estrés abiótico debido a que está formada por proteínas que regulan la fosforilación de residuos

serina/treonina de proteínas específicas con las células. Un ejemplo de esto es el incremento de la expresión de estos genes en plantas transgénicas de tomate en presencia de estrés abiótico (Pineda et al. 2012)

Se identificaron veinte secuencias que codifican a dos subfamilias de expansinas, de las cuales quince pertenecen a α -expansinas y cinco a β -expansinas. Las expansinas son proteínas que modifican la pared celular de los tejidos vegetales. A cada una de las secuencias de expansinas encontradas en el transcriptoma, se le realizó un análisis bioinformático. Kan-Ming et al (2020), mencionan que las expansinas deben de contener 2 dominios, el primero es el DPBB, que codifica para dominio N-terminal que tiene una estructura de barril β de doble, que es un alérgeno de polen y el dominio GH45 que codifica a las endoglucanasas glucósido hidrolasas, Dominios que tienen las secuencias MbEXPA y MbEXPB, los cuales se pueden observar en la tabla 14.

En la Figura 29, se observa el alineamiento múltiple de estas con las de *Z. mays* y *A. thaliana*, se observa que se obtuvieron seis secuencias de los motivos conservados para cada una de las secuencias de las subfamilias de α -expansinas y 3 secuencias consenso para la subfamilia de β -expansinas (Tabla 15) encontrados en todas las expansinas. De igual manera cada una de la arquitectura de los dominios están representados en la Figura 31, en dónde se identifica la diferencia entre las α -expansinas que tienen en su estructura 6 motivos y las β -expansinas que tienen en su arquitectura únicamente 3 dominios, así como lo menciona Cosgrove en (2017).

Se realizó el análisis fisicoquímico de las secuencias (Tabla 16) y un modelado 3D hipotético de las α -expansinas y β -expansinas. En la figura 32, inciso A, se observa la proteína hipotética MbEXPA8.1, de color Naranja el dominio 1 DPBB; de Azul el dominio 2 Alérgeno del Pólen; Verde, parejas de puentes disulfuro, a partir del templete 1N10, la cual se obtuvo a partir de la estructura cristalina de Phl p 1, que codifica a un alérgeno de polen de hierba de Timothy (*Phleum pratense*).

En la figura 32, inciso B, se observa que el modelado de MbEXPA8.1 en color naranja está el dominio DPBB, en azul el dominio Alérgeno del polen, en rosa se destaca el péptido señal y en verde se aprecia la presencia de puentes disulfuro. Para la proteína MbEXPB1.1 se obtuvo el diseño a partir del templete 2HCZ, la cual se obtuvo a partir de la estructura cristalina de EXPB1, una β -expansina y alérgeno de polen del grupo 1 del maíz (*Zea mays* La importancia de su identificación radica en que están involucradas en varios procesos

fisiológicos importantes en la planta para su óptimo desarrollo, por ejemplo, la plasticidad celular, apertura y cierre estomático, maduración del fruto, procesos de abscisión, entre otros (Cosgrove, 2012).



9 Conclusiones

Se generó el transcriptoma de *M. bombycina* bajo condiciones *in vitro* donde se obtuvieron 47406 secuencias de genes y los cuales fueron depositados en la base de datos SRA de NCBI con el número de acceso PRJNA764261. En este transcriptoma, se identificaron y se estudiaron bioinformáticamente genes para glioxalosas, algunos factores de transcripción y de expansinas.

En lo que respecta a las glioxalosas, se identificaron dieciseis secuencias; diez para MbGLYI y cinco para GLYII y una para MbDJ-1. El análisis de expresión génica reveló que la MbGLYI_1, MbGLYII_1 y MbDJ-1 bajo los tratamientos de estrés abiótico, tienen un aumento en la expresión con respecto al control.

Para los factores de transcripción, se identificaron y analizaron tres secuencias que codifican para los factores DREB y CDPK. En el análisis de expresión por qRT-PCR, se observó un aumento en la expresión del gen MbDREB1 y MbCDPK1 bajo tratamientos de estrés por frío y en presencia de ABA.

En lo referente a las expansinas, se identificaron veinte secuencias que codifican para las subfamilias de α y β -expansinas. El análisis bioinformático reveló que cumplen con todas las características de las α y β -expansinas.

10 Bibliografía

- Aceituno, F. F., Mosekyo, N., Rhee, S. (2008). The rules of gene expression in plants: Organ identity and gene body methylation are key factors for regulation of gene expression in *Arabidopsis thaliana*. *BMC Genomics*.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell*, 4th edition. New York: Garland Science; 2002. ISBN-10: 0-8153-3218-1 ISBN-10: 0-8153-4072-9
- Alharbi, B.M., Elhakem, A.H., Alnusairi, G.S.H., Mona, H., Soliman, M.H., Hakeem, K.R., Hasan Md.M., Abdelhamid M.T. (2021). Exogenous application of melatonin alleviates salt stress-induced decline in growth and photosynthesis in *Glycine max* (L.) seedlings by improving mineral uptake, antioxidant and glyoxalase system. *Plant Soil Environ.*, 67: 208–220.
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Anderson, E. F. (2001). *The cactus family*. Timber Press, Portland, Oregon, USA.
- Agarwal, P.K. & Jha. B. (2010). Transcription factors in plants and ABA dependent and independent abiotic stress signalling. *Biologia Plantarum*. 54, 201– 212.
- Alfred, J. & Baldwin. I.T. (2015). New opportunities at the wild frontier. *eLife* 4: e06956.
- Apel, K. & Hirt, H. (200) Reactive oxygen species: Metabolism, oxidative stress and signal transduction. *Annu. Rev. Plant Mol. Biol.* 2004, 55, 373–399.
- Barrell, D., Dimmer, E., Huntley. R.P., Binns, D., O'Donovan, C. & Apweiler R. (2009). The GOA database in 2009--an integrated Gene Ontology Annotation resource. *Nucleic Acids Res* 2009; 37: D396-403.
- Babraham Institute. (2000). Babraham Bioinformatics. Obtenido de <http://www.bioinformatics.babraham.ac.uk/index.html>
- Bhowal, B., Singla-Pareek, S.L. & Sopory, S.K. (2020). From methylglyoxal to pyruvate: a genome-wide study for the identification of glyoxalases and D-lactate dehydrogenases in *Sorghum bicolor*. *BMC Genomics* 21, 145. <https://doi.org/10.1186/s12864-020-6547-7>

- Bartles, D. & Kotchoni, S.O. (2003). Water stress induces the up-regulation of a specific set of gens in plants: aldehyde dehydrogenases as an example. *Bulg. J. Plant Physiol*, Special Issue, 37-51.
- Birol, Inanç., Jackman, S. D., Nielsen, C. B., Qian, J. Q., Varhol, R., Stazyk, G., Morin, R. D., Zhao, Y., Hirst, M., Schein, J. E., Horsman, D. E., Connors, J. M., Gascoyne, R. D., Marra, M. A. & Jones, S. J. M. (2009). De novo transcriptome assembly with ABySS. *10.1093/bioinformatics/btp367*. *Bioinformatics*. 25 (21): 2872-2877. <https://doi.org/10.1093/bioinformatics/btp367>
- Bruijn, N.G. (1946). A combinatorial problem *Proceedings of the Section of Sciences of the Royal Nethe43rlands Academy of Arts and Sciences in Amsterdam*, 49 (7), 758-764.
- Chaisson, M.J., Huddleston, J., Dennis, MY., Sudmant, P.H., Malig, M., Hormozdiari, F., Antonacci, F., Surti, U., Sandstrom, R., Boitano, M., Landolin, J.M., Stamatoyannopoulos, J.A., Hunkapiller, M.W., Korlach, J. & Eichler, E.E. (2015). Resolving the complexity of the human genome using single-molecule sequencing. *Nature*. 517 (7536), 608–611. <https://doi.org/10.1038/nature13907>
- Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Instituto del Medio Ambiente del Estado de Aguascalientes (IMAE), Universidad Autónoma de Aguascalientes (UAA). 2008. *La Biodiversidad en Aguascalientes: Estudio de Estado*. México. 385 pp
- Cosgrove D. J. (2017). Microbial Expansins. *Annual Review of Microbiology*, 71(1).
- Concepción J., Puerta B., Claudia P. y Ureña P. (2005). *Prácticas de biología molecular*. Ed. Pontificia Universidad Javeriana. Colombia. 100 pp
- Deswal, R. & Sopory, S.K. (1998) Biochemical and immunochemical characterization of *Brassica juncea* glyoxalase I. *Phytochemistry*, 49, 2245–2253.
- Dubois, M., Claeys, H., Van den Broeck, L., & Inzé, D. (2017). Time of day determines *Arabidopsis* transcriptome and growth dynamics under mild drought. *Plant, cell & environment*, 40(2), 180–189. <https://doi.org/10.1111/pce.12809>
- Egan, A.; Schluete, J.; & Spooner, D., (2012). Applications of Next-Generation Sequencing in Plant Biology», *American Journal of Botany*. 99 (2): 175-185.

- Fedoroff, N.V., Battisti, D.S., Beachy, R.N., Cooper, P.J., Fischhoff, D.A., Hodges, C.N., Knauf, V.C., Lobell, D., Mazur, B.J. & Molden, D. (2010) Radically rethinking agriculture for the 21st century. *Science*: 327:833–834. [PubMed: 20150494]
- Garzón-Martínez, G.A., Zhu, Z.I. & Landsman, D. (2012). The *Physalis peruviana* leaf transcriptome: assembly, annotation and gene model prediction. *BMC Genomics* **13**, 151. <https://doi.org/10.1186/1471-2164-13-151>
- Gill, S.S.; Tuteja, N. (2010). Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol. Biochem.* (48): 909–930.
- Gibson, A. C., and P. S. Nobel. (2010). *The cactus primer*. Harvard University Press, Cambridge, Massachusetts, USA.
- Ghosh, A.; Pareek, A.; Sopory, S.K. & Singla-Pareek, S.L. (2014). A glutathione responsive rice glyoxalase II, OsGLYII-2, functions in salinity adaptation by maintaining better photosynthesis efficiency and anti-oxidant pool. *Plant J.* 80, 93–105
- Ghosh, A.; Kushwaha, H.R.; Hasan, M.R.; Pareek, A.; Sopory, S.K. & Singla-Pareek, S.L. (2016). Presence of unique glyoxalase III proteins in plants indicates the existence of shorter route for methylglyoxal detoxification. *Sci. Rep.* 6, 18358.
- Ghosh A. (2017). Genome-Wide Identification of Glyoxalase Genes in *Medicago truncatula* and Their Expression Profiling in Response to Various Developmental and Environmental Stimuli. *Front Plant Sci.* 8: 836.
- Gladfelter A. S. (2015). How nontraditional model systems can save us. *Molecular biology of the cell*, 26(21), 3687–3689. <https://doi.org/10.1091/mbc.E15-06-0429>
- González, J.M., Agostini, R.B., Alvarez, C.E., Klinke, S., Andreo, C.S. & Campos, V.A. (2019). Deciphering the number and location of active sites in monomeric glyoxalase I from *Zea mays*. *FEBS J*, 286: 3255-3271. <https://doi.org/10.1111/febs.14855>
- Grabherr, M., Haas, B., Yassour, M. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29, 644–652 <https://doi.org/10.1038/nbt.1883>

- Golldack D., Lüking I., Yang O. (2011). Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. *Plant Cell Reports*. 30, 1383–1391.
- Green, S., Dietrich, M.R., & Leonelli, S. (2018). Extreme' organisms and the problem of generalization: interpretation of Krogh's principle. *HPLS* 40, 65. <https://doi.org/10.1007/s40656-018-0231-0>
- He, S. L., & Green, R. (2013). Northern blotting. *Methods in enzymology*, 530, 75–87. <https://doi.org/10.1016/B978-0-12-420037-1.00003-8>.
- Hirayama, T. & Shinozaki, K. (2010). Research on plant abiotic stress responses in the post-genome era: past, present and future. *Plant J.* 61:1041-1052.
- Ibarra, L. E., Zamudio, H. F., Pérez, T.C., Albert, V. A., Ramírez, C.E., Molina, T. J., Fernández, C. A., Calderón, V. C., Olivares, J. L., Herrera, E. A., & Herrera, E. L. (2015) *De novo* sequencing and analysis of *Lophophora williamsii* transcriptome, and searching for putative genes involved in mescaline biosynthesis. Ibarra-Laclette et al. *BMC Genomics*. 16:657 DOI 10.1186/s12864-015-1821-9
- Islam, T. & Ghosh, A. (2018). Genome-wide dissection and expression profiling of single glyoxalase III genes in soybean reveal the differential pattern of transcriptional regulation. *Sci Rep-Reino Unido*. 8 (1): 4848.
- Jain, M.; Batth, R.; Kumari, S.; Mustafiz, A. *Arabidopsis thaliana* contains both Ni²⁺ and Zn²⁺ dependent Glyoxalase I enzymes and ectopic expression of the latter contributes more towards abiotic stress tolerance in *E. coli*. *PLoS ONE* 2016, 11, e0159348.
- Jian-Kang Zhu. (2016) Abiotic stress signaling and responses in plants. *Cell*. 167(2): 313–324.2016.08.029. doi:10.1016/j.cell
- Jochen, B. & Wolf, W. (2013). Principles of transcriptome analysis and gene expresión quantification: an RNA-seq tutorial. *Molecular Ecology Resources*. 13, 559–572. doi: 10.1111/1755-0998.12109
- Kang, M. J., Ren Y. Z., Dong, X., Yu, J. W., Hui-Jin F., Bi, Y. H. & Gui, R. Q. (2020). Genome-Wide Identification of the Expansin Gene Family and Its Potential Association with Drought Stress in *Moso Bamboo*. *International Journal of Molecular Sciences*; 21, 9491; d oi:10.3390/ijms21249491.

- TESIS TESIS TESIS TESIS TESIS
- Kaur, C., Ghosh, A., Pareek, A., Sopory, S.K., & Singla-Pareek, S.L. (2014). Glyoxalases and stress tolerance in plants. *Biochem. Soc. Trans.* 42, 485–490. doi: 10.1042/BST20130242
- Kaur, C.; Kushwaha, H.R.; Mustafiz, A.; Pareek, A.; Sopory, S.K.; Singla-Pareek, S.L. (2015). Analysis of global gene expression profile of rice in response to methylglyoxal indicates its possible role as a stress signal molecule. *Front. Plant Sci.* 6, 682.
- Kaur, C.; Sharma, S.; Singla-Pareek, S.L.; Sopory, S.K. (2016). Methylglyoxal detoxification in plants: Role of glyoxalase pathway. *Ind. J. Plant Physiol.* 21, 377–390
- Kaur, C., Sharma, S., Hasan, M. R., Paree, A., Singla-Pareek S. L. & Sopory S. K. (2017). Characteristic variations and similarities in biochemical, molecular, and functional properties of glyoxalases across prokaryotes and eukaryotes. *Int J Mol Sci.* 18 (4): 250.
- Kim, Y. H., Yang, K. S., Ryu, S. H., Kim, K. Y., Song, W. K., Kwon, S. Y., Lee, H. S., Bang, J. W., & Kwak, S. S. (2008). Molecular characterization of a cDNA encoding DRE-binding transcription factor from dehydration-treated fibrous roots of sweetpotato. *Plant physiology and biochemistr.:* PPB, 46(2), 196–204. <https://doi.org/10.1016/j.plaphy>.
- Kwon, K.; Choi, D.; Hyun, J.K.; Jung, H.S.; Baek, K.; Park, C.(2013). Novel glyoxalases from *Arabidopsis thaliana*. *FEBS J.* 280, 3328–3339.
- Lai Z.X. & Lin, Y.L. (2013). Global transcriptome analysis of embryogenic longan callus (*Dimocarpus longan* Lour) using Illumina paired-end sequencing. *BMC genomics* 14: 561.
- Lin, F.; Xu, J.; Shi, J.; Li, H.& Li, B. (2010). Molecular cloning and characterization of a novel glyoxalase I gene TaGLY I in wheat (*Triticum aestivum* L.). *Mol. Biol. Rep.* 37, 729–735.
- Li, B., Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. <https://doi.org/10.1186/1471-2105-12-323>

- Li, T., Cheng, X. & Wang, Y. (2019). Genome-wide analysis of glyoxalase-like gene families in grape (*Vitis vinifera* L.) and their expression profile in response to downy mildew infection BMC Genomics 20, 362. <https://doi.org/10.1186/s12864-019-5733-y>
- Liu, S., Kuang, H. y Lai, Z. (2014). Illumina transcriptome analysis with pairwise high-end sequencing reveals complexity of differential gene expression during in vitro seedling growth and flowering in *Amaranthus tricolor* L. *PLoS one*, 9 (6), e100919. <https://doi.org/10.1371/journal.pone.0100919>.
- Leoncini, G. (1979). Therole of alpha- oxoaldehydes in biological systems. *Ital.J. Biochem.* 28, 285–294.
- Limphong, P.; McKinney, R.; Adams, N.; Makaroff, C.; Bennett, B.; Crowder, M. (2010). The metal ion requirements of *Arabidopsis thaliana* Glx2–2 for catalytic activity. *J. Biol. Inorg. Chem.* 15, 249–258.
- Lockhart, D.; & Winzeler, E: (2000). Genomics, Gene Expression and DNA Arrays. *Nature*. 2000. 415 :827-836.
- Lüttge, U. (2004). Ecophysiology of crassulacean acid metabolism. *Ann. Bot.* 93, 629-652.
- Lustre Sánchez H & Manzanero, M. G I. (2012). Germinación y latencia comparativa de especies del género *Mammillaria* (*Cactaceae*) del Valle de Tehuacán-Cuicatlán, Puebla, México. *Cact Suc Mex* 57:4-15.
- Malhotra, S. & Sowdhamini, R. (2014). Interactions among plant transcription factors regulating expression of stress-responsive genes. *Bioinformatics and Biology Insights*. 8, 193–198.
- Martí, M.C., Camejo, D., Fernandez-García, N., Rellán-Aí lvarez, R., Marques, S., Sevilla, F. & Jiménez, A. (2009). Effect of oil refinery sludges on the growth and antioxidant system of alfalfa plants. *Journal of Hazardous Materials* 172, 879-885.
- Matsukura, S., Mizoi, J. & Yoshida, T. (2010). Comprehensive analysis of rice DREB2-like genes encoding transcription factors involved in the expression of response genes to abiotic stress. *Mol Genet Genomics* 283, 185–196. <https://doi.org/10.1007/s00438-009-0506-y>

- Meza, R., Tafoya, E., Lindig, C. F., Sigala, R.R., & Pérez-Molphe -Balch, E. (2014). Present and potential distribution of the cacti *Ferocactus histrix*, *Mammillaria bombycina* and *M. perezdelarosae* in the state of Aguascalientes, Mexico. *Acta botánica mexicana*, (108), 67-80. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-71512014000300005&lng=es&tlng=en.
- Mochida, K.; K. Shinozaki. (2011). Advances in Omics and Bioinformatics Tools for Systems Analyses of Plant Functions», *Plant and Cell Physiology*. 52, (12): 2017-2038,
- Morales J., Soria R., Alpuche A., Garcidueñas C. (2013). Expresión de una α -expansina implicada en la maduración del fruto de guayaba (*Psidium guajava* L.). Investigación y Ciencia de la Universidad Autónoma de Aguascalientes; 58, 5-11.
- Mustafiz, A.; Singh, A.K.; Pareek, A.; Sopory, S.K.; & Singla-Pareek, S.L. (2011). Genome-wide analysis of rice and Arabidopsis identifies two glyoxalase genes that are highly expressed in abiotic stresses. *Funct. Integr. Genom.* 11, 293–305.
- Mustafiz, A.; Ghosh, A.; Tripathi, A.K.; Kaur, C.; Ganguly, A.K.; Bhavesh, N.S.; Tripathi, J.K.; Pareek, A.; Sopory, S.K.; & Singla-Pareek, S.L. (2014) A unique Ni²⁺-dependent and methylglyoxal-inducible rice glyoxalase I possesses a single active site and functions in abiotic stress response. *Plant J.* 78, 951–963
- Nahar, K., Hasanuzzaman, M., Alam, M.M., & Fujita, M. (2015). Glutathione-induced drought stress tolerance in mung bean coordinated roles of the antioxidant defence and methylglyoxal detoxification systems. *AoB Plants* 7, lv069. doi:10.1093/aobpla/plv069
- Nie, L. & Wang, R. & Xia, Y. & Li, Guojing. (2015). CDPK1, an Arabidopsis thaliana calcium-dependent protein kinase, is involved in plant defense response. *Russian Journal of Plant Physiology*. 62. 866-874. 10.1134/S1021443715070018
- Hernández, M., Narciso, M. Quijada, D. R. & Eiros, J. M. (2019). Application of massive sequencing and bioinformatics to clinical microbiological diagnosis. *RAM-354*; No. of Pages 12. 0325-7541. <https://doi.org/10.1016/j.ram.2019.06.003>
- Méndez Espinoza, C., & Vallejo Reyna, M. Ángel. (2019). Mecanismos de respuesta al estrés abiótico: hacia una perspectiva de las especies forestales. *Revista Mexicana De Ciencias Forestales*, 10(56). <https://doi.org/10.29298/rmcf.v10i56.567>

- Mostofa, M.G., Hossain, M.A., & Fujita, M. (2015). Trehalose pretreatment induces salt tolerance in rice (*Oryzasativa L.*) seedlings: oxidative damage and co-induction of antioxidant defense and glyoxalase systems. *Protoplasma* 252, 461–475. doi:10.1007/s00709-014-0691-3
- Morozovao, Hirstm, Marram. (2009). Applications of NewSequencing Technologies for Transcriptome Analysis. *Annu Rev Genomics Hum Genet.* 10:135-151.
- Marasinghe, G.P.; Sander, I.M.; Bennett, B.; Periyannan, G.; Yang, K.W.; Makaroff, C.A. & Crowder, M.W. (2005). Structural studies on a mitochondrial glyoxalase II. *J. Biol. Chem.* 280, 40668–40675
- Murashige, T. & Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco cultures. *Physiologia Plantarum* 15: 473- 497.
- Pandey, S. K.; Nookarajua, A. Upadhyaya, Ch. P.; Gururani, M. A.; Venkatesh, J.; Kim, D. H. & Park, S. W. (2011). An Update on Biotechnological Aproaches for Improving Abiotic Stress Tolerance in Tomato. *Crop Sci.* vol. 51, pp. 2303-2324
- Pérez-Molphe, B. E. & Dávila, F.C. (2002). In vitro Propagation of *Pelecyphora aselliformis* Ehrenberg and *P. strobiliformis* Werdermann (cactaceae11119. *In vitro Cell. Dev. Biol.- Plant* 38: 73-78.
- Pineda, B.; García-Abellán, J. O. & Antón, T. (2012). Tomato: Genomic Aproaches for Salt and Drought Stress Tolerance. *Improving Crop Resistance to Abiotic Stress*, vol. 1 & 2, pp. 1085-1120
- Proudfoot, N.; A. Furger; & M. Dye: (2002). Integrating mRNA Processing with Transcription, *Cell.* 108, 501-512.
- Quiroz-Castañeda R. & Folch J. (2011). Proteins that remodel and degrade the plant cell wall: current perspectives. *Rev. Biotecnología Aplicada:* vol. 28.
- Rabbani, N., & Thornalley,P.J. (2014). Dicarbonyl proteome and genome damage in metabolic and vascular disease. *Biochem. Soc. Trans.* 42, 425–432. doi: 10.1042/BST20140018

- Rehman, S. & Mahmood, T. (2015). Functional role of transcription factors DREB and ERF: regulation of the stress response network in plants. *Acta Physiol Plant* 37, 178. <https://doi.org/10.1007/s11738-015-1929-1>
- Rezaei, M. A.; Jokar, I.; Ghorbanli, M.; Kaviani, B. & Kharabian-Masouleh, A. (2012). Morpho-physiological improving effects of exogenous glycine betaine on tomato (*Lycopersicon esculentum Mill.*) cv. PS under drought stress conditions. *POJ.*, vol. 5, no. 2, pp. 79-86.
- Rick W Ye, Tao Wang, & Kevin M. C.. (2001). DNA microarray applications in microbial systems. Elsevier. *Revista de métodos microbiológicos*. Volumen 47, Número 3, 257-272 [https://doi.org/10.1016/S0167-7012\(01\)00308-6](https://doi.org/10.1016/S0167-7012(01)00308-6)
- Rivera-Martínez, R., & Ocampo, G. (2018). La familia Cactaceae en el Área Natural Protegida Monumento Natural Cerro del Muerto, Aguascalientes, México. *Investigación y Ciencia de la Universidad Autónoma de Aguascalientes*, 26(75), 28-37.
- Rodríguez-Alonso G. & Shishkova S. (2018). Study of the transcriptome by RNA-seq with emphasis on non-model plant species. *Revista De Educación Bioquímica (REB)* 37(3):75-88, 2018
- Rodríguez-Escriba, R. C. & Rodríguez-Cartaya, D. (2016). Water effect deficit on morpho-physiological and biochemical changes in 'MD-2' micro-propagated pineapple plants at the end of acclimatization stage. *Cultivos Tropicales*. vol. 37, no. especial, pp. 64-73. <http://dx.doi.org/10.13140/RG.2.1.3028.4406>.
- Rubluo A. (1997). Micropropagation of Mammillaria species (Cactaceae). 193-205. En Bajaj YPS (ed). *High-Tech and Micropropagation VI*. Springer-Verlag. Berlin
- Russell, J.J., Theriot, J.A., Sood, P., Marshall, W.F., Landweber, L.F., Fritz-Laylin, L., Polka, J.K., Oliferenko, S., Gerbich, T., Gladfelter, A., Umen, J., Bezanilla, M., Lancaster, M.A., He, S., Gibson, M.C., Goldstein, B., Tanaka, E.M., Hu C.K. & Brunet A (2017) Non-model model organisms. *BMC Biol* 15:55.
- Sampedro, J. & Cosgrove, D.J. (2005). The expansin superfamily. *Genome Biol* 6, 242. <https://doi.org/10.1186/gb-2005-6-12-242>

- Sayadi, A., Immonen, E., Bayram, H. & Arnqvist, G. (2016). The De Novo Transcriptome and Its Functional Annotation in the Seed Beetle *Callosobruchus maculatus*. *PLoS One*. 11(7): e0158565. doi:10.1371/journal.pone.0158565
- Sarah, C. D., June, S., Gil, K. C, Niechayev, N. A., Van-Tongerlo, E., Hurtado-Castano, N., V Dever, L. & Búrquez A. (2019). Underappreciated potential of crassulaceae acid metabolism for current and future agricultural production, *Journal of Experimental Botany*, volumen 70, número 22, páginas 6521–6537, <https://doi.org/10.1093/jxb/erz223>
- Schilling, O.;Wenzel, N.; Naylor, M.; Vogel, A.; Crowder, M.; Makaroff, C.; Meyer-Klaucke,W. (2003). Flexible metal binding of the metallo- β -lactamase domain: Glyoxalase II incorporates iron, manganese, and zinc in vivo. *Biochemistry*, 42, 11777–11786
- Shi, C.Y., Yang, H., Wei, C.L., Yu, O. & Zhang, Z.Z., (2011). Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways for tea-specific compounds. *Genómica de BMC* 12 (1): 131.
- Shifu, C., Yanqing, Z., Yaru, C. & Jia, G. (2018). fastp: an ultrafast all-in-one FASTQ preprocessor, *Bioinformatics*, Volumen 34, Número 17, i884 – i890, <https://doi.org/10.1093/bioinformatics/bty560>
- Shi, C.Y., Yang, H., Wei, C.L., Yu, O., Zhang, Z.Z., Jiang, C.J., Sun, J., Li, Y.Y., Chen, Q. & Xia, T., (2011). Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics*. 12: 131. 10.1186/1471-2164-12-131
- Shishkova, S., Rost T. L. & Dubrovsky J. G. (2008). Determinate root growth and meristem maintenance in angiosperms. *Annals of Botany* 101: 319 – 340.
- Shinozaki, K., & Shinozaki, K. (2006). Transcriptional regulatory networks in cellular responses Yamaguchi and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.*,57, 781-803.
- Singla-Pareek, S.L.; Reddy, M.K. & Sopory, S.K. (2003). Genetic engineering of the glyoxalase pathway in tobacco leads to enhanced salinity tolerance. *Proc. Natl. Acad. Sci. USA*, 100, 14672–14677.

- Soto Sedano, J. C. & López Carrascal, C. E. (2012). RNA-seq: herramienta transcriptómica útil para el estudio de interacciones planta-patógeno. *Fitosanidad* 16(2) 101-113
- Sukdeo, N. & Honek, J.K. (2007). *Pseudomonas aeruginosa* contains multiple glyoxalase I-encoding genes from both metal activation classes. *Biochimica et biophysica acta*, 1774(6), 756–763. <https://doi.org/10.1016/j.bbapap>.
- Subedi, K.P., Choi, D., Kim, I., Min, B. & Park, C. (2011). Hsp31 of *Escherichia coli* K-12 is glyoxalase III. *Mol. Microbiol.* 81, 926–936.
- Strickler, S. R., Bombarely, A., & Mueller, L. A. (2012). Designing a transcriptome next-generation sequencing project for a nonmodel plant species. *American Journal of Botany* 99(2): 257–266.
- Tahsina, S., Hoque, M. A., Hossain, M. G., Mostofa, D. J., Burritt, M. F. and Lam-Son, P. T. (2016). Methylglyoxal: An Emerging Signaling Molecule in Plant Abiotic Stress Responses and Tolerance. doi: 10.3389/fpls.2016.01341
- Thornalley, P.J. (2003) Glyoxalase I-structure, function and a critical role in the enzymatic defence against glycation. *Biochem. Soc. Trans.* 31, 1343–1348
- Tovar O. (2017). Análisis bioquímico y estructural de la proteína scexlx1 de *Schizophyllum commune*. Universidad Autónoma de Nuevo León.
- Umeda, M., Hara, C., Matsubayashi, Y., Li, H. H., Liu, Q., Tadokoro, F. & Uchimiya, H. (1994). Expressed sequence tags from cultured cells of rice (*Oryza sativa* L.) under stressed conditions: analysis of transcripts of genes engaged in ATP-generating pathways. *Plant Molecular Biology*, 25(3), 469–478. <http://www.ncbi.nlm.nih.gov/pubmed/8049371>
- Torres-Silva G. & Correia-Ludmila N. F. (2021). Transcriptome Analysis of *Melocactus glaucescens* (Cactaceae) Reveals Metabolic Changes During *In vitro* Shoot Organogenesis Induction. *Frontiers in Plant Science*. Vol.12. <https://www.frontiersin.org/article/10.3389/fpls.2021.697556>
- Trapnell, C. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.

- Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J. & Geschwind, D.H. (2011) Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Natur.* 474(7351), 380–384. <https://doi.org/10.1038/nature10110>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1), 57–63. <https://doi.org/10.1038/nrg2484>
- Xiong L. & Zhu J.K. (2001). Abiotic stress signal transduction in plants: Molecular and genetic perspectives. *Physiologia Plantarum* 112,152-66.
- Xu, M., Liu, CL., Luo, J. (2019). Análisis transcriptómico de novo de la enfermedad del cancro de pitaya (*Hylocereus polyrhizus*) causada por *Neoscytalidium dimidiatum* . *BMC Genomics*. <https://doi.org/10.1186/s12864-018-5343-0>
- Yan, G.; Lv, X.; Gao, G.; Li, F.; Li, J.; Qiao, J.; Xu, K.; Chen, B.; Wang, L. & Xiao, X. (2016). Identification and characterization of a glyoxalase I gene in a Rapeseed cultivar with seed thermotolerance. *Front. Plant Sci.* 7, 150.
- Yadav, S.K., Singla-Pareek, S.L., Ray, M., Reddy, M.K. & Sopory, S.K., (2005). Methylglyoxal levels in plants under salinity stress are dependent on glyoxalase I and glutathione. *Biochem. Biophys. Res. Commun.*, 337,61–67.
- Zhang, S., Wang, B., Wan, L., & Li, L. M. (2017). Estimating Phred scores of Illumina base calls by logistic regression and sparse modeling. *BMC bioinformatics*, 18(1), 335. <https://doi.org/10.1186/s12859-017-1743-4>
- Zhao, Q., Wang, Y. & Kong, Y. (2011). Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics* 12, S2. <https://doi.org/10.1186/1471-2105-12-S14-S2>
- Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS one*, 9(1), e78644. <https://doi.org/10.1371/journal.pone.0078644>
- Zhu JK. (2002). Salt and drought stress signal transduction in plants. *Ann. Rev. Plant Biol.* 2002; 53:247–273. [PubMed: 12221975].

Zhu., J.-K. (2016). Abiotic stress signaling and responses in plants. *Cell*, 313–324.
doi:10.1016/j.cell.2016.08.029.



11 Anexo

Medio de cultivo Murashige y Skoog (MS):

Tabla 16. Preparación de las soluciones concentradas para el medio MS.

SOLUCIÓN	REACTIVO	CANTIDAD
SOLUCION A Concentración: 1000X. Volumen: 50 mL	Cloruro de calcio (CaCl ₂ -2H ₂ O)	22.0 g
SOLUCION B. Concentración: 1000X. Volumen: 50 mL	Yoduro de potasio (KI) Cloruro de cobalto (CoCl ₂ -6 H ₂ O)	41.50 mg 1.25 mg
SOLUCION C. Concentración: 400 X. Volumen: 50 mL	Fosfato monobásico de K (KH ₂ PO ₄) Ac. bórico (H ₃ BO ₃) Molibdato de sodio (NaMoO ₄)	3.40g 0.124 g 0.005 g
SOLUCION D. Concentración: 400 X. Volumen: 50 mL	Sulfato de magnesio (MgSO ₄ -7H ₂ O) Sulfato de manganeso (MnSO ₄ -H ₂ O) Sulfato de zinc (ZnSO ₄ -7H ₂ O) Sulfato de cobre (CuSO ₄ -5 H ₂ O)	7.40 g 0.340 g 0.172 g 0.50 mg
SOLUCION E. Concentración: 200 X. Volumen: 100 mL	Sulfato ferroso (FeSO ₄ -7H ₂ O) EDTA disódico (Na ₂ EDTA)	0.557 g 0.745 g
SOLUCION F. Concentración 100 X. Volumen: 100 mL	Glicina Piridoxina HCl Ac. nicotínico Tiamina HCl Mio inositol	20.00 mg 5.00 mg 5.00 mg 1.00 mg 1.00 g
Observaciones	<p>En la Solución E Disolver ambos componentes por separado, para lo cual puede requerirse calor. Agregar poco a poco la solución de Fe a la de EDTA y aforar. Debe de quedar de color amarillo sin precipitados.</p> <p>Almacenar estas soluciones en frascos limpios (lavados y enjuagados con agua destilada) y en refrigeración.</p>	

Para preparar 1 litro de medio MS se agregan las siguientes cantidades de las soluciones anteriormente preparadas:

SOLUCIÓN	VOLUMEN MI
A	1
B	1
C	2.5
D	2.5
E	5
F	10

Posteriormente se pesan y añaden 30.0 g de sacarosa, 1.90 g de Nitrato de potasio y 1.65 g de Nitrato de amonio. Se ajusta el pH del medio a 5.7 con NaOH o HCl 0.1 N. Se afora a 1 L con agua destilada y se añaden 8 g/L de Agar como gelificante. El gelificante se disuelve calentando el medio en horno de microondas y agitando posteriormente. El medio se distribuye en frascos de cultivo (dependiendo de su capacidad) se tapan y esterilizan a 121 °C por 20 min.

A1 Estadística básica en transcriptoma

En los datos estadísticos básicos se encuentra la información general de la secuencia obtenida. En la Figura 33 se muestra una comparación de los resultados obtenidos, la cual nos indica que en ambas secuencias el sistema utilizado fue Illumina, se obtuvieron 6,089,759 de lecturas con 0 secuencias de mala calidad, que la longitud de cada una de las secuencias es de 76 bases, la única diferencia que existe es el contenido es el %GC, ya que, como se observa en el inciso A tiene un %49 y en el inciso B 50%.

A	Medida	Valor	B	Medida	Valor
	Nombre del archivo	MB1_S11_R1_001.fastq		Nombre del archivo	MB1_S11_R2_001.fastq
	Tipo de archivo	Llamadas base convencionales		Tipo de archivo	Llamadas base convencionales
	Codificación	Sanger / Illumina 1.9		Codificación	Sanger / Illumina 1.9
	Secuencias totales	6089759		Secuencias totales	6089759
	Secuencias marcadas como de baja calidad.	0 0		Secuencias marcadas como de baja calidad.	0 0
	Longitud de secuencia	76		Longitud de secuencia	76
	% GC	49		% GC	50

Figura 33. Estadística básica. comparación de la secuenciación: A; pertenece a la muestra MB1_S11_R1_001 y B; pertenece a la muestra MB1_S11_R2_001.

ANEXO C. Análisis ContingX

Resultados obtenidos del análisis de ContingX para cada uno de los ensamblados realizados.

Trinity.fasta FastX

#####

Counts of transcripts, etc.

#####

Total trinity 'genes': 45743
Total trinity transcripts: 74777
Percent GC: 44.30

#####

Stats based on ALL transcript contigs:

#####

Contig N10: 3509
Contig N20: 2658
Contig N30: 2158
Contig N40: 1821
Contig N50: 1530

Median contig length: 632
Average contig: 962.33
Total assembled bases: 71960190

#####

Stats based on ONLY LONGEST ISOFORM per 'GENE':

#####

Contig N10: 3290
Contig N20: 2492
Contig N30: 2038
Contig N40: 1701
Contig N50: 1401

Median contig length: 424
Average contig: 793.19
Total assembled bases: 36283076

Trinity.fasta Trimfilter

#####

```
## Counts of transcripts, etc.
#####
Total trinity 'genes': 40884
Total trinity transcripts: 61126
Percent GC: 44.72
```

```
#####
Stats based on ALL transcript contigs:
#####
```

```
Contig N10: 3131
Contig N20: 2400
Contig N30: 1964
Contig N40: 1651
Contig N50: 1380

Median contig length: 584
Average contig: 883.93
Total assembled bases: 54030836
```

```
#####
## Stats based on ONLY LONGEST ISOFORM per 'GENE':
#####
```

```
Contig N10: 2995
Contig N20: 2279
Contig N30: 1864
Contig N40: 1553
Contig N50: 1273

Median contig length: 445
Average contig: 769.36
Total assembled bases: 31454395
```

Tabla 18. Diseño de oligonucleótidos para los genes Glioxalasa GlyI, GlyII, GlyIII y el gen de referencia 25S, se puede observar el nombre del oligonucleótido, la secuencia de la que se compone y la temperatura a la que amplifica.

	Oligonucleótido	Secuencia	TM
GlyI	FMbGlyI.1	CTATTTCTTCGTTTAGGTGC	60
	RMbGlyI.1	GAGAACAACCTGAAACCGAT	
	FMbGlyI.4	TCTAATAAAGGCCAAAGGGG	60
	RMbGlyI.4	CACAATGGCTCAGGTGTAGG	
	FMbGlyI.6	ATCTAATAAAGGCCAAAGGG	57
	RMbGlyI.6	TAACTTGGCACAATGGCTCA	
	FMbGlyI.9	AATGGTTGAAGACATTCGTG	60
	RMbGlyI.9	ATTACTTGGCAGAGTGGTTC	
GlyII	FMbGly2.1	CGTCCTTAACAGATGAGACA	60
	RMbGly2.1	TACTGCGTAAATGAGCTACG	
	FMbGly2.3	AGCCGACTAGAGTTTATTGT	60
	RMbGly2.3	CAATGGTTGAAGGAACAGTT	
GlyIII	FMbGly3.1	GGAGCCGGAGCCAATAGACC	60
	RMbGly3.1	TCACTCAGAATGTTACACGGATGC	
Referencia	F25S	CGTAAGGCGTAAGGAAGCTG	59
	R25S	TCGGAGGGAACCAGCTACTA	

12 Artículo

De Novo Transcriptome of *Mammillaria bombycina* (Cactaceae) under In Vitro Conditions and Identification of Glyoxalase Genes

Plants 2022, 11(3), 399; <https://doi.org/10.3390/plants11030399>



Article

De Novo Transcriptome of *Mammillaria bombycina* (Cactaceae) under In Vitro Conditions and Identification of Glyoxalase Genes

Carolina Enriquez-González ¹, Cristina Garcidueñas-Piña ¹, Osvaldo Adrián Castellanos-Hernández ², Sergio Enriquez-Aranda ³, Abraham Loera-Muro ⁴, Gilberto Ocampo ⁵, Eugenio Pérez-Molphe Balch ¹ and José Francisco Morales-Domínguez ^{1,*}

- ¹ Centro de Ciencias Básicas, Departamento de Química, Universidad Autónoma de Aguascalientes, Av. Universidad 940, Aguascalientes 20100, Mexico; al687@edu.uaa.mx (C.E.-G.); cristina.garcidueñas@edu.uaa.mx (C.G.-P.); eperezmb@gmail.com (E.P.-M.B.)
- ² Centro de Investigación en Biología Molecular Vegetal, Centro Universitario de la Ciénega, Universidad de Guadalajara, Av. Universidad, 1115, Linda Vista, Ocotlán 47810, Mexico; ocnoscr@gmail.com
- ³ Centro de Ciencias Básicas, Departamento de Sistemas de la Información, Universidad Autónoma de Aguascalientes, Av. Universidad 940, Aguascalientes 20100, Mexico; sergio.enriquez@edu.uaa.mx
- ⁴ CONACYT Centro de Investigaciones Biológicas del Noroeste S.C., Av. Instituto Politécnico Nacional 195, Playa Palo de Santa Rita Sur, La Paz 23096, Mexico; aloera@cibnor.mx
- ⁵ Centro de Ciencias Básicas, Departamento de Biología, Universidad Autónoma de Aguascalientes, Av. Universidad 940, Aguascalientes 20100, Mexico; gilberto.ocampo@edu.uaa.mx
- * Correspondence: francisco.morales@edu.uaa.mx



Citation: Enriquez-González, C.; Garcidueñas-Piña, C.; Castellanos-Hernández, O.A.; Enriquez-Aranda, S.; Loera-Muro, A.; Ocampo, G.; Pérez-Molphe Balch, E.; Morales-Domínguez, J.F. De Novo Transcriptome of *Mammillaria bombycina* (Cactaceae) under In Vitro Conditions and Identification of Glyoxalase Genes. *Plants* **2022**, *11*, 399. <https://doi.org/10.3390/plants11030399>

Academic Editor: Thomas Nussbaumer

Received: 12 November 2021
Accepted: 18 January 2022
Published: 31 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: *Mammillaria bombycina* is a cactus distributed in the central region of Mexico. Cactaceae have the particularity of surviving drought and high temperatures, which is why in vitro propagation studies have been carried out successfully to preserve this species and use it as a study model in cacti. In this contribution, a de novo transcriptome of *M. bombycina* was produced under in vitro conditions for the identification and expression of genes related to abiotic stress. Samples were sequenced using an Illumina platform, averaging 24 million clean readings. From assembly and annotation, 84,975 transcripts were generated, 55% of which were unigenes. Among these, the presence of 13 isoforms of genes belonging to glyoxalase I, II and III were identified. An analysis of the qRT-PCR expression of these genes was performed under in vitro and ex vitro conditions and dehydration at 6 and 24 h. The highest expression was observed under greenhouse conditions and dehydration at 24 h, according to the control. The de novo assembly of the *M. bombycina* transcriptome remains a study model for future work in cacti.

Keywords: Cactaceae; Glyoxalases; D-lactate; methylglyoxal; RNA-Seq

1. Introduction

Mammillaria bombycina is located in the mountainous area between the states of Jalisco and Aguascalientes, Mexico in the subtropical scrub, and in the low deciduous forest of this region [1]. It is generally small, globose, with abundant thorns and very colorful flowers (Figure 1), which makes it highly attractive [2]. Overexploitation, overgrazing, and climate change have considerably reduced its spread and natural distribution [3], which is why *M. bombycina* has been classified as endangered by Mexican authorities in NOM-059.-SEMARNAT-2010 [4]. As measures to preserve this species, a large number of in vitro propagation protocols have been developed (Figure 1a) as well as a germplasm bank [5]. *M. bombycina* seedlings grown in vitro can adapt to an ex vitro medium, and their cultures can be maintained under greenhouse conditions (Figure 1b), enabling studies of the identification and expression of genes related to abiotic stress to be developed [6] and research related to its metabolism to continue.

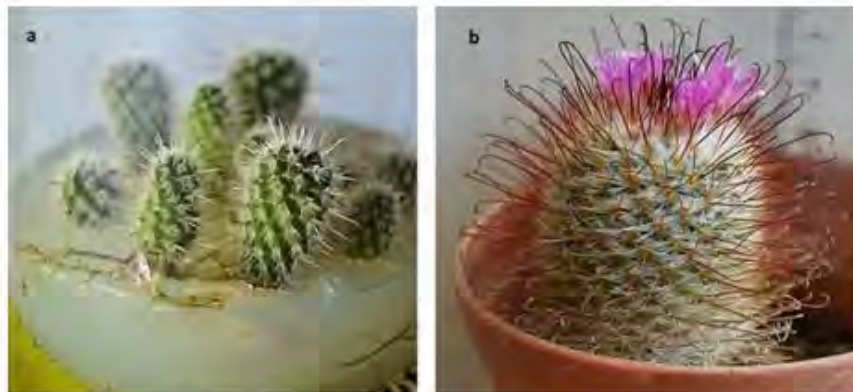


Figure 1. *Mammillaria bombycina*. (a) specimens grown in vitro for 18 months and (b) ex vitro in a greenhouse.

M. bombycina molecular analyses can provide information on the mechanisms of adaptation of the species to extreme conditions. In other plant species, proteins, transcription factors, osmoprotectors, and secondary metabolites have been identified as important components in abiotic stress response [7,8]. Under these conditions, the overproduction of glyoxalases has been detected [9]; these enzymes are responsible for metabolizing methylglyoxal (MG) to D-lactate [10]. MG is a component that is synthesized from metabolic intermediates of photosynthesis, glycolysis, protein metabolism, and lipid peroxidation [11]. However, when plants are under some types of stress, large amounts of MG form, which is toxic to cells due to its ability to act as a glycation agent and increase reactive oxygen species (ROS) [9,12,13]. There are two systems of glyoxalases: glyoxalase I (GLYI; lactoylglutathione lyase) and glyoxalase II (GLYII; hydroxyacylglutathione hydrolase) that, together, catalyze the isomerization of MG to D-lactate, using reduced glutathione (GSH) as a cofactor. On the other hand, glyoxalase III (GLYIII) converts MG to D-lactate in a single step, without the need to reduce glutathione or requiring the intervention of a cofactor or metal ions for its activity [14].

The application of next-generation sequencing (NGS) technology has brought substantial advances in the genomic research of many species. This technology has also been applied for more than a decade to the analysis of transcriptomes for the discovery and analysis of genes [15]. This ability has made it possible to determine and understand the expression patterns of genes in response to different stress conditions. When information on reference genomes or transcriptomes is lacking in the expression analysis of non-model plants, it is highly important to analyze the de novo transcriptome because, by means of bioinformatics tools, it is possible to reach the assembly of short sequences and obtain complete information on the expressed genes [16,17]. Several de novo transcriptomes have been reported from non-model organisms, particularly from succulent plants; for example, in *Lophophora williamsii* for the search for genes involved in the synthesis of mescaline [18], in *Agave sisalana* for the study of genes under drought stress [16], in *Agave H11648* in the identification of cellulose synthase genes [19], in *Hylocereus undatus* for the analysis of floral induction [20], in *Agave deserti* and *Agave tequilana* in the search for genes related to drought tolerance [21], and in *Pachycereus pringlei* for the regulation of root apical meristems [22].

In this study, we report the de novo transcriptome and identification of genes expressed under in vitro conditions in *M. bombycina*. Using the transcriptome assembly, genes were compared with different de novo transcriptomes. In addition, genes coding for glyoxalases were identified. A detailed bioinformatics study was conducted as well as an analysis of expression in vitro, ex vitro, and in dehydration conditions at 6 and 24 h. This is the first

transcriptome under in vitro conditions of *M. bombycina* to be studied and could, therefore, expand our knowledge to identify and analyze an extensive number of genes in this or other species of cacti under different conditions.

2. Results

2.1. Illumina Sequencing, Trimming, and Filtering of the Readings Obtained

From the total RNA of the aerial part of *M. bombycina* plants in vitro, the libraries were constructed for each of the three biological repeats for their sequencing by Illumina. From these libraries, approximately 24 million readings were obtained (Table S2). The quality and integrity of the raw reads of the sequences were determined using the FastQC tool and showed a quality score >35 Phred (average score in all sequences). Trimming and filtering of the readings were performed with the Fastp program, with which adapter remnants were eliminated, and automatic PloyG trimming was carried out (Table S3). The information discarded from each of the sequences was highly accurate because approximately 90% of the readings were retained and used in the assembly (Table S4).

2.2. Assembling the De Novo Transcriptome with the Trinity Program

Filtered reads were concatenated on paired endpoints (in single files 'left.fq' and 'right.fq', respectively) so that they could be digitally standardized and assembled de novo. This approach resulted in a total of 79,881 Trinity-generated transcripts, with a total of 78,412,335 assembled bases. Of the total transcripts, 47,406 correspond to unique genes. Based on the lengths of the assembled contigs of the transcriptome, the standard Nx length statistic was calculated. In this way, it was determined that 50% of the nucleotides of the assembly are found in contigs that are at least 1574 bases long, with an average of 981 bases per contig. As for the longest or most representative isoforms per gene, there are a total of 37,675,612 assembled bases, of which 50% of the nucleotides are in contig that are at least 1417 bases long, with an average of 794 bases (Table 1).

Table 1. Statistical analysis of *M. bombycina* transcriptome assembly evaluation.

Counts of Transcripts, etc.	
Total trinity 'genes'	47,406
Total trinity transcripts	79,881
Percent GC	44.22
Stats based on ALL transcript contigs	
Contig N10	3539
Contig N20	2742
Contig N30	2227
Contig N40	1870
Contig N50	1574
Median contig length	641
Average contig	981.61
Total assembled bases	78,412,335
Stats based on ONLY LONGEST ISOFORM per 'GENE'	
Contig N10	3337
Contig N20	2547
Contig N30	2077
Contig N40	1730
Contig N50	1417
Median contig length	418
Average contig	794.74
Total assembled bases	37,675,612

2.3. Annotation of *M. bombycina* Transcriptome

Table 2 describes the results obtained from the functional annotation of the de novo transcriptome of *M. bombycina*, where the statistical summary generated from the Trinotate.xls database is observed. A total of 84,975 transcripts were obtained, 55% of which are unique (47,406 unigenes). In the results obtained for Blastx, a total of 43,640 hits of nucleotide sequences similar to others already annotated in the aforementioned databases were generated, of which 38,175 are unigenes. In the case of Blastp, a total of 34,192 hits were obtained for putative peptide sequences that are similar to other proteins, whereas 28,201 sequences are unique. A total of 44,629 of transmembrane protein sequences were identified, of which 1729 belong to unique sequences.

Table 2. Statistical analysis of *M. bombycina* transcriptome annotations.

	Unique	Total
gene_id	47,406	84,975
transcript_id	79,881	84,975
prot_id	44,629	44,629
prot_coords	32,539	44,629
TmHMM	1729	44,629
sprot_Top_BLASTX_hit	38,175	43,640
gene_ontology_BLASTX	9638	42,361
Kegg	11,468	39,241
sprot_Top_BLASTP_hit	28,201	34,192
gene_ontology_BLASTP	8947	33,218
Pfam	25,403	30,825
gene_ontology_Pfam	1547	19,238
SignalP	1284	2115
egglog	205	988

Regarding the results obtained in the KEGG platform, 49,241 codes were assigned to metabolic routes in this database, of which 11,468 belong to a unique route.

2.4. EGGNOG Function Classification and Ontological Genes

The description and functional classification of orthologous proteins was carried out in the EGGNOG database. Nine hundred and eighty-eight orthologous protein sequences were assigned from other species already annotated, such as that of *A. thaliana*. These sequences were classified into 17 different functional groups, among which the category translation, ribosomal structure, and biogenesis identified with section J. stand out, with 85% of the total hits (Figure 2). In contrast, the defense mechanisms category (V) contains the lowest number of sequences expressed with 5%.

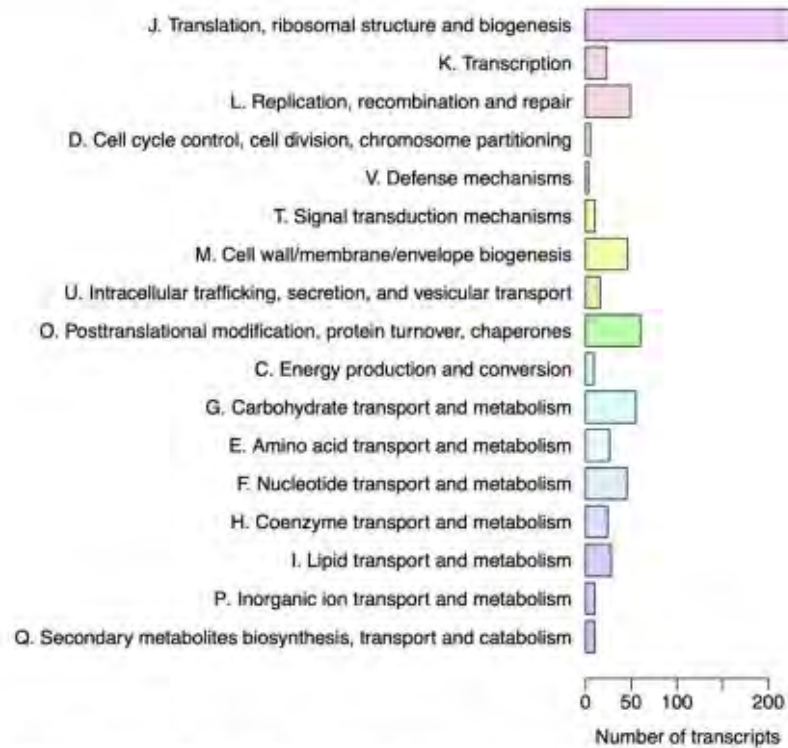


Figure 2. Functional classification of orthologous proteins obtained from the *M. bombycina* transcriptome. The figure shows that the highest number of hits was for translation, ribosomal structure, and biogenesis, whereas the lowest number of hits was obtained for the defense mechanism.

Regarding the classification of gene ontology (GO), a total of 43,640 transcripts were assigned, distributed in 99 different functional groups, and divided into three main categories: (a) cellular component, having around 60% of the total expressed gene sequences, of which, the highest number of hits correspond to genes related to nuclear function; (b) molecular function, corresponding to 30% of the transcripts, most of which are related to ATP binding, and (c) biological processes, which account for approximately 10% of the total hits; among them, the category with the highest number of Gene sequences are those related to the function of the regulation of transcription (Figure 3). GO terms were inferred using the annotation of genes that had blast matches with proteins belonging to the UniprotKB database.

For the distribution of hits in the Nr database (Figure S1), it was found that 90% of the functionally annotated transcripts coincide with *A. thaliana*. The second taxon most closely related to the annotations obtained was *Oryza sativa*, followed by other species that were in the repository. The remaining hits for which a match or a specific hit was not found may be the result of contaminations or genes that have not been well characterized.

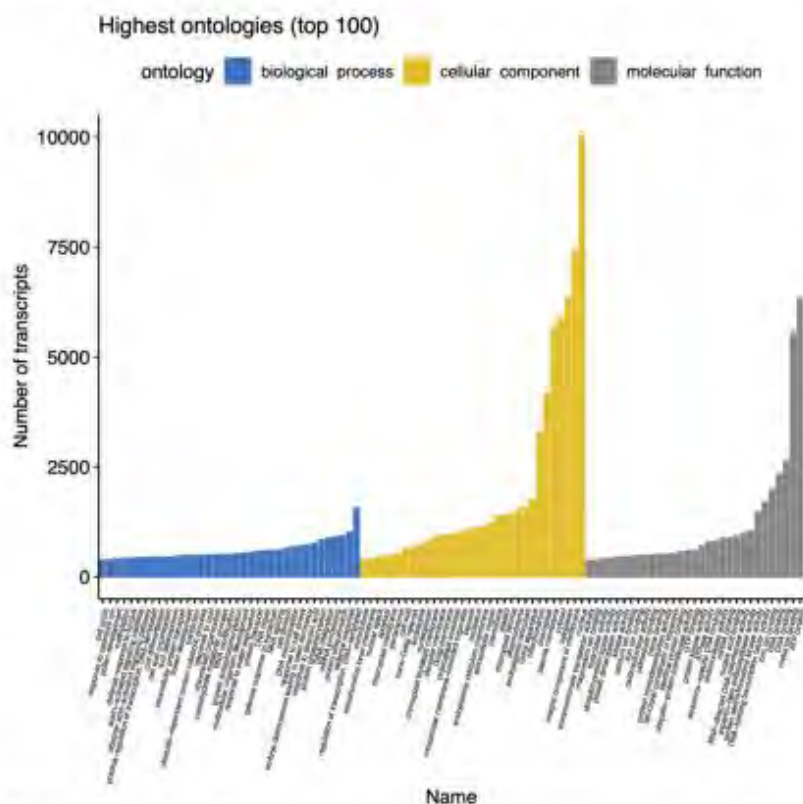


Figure 3. Classification of GO terms obtained from the *M. bombycina* transcriptome. In the biological process classification, the lowest number of hits was found in the cell cycle category, whereas more hits were found for the defense response. In the cell component classification, the lowest was found in the category of microtubule-related transcripts, and in a higher proportion were transcripts associated with nuclear location. Finally, in the classification of molecular function are transcripts related to membrane transport activity, and at a higher proportion are those transcripts related to ATP binding.

2.5. Identification and Analysis of GLYI, GLYII, and GLYIII Genes

All members of the GLYI, GLYII, and GLYIII family of the *M. bombycina* transcriptome were analyzed in detail. Eight sequences similar to the GLYI protein family were identified; however, the putative protein sequence showed that only eight of these contain the Glyoxalase/Bleomycin-resistant domain (PF00903). These sequences were named and classified as MbGLYI-1 to MbGLYI-8 (Table 3). The size of the CDS for the GLYI was varied and with a length ranging from 564 bp (MbGLYI-3) to 1092 bp (MbGLYI-4), with an average size of 858 bp. In the same way, the size of the largest protein was 363 aa for MbGLYI-3, weighing 40.21 kDa, and the shortest was for MbGLYI-2, with 187 aa and 20.88 kDa. Most of these proteins (MgGLYI-2 to MgGLYI-8) have an acidic isoelectric point (pI), and only MgGLYI-1 is basic at pI 8.8 (Table 3). Subcellular location was also analyzed, found in chloroplasts, mitochondria, cytoplasm, and nucleus for MbGLYI-1. The putative protein sequence of MbGLYI-1 was aligned with the homologs of *A. thaliana* and *O. sativa* because these have been reported with nuclear activity and localization, having a protective activity for DNA when MG is present in the cell (Figure S2).

Table 3. Physicochemical characterization and cellular localization of MbGLYI, MbGLYII, and MbDJ-1 putative genes found in the *M. bombycina* transcriptome.

Name Gene	Transcript	Transcript Length (pb)	CDS (pb)	Protein			Localization
				Putative Protein	MW (kDa)	pI	
MbGLYI-1	TRINITY_DN14362_c0_g1_i4	1161	708	235	26.35703	8.82	Chlo ^{ac} , nucl ^b
MbGLYI-2	TRINITY_DN14362_c0_g1_i7	1023	564	187	20.88262	5.43	Chlo
MbGLYI-3	TRINITY_DN14567_c0_g1_i2	1701	1092	363	40.21782	6.64	Chlo ^{ac} , mit ^b
MbGLYI-4	TRINITY_DN14567_c0_g1_i5	2457	792	263	29.53397	5.13	Nucl ^a , cyt ^b
MbGLYI-5	TRINITY_DN14567_c0_g2_i1	1400	1071	356	39.31778	5.80	Chlo ^{abc}
MbGLYI-6	TRINITY_DN15304_c0_g1_i1	1219	870	289	32.68124	5.32	Nucl ^a Nucl ^c
MbGLYI-7	TRINITY_DN15304_c0_g1_i2	1024	903	300	33.65052	5.70	Cyto ^{ac}
MbGLYI-8	TRINITY_DN15304_c0_g2_i1	1386	870	289	32.43995	5.26	Cyto ^{ac}
MbGLYII_1	TRINITY_DN9859_c0_g1_i1	1112	NA	334	37.19777	8.72	Chlo ^a , mito ^c
MbGLYII_2	TRINITY_DN9859_c0_g1_i2	1073	NA	321	35.72691	8.45	Chlo ^{ab} , mito ^c
MbGLYII_3	TRINITY_DN14689_c0_g1_i3	1306	NA	348	38.42322	7.30	Chlo ^a , cyto ^b Nucl ^c
MbGLYII_4	TRINITY_DN15347_c0_g1_i3	1079	NA	290	32.34300	6.55	Cyto ^a , mit ^c
MbDJ-1	TRINITY_DN10140_c0_g1_i1	1617	1176	392	41.974.14	5.57	Cyto ^{ac}

Abbreviations: CDS coding DNA sequence, kDa kilo Daltons, pI isoelectric point, Chlo chloroplast, Cyto cytoplasm, NA does not apply. Prediction of the location: ^a: Wolf Psort, ^b: Localizer and ^c: CELLO. pIs and molecular weights were obtained by using the ProtParam program.

For the GLY II, four partial coding genes were obtained because these were missing parts towards the 5' end. The putative partial sequences of aa contain the glyoxalase domains. The GLYII has an average mass of 35.91 kDa; MbGLYII-1 and MbGLYII-2 have a basic pI. The putative MbGLYII-3 protein has a slightly neutral pI, whereas the MbGLYII-4 pI is acidic. Its subcellular location is cytoplasmic, chloroplast, mitochondrial, and MbGLYII-3 is possibly nuclear because it presents a signal sequence in the cNLSMapper program [23] (Table 3). Regarding GLYIII (MbDJ-1), only one coding gene was found with a CDS of 1176 bp and a sequence of 392 aa, an acid pI of 5.5, and it is a cytoplasmic protein (Table 3).

2.6. Multiple Alignment and Domain Architecture

All putative proteins for glyoxalases from the *M. bombycina* transcriptome were analyzed to identify conserved domains. Figure S3 shows the alignment of all putative proteins and of other species already characterized; similarly, they all present the conserved domains and motifs and their catalytic sites. The putative proteins MbGLYI-1 and MbGLYI-2 present the domain of glyoxalase/bleomycin resistance protein/dioxygenase superfamily (PF00903), which is present in the dependent Zn²⁺; in the shaded boxes are the active aa and, in the star, the active site. For MbGLYI-3 to MbGLYI-8, the same domain (blue arrow) is present twice, making them Ni²⁺ dependent (Figure S3a); in the shaded boxes are the active aa and, in the star, the active site.

In the case of GLYII, although they are partial sequences, it was found that they all present the two domains of GLYII: (1) Metallo-beta-lactamase (PF00753) towards the terminal carboxyl (pink arrow) and (2) Hydroxyacylglutathione hydrolase C-terminal (PF16123) towards the amino terminal (green arrow); in addition, they contain the metal-binding sites (diamonds) and the conserved region THXHXDH/H/D/H, an active site (C/GHT), and seven conserved GSH-binding sites: (C/K(R)/F(Y)/Y/N/R/K) (Figure S3b).

In the case of GLYIII, only one sequence was found. The alignment with ATYL5S and AtDJ-1D shows that MbGLYIII-1 contains twice the DJ-1 domain (yellow arrows) that is characteristic of GLYIII and also presents the motifs conserved within these proteins (Figure S3c).

2.7. Hypothetical Modeling of the MbGLY Family

To determine the folding of the composition of the aa of the MbGLY family, a homology model of the putative proteins of MbGLYI-1, MbGLYI-2, MbGLYII-1, and MbDJ-1 was constructed using the AlphaFold Colab program. Figure 4a shows the hypothetical structure of MbGLYI-1 with the glyoxalase/lactoylglutathione lyase domain for Ni^{2+} -dependent MbGLYI-2 (Figure 4b). In the case of MbGLYII-1, the structure of the protein belongs to the family of metallo- β -lactamases (Figure 4c). In Figure 4d, the hypothetical modeling of MbDJ-1 is shown. We identified glyoxalases proteins from other plant species using the AlphaFold program and Uniprot Database (Figure 4A–D); (A) modeling of the Zn^{2+} -dependent Lactoylglutathione lyase from *Brassica juncea*, (B) Lactoylglutathione lyase-dependent Ni^{2+} from *Zea Mays*, (C) modeling of the GLX2-4 from *Arabidopsis thaliana*, and (D) DJI-D from *Arabidopsis thaliana*. In each model is shown in navy blue the highest degree of confidence (>90).

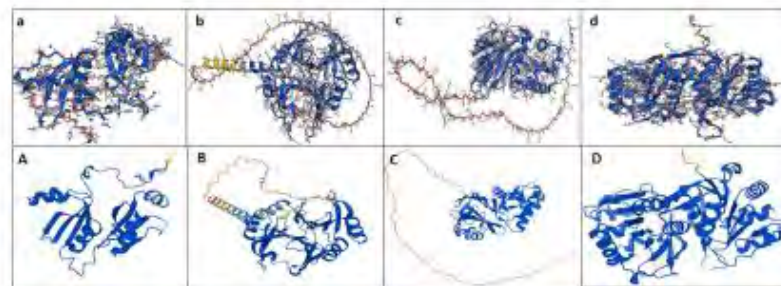


Figure 4. Comparison of the hypothetical modeling of the 3D structures of the putative proteins of MbGLYI, MbGLYII and MbGLYIII of *M. bombycina* with the proteins Glyoxalase I, Glyoxalase II and DJI of the AlphaFold platform. (a) MbGLYI-1 shows a Zn^{2+} dependent region. (A) Modeling of the Zn^{2+} -dependent Lactoylglutathione lyase protein from *Brassica juncea* (O04885 LGUL_BRAJU). (b) Ni^{2+} dependent MbGLYI-3. (B) Lactoylglutathione lyase-dependent Ni^{2+} protein from *Zea Mays* (A0A1D6IYD6_MAIZE). (c) MbGLYII-1. (C) Modeling of the GLX2-4 protein from *Arabidopsis thaliana* (Q8LDW8, GLO2D_ARATH). (d) MbDJ-1. (D) *Arabidopsis thaliana* DJI-D protein (Q9M8R4, DJ1D_ARATH).

2.8. Phylogenetic Analysis of Glyoxalase from *M. bombycina* and Other Plants

Phylogenetic analysis for all members of MbGLY and other species shows three main groups (Figure 5). In the CI group (orange) are all the GLYI, and this group is divided into two subgroups, in the first one are those dependent on Zn^{2+} (MbGLYI-1 and MbGLYI-2), and they are closely related to OsGLYI-8, AtGLYI-2, AtGLYI-12, VvGLYI-2, and GmGLYI-16 (Figure 5). In subgroup 2 of the CI group, the Ni^{2+} -dependent GLYI are found, which in turn are distributed in two groups. In the first group are MbGLYI-3, MbGLYI-4, and MbGLYI-5, very close to those of VvGlyI-1 and VvGlyI-4 *Vitis vinifera*. In contrast, MbGLYI-6, MbGLYI-7, and MbGLYI-8 are closely related to GmGLYI-10 and GmGLYI-21 of glycine max. In group CII (yellow), members of the MbGLYII family are seen. In subgroup 1 are MbGLYII-1, MbGLYII-2, and MbGLYII-4, which are closely related to AtGLYII-4 and VvGLYII-1, whereas in subgroup 2, MbGLYII-3 is related to VvGLYII-2. In group CIII (green) are the GLYIII, and here is located MbDJ-1, which has a close relationship with AtDJ1D of *A. thaliana*.

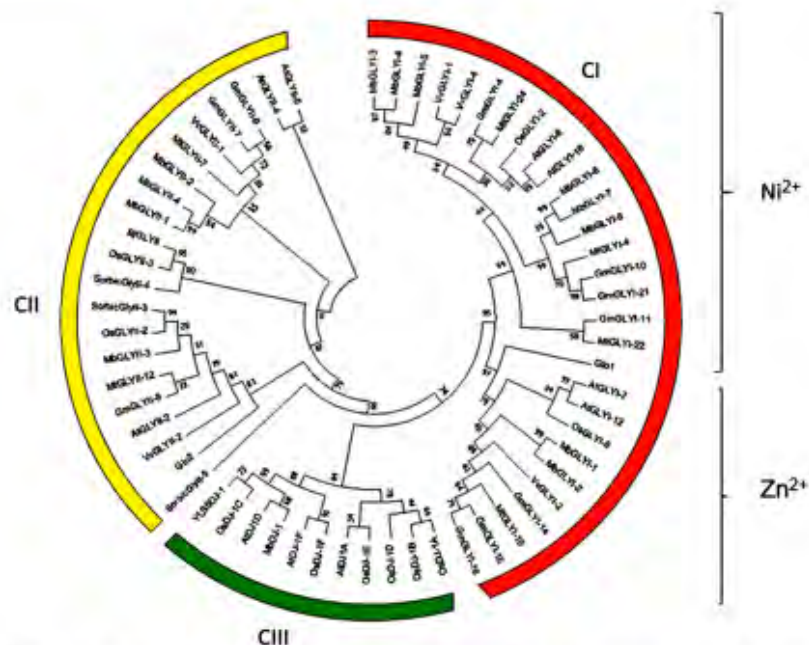


Figure 5. Phylogenetic analysis of the putative proteins MbGLYI, MbGLYII, and MbDJ-1. The phylogeny shows three groups; in CI (orange) are the putative protein sequences belonging to MbGLYI, which in turn are divided into Ni²⁺- and Zn²⁺-dependent. The CII group (yellow) contains the putative protein sequences of MbGLYII, whereas the putative protein sequence MbDJ-1 is located in group CIII (green).

2.9. In Silico Analysis of Co-Expression and Expression Analysis (qRT-PCR) of MbGLYI, MbGLYII, and MbDJ-1 Genes

The IPP results were obtained from the orthologs for At1g8110 (MbGLYI-1), At1g67280 (MbGLYI-2), Glx2-5 (MbGLYII-1), Gly2 (MbGLYII-4), and DJ1D (MbDJ-1) (Figure 6). The five query proteins of *M. bombycina* interact with five others: 4 glyoxalases and a glutamine amido-transferase-like enzyme (YLS5). The analysis of the IPP in the STRING database shows that there are 11 nodes connected with 30 different edges, and the average node degree was 5.45, which means that at least one node has 5.45 interacting nodes. The IPP network shows that all proteins interact with each other (except for DGK1 and DGK2 which are the negative controls). The closest interaction between each protein is based on its score; thus, At1g8110 (MbGLYI-1) has an interaction with GLY2 with a score of 0.993. At1g67280 (MbGLYI-2) interacts with GLY2-1, GLY2-4, and GLY2-5, all with a score of 0.992. GLY2-5 interacts with At1g8110 (MbGLYI-1) with a score of 0.992. GLY2 interacts with At1g8110 (MbGLYI-1) with a score of 0.993. GLY-2 (MbGLYII-4) and DJ1-D (MbDJ-1) interact with At1g8110 (MbGLYI-1) with a score of 0.965. Two sequences of the *M. bombycina* transcriptome encoding a two isoforms of Diacylglycerol kinase 2 was used as a negative control (DGK1 and DGK2).

To analyze real-time expression, the plants generated in vitro were used as a control. The expression of MbGLYI-1 was 1.8-fold higher due to dehydration at 24 h, and in the other treatments, there were no significant changes compared to the control (Figure 7). For MbGLYI-3, 4, and 6, the expression levels decreased drastically compared to the control. With regard to MbGLYII-1, a slight increase in expression was observed in plants ex vitro and by dehydration at 24 h. For MbDJ-1, it was observed that the expression decreased by

drying at 6 h compared to the control, and in the other conditions, there were no statistically significant changes in expression (Figure 7).

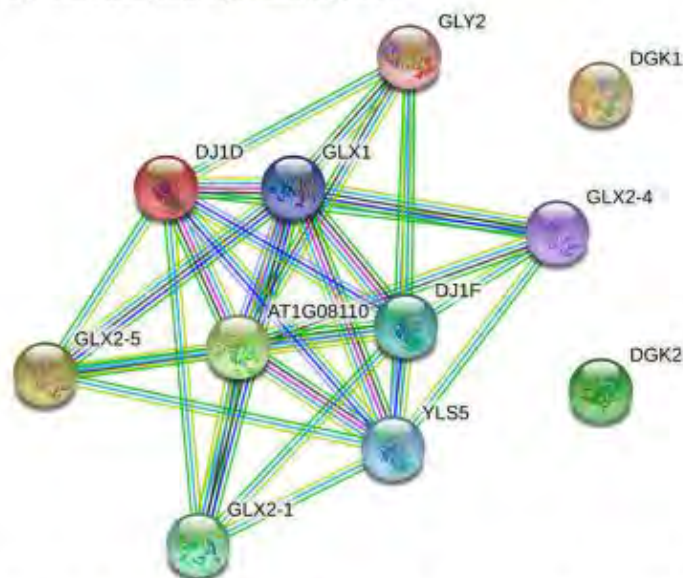


Figure 6. Map of protein–protein interactions for MbGLYI-1, MbGLYI-2, GLX2-5 MbGLYII-1, MbGLYII-4, and MbDJ1. The IPP network shows that all proteins interact with each other.

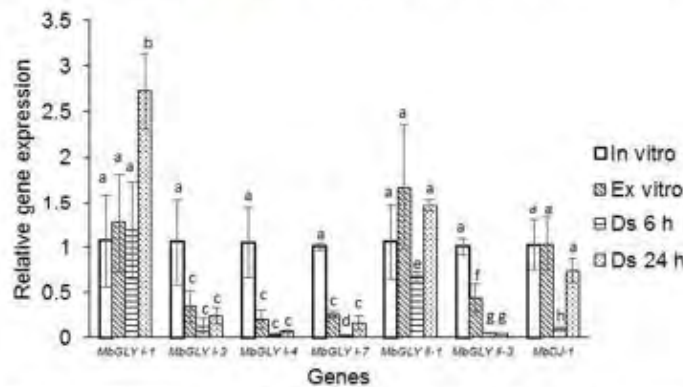


Figure 7. Expression analysis of MbGLYI, MbGLYII, and MbDJ-1 genes. The MbGLYI-1 gene shows a higher expression in the 24 h dehydration treatment, whereas MbGLYI-3, MbGLYI-4, and MbGLYI-7 expression decreased in all treatments compared to the control. For MbGLYII-1, a higher expression was presented in the ex vitro treatments and dehydration 24 h compared to the control. The expression in MbGLYII-3 is significantly lower in all treatments with respect to the control. In the case of MbDJ-1, the expression was maintained for the ex vitro treatments and dehydration at 24 h compared to the control; however, in the dehydration treatment 6 h, there was a significant decrease in the expression compared to the control. Different letters above the bars (a–h) represent a significant difference ($p = 0.05$) in Tukey test.

3. Discussion

Mammillaria bombycina belongs to the Cactaceae family. It is subject to special protection because it has been looted from its natural habitat to be used as an ornamental plant. However, the optimal conditions for its cultivation in vitro have been established, which has facilitated the study of some genes related to abiotic stress response [6]. The use of this species under in vitro conditions for its study at the gene level is of substantial relevance for obtaining genes of interest, as well as the knowledge of their regulatory and functional pathways.

In this work, the annotations of the de novo transcriptome of *M. bombycina* were made under in vitro conditions because there was no previous analysis of new-generation sequencing in this species. Therefore, *M. bombycina* is proposed as a model plant for study in cacti.

A total of 24 million clean reads were obtained, of which 84,975 transcripts found in public databases were noted, as shown in Table 2. Fifty-five percent are unique (47,406 unigenes), and most of these come from *A. thaliana* (Figure 4). Similar results were obtained from *Dimocarpus longan* [24] where, out of 13 samples, the authors generated 64,876,258 clean readings; however, they only obtained 68,925 unigenes. Similarly, in *Agave sisalana* [16], 276.8 million readings were obtained from 9 samples, recording 93,141 genes, of which 67,328 are unigenes. Therefore, our *M. bombycina* de novo transcriptome annotation is a robust and good-quality analysis.

Various conditions such as salinity, drought, high temperature, and exposure to heavy metals are responsible for producing abiotic stress and causing considerable losses in plant productivity [25]. Under stress conditions, plants trigger defense mechanisms based on the synthesis of different enzymes, among which glyoxalases stand out. Glyoxalases are responsible for biotransforming MG, a by-product of metabolism, which under normal conditions is produced in small amounts, whereas under some types of stress, its concentration increases up to six times, causing toxicity to the cell [13,26]. Thus, genetic studies of glyoxalases have been carried out in plants such as *Arabidopsis* [27], rice [28], sorghum [11], and soybeans [14]. However, until now, there have been no reports of *Gly* genes in cacti, despite their importance in adaptation to extreme climates. Hence, in this work, a search for GLY genes in the *M. bombycina* transcriptome was carried out.

In this study, 13 glyoxalase sequences were found: eight from GLYI (MbGLYI), four partial sequences from GLYII (MbGLYII), and one from GLYIII or DJ-1 (MbDJ-1). Bioinformatic analysis and cluster analysis showed that, for MbGLYI, there are two Zn²⁺-dependent genes and six Ni²⁺-dependent genes. *O. sativa*, *S. bicolor* and *A. thaliana* plants have also been reported to contain more Ni²⁺-dependent *GLYI* genes [11]. Lee et al. [29] argues that the higher number of Ni²⁺-dependent genes in plants is due to two evolutionary phenomena—gene duplication and divergence. The partial sequences of MbGLYII contain all the characteristics of glyoxalases II, and the putative protein sequence of MbDJ-1 showed all the characteristics of being a protein with functional activity, as described by Quigley et al. [30]. The cluster analysis showed that the analyzed sequences are closely related to glyoxalases that are expressed under different types of stress [11,14,28]. The putative protein sequence of MbGLYI-1 was found to contain a nuclear signal, indicating that it has the same function as the *A. thaliana* (AtGLYI-2) and *O. sativa* (OsGLYI-8) homologs because they have been reported with nuclear activity and localization [28]. With these results, a hypothetical model of the action of *M. bombycina* glyoxalases was proposed (Figure 8).

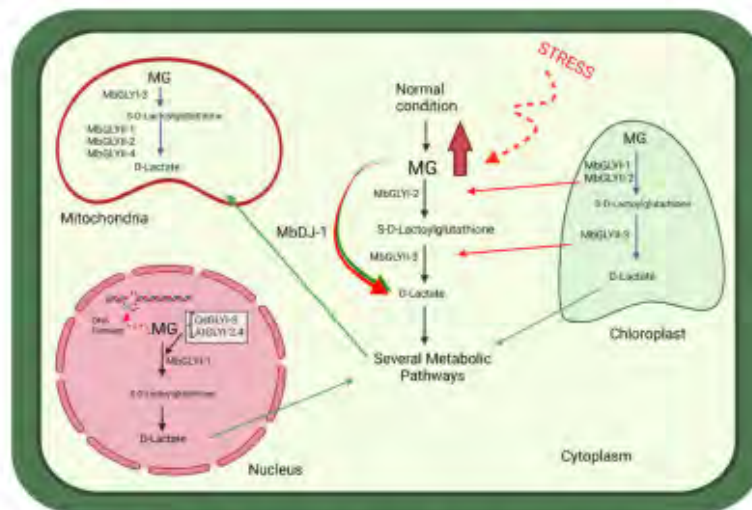


Figure 8. Hypothetical model proposed for the detoxification of MG through the glyoxalase pathway in various subcellular organelles in *M. bombycina*. The role hypothetically played by each of the putative proteins MbGLYI, MbGLYII, and MbDJ-1 in vegetal cells is observed.

The hypothetical modeling of the 3D structures of the putative proteins of MbGLYI, MbGLYII and MbGLYIII of *M. bombycina* show a very similar structure with other glyoxalases models existing in the AlphaFold database (Figure 4). The confidence score in AlphaFold is produced by a per-residue confidence score (pLDDT) with values between 0 and 100. These values are shown in the modelling with different color; navy blue > 90, yellow < 70 and orange < 50 (this region may be unstructured in isolation). AlphaFold protein structure database provides a computational method to predict protein structures with atomic accuracy. This method has the advantage of predicting structures when a model structure is not known, it also shows great accuracy with models reported with other laboratory methods [31,32].

The IPP map shows that all MbGLY interact with each other and with other glyoxalases of the *Arabidopsis* genome and that all the proteins involved are related to abiotic stress resistance, including YLS5 [33]. The expression analysis shows that the MbGLY genes have a similar pattern of expression to those reported in other plant species [11,14,34]. However, it has also been observed that the expression of GLYs genes responds differently depending on the plant and type of stress [27]. With these results, a line of research is opened regarding the behavior of this gene family in *M. bombycina* under in vitro, ex vitro, and dehydration conditions.

4. Materials and Methods

4.1. Vegetal Material

M. bombycina seedlings were obtained from the in vitro Germplasm Bank of the Plant Biotechnology Unit of the Universidad Autónoma de Aguascalientes (Aguascalientes, México). The vegetal material was propagated on MS culture medium [35] with 0.5 m/L of BA (Benzyladenine) [6].

4.2. RNA Extraction and Sequencing by the Illumina Method

Total RNA was extracted from the aerial parts of *M. bombycina*. The tissue was immediately cut and pulverized with liquid nitrogen. Two g of powdered tissue were used, and RNA was extracted using the commercial Trizol kit (Invitrogen) following the

manufacturer's recommendations. RNA purity was verified using an Agilent Bioanalyzer electropherogram (Agilent Technologies, Santa Clara, CA, USA). The construction of the libraries was done with an Illumina TruSeq Stranded mRNA Sample Preparation Kit. Sequencing was carried out with NextSeq 500 equipment from the Illumina company, using a NextSeq 500/550 v2.5 High Output Kit, at the Biotechnology Institute of UNAM at the Unidad Universitaria de Secuenciación Masiva de Datos y Bioinformática (UUSMB) (Cuernavaca, Morelos, Mexico).

4.3. De Novo Transcriptome Assembly and *M. bombycina* Functional Annotation

The FastQC tool (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed on 14 April 2021)) was used to evaluate the quality of raw data from high-throughput sequencing, whereas trimming and filtering of the readings were performed using the Fastp program (<http://opengene.org/fastp/fastp> (accessed on 15 April 2021)). Digital normalization and de novo transcriptome assembly were carried out using Trinity 2.0.5 software with default values [36]. The annotation of the putative genes assembled with the Trinity program was performed using a BlastT analysis in several databases, including NCBI (not redundant; Nr) (<ftp://ftp.ncbi.nih.gov/blast/db/> (accessed on 17 April 2021)), the Swissprot-Uniprot database [37], and the Kyoto encyclopedia of genes and genomes (KEGG) [38], with an E cut-off value set at 10⁻⁵. Analysis of the open reading frames (ORFs) of genes was done with the program Transdecoder v.2.0.1 (<https://github.com/TransDecoder> (accessed on 17 April 2021)), with a cut-off value E of 10⁻⁵ (<http://transdecoder.sourceforge.net/> (accessed on 17 April 2021)). The remaining functional annotation was carried out with Trinotate (<https://trinotate.github.io/> (accessed on 18 April 2021)), through various programs such as Hmmer v.3.1b1 [39], Tmhm v.2 [40], SignalP v.4.1 [41], GOseq [42], and EGGNOG v.3.0. [43]. The summary of results was made from the trinotate.xls database from R version 3.5.0 (<https://cran.r-project.org/bin/windows/base/old/3.5.0/> (accessed on 21 April 2021)).

4.4. In Silico Analysis of Glyoxalase Genes (GLY)

The identification of the putative gene sequences of GLYI, GLYII, and GLYIII was obtained from the *M. bombycina* transcriptome annotations and confirmed by Blastx in the NCBI database (<https://www.ncbi.nlm.nih.gov/> (accessed on 22 April 2021)). The identification of the domains was carried out in the Pfam database (<http://pfam.xfam.org/> (accessed on 22 April 2021)) and Prosite (<https://prosite.expasy.org> (accessed on 22 April 2021)).

The physicochemical analysis of the putative amino acid (aa) sequences was performed with the ProtParam tool [<https://web.expasy.org/protparam> (accessed on 4 May 2021)] [44]. The subcellular localization was predicted with the WolfPSORT programs [<https://www.genscript.com/wolf-psort.html> (accessed on 4 May 2021)] [45], CELLO [<http://cello.life.nctu.edu.tw/> (accessed on 4 May 2021)] [46], and LOCALIZER [<http://localizer.csiro.au/> (accessed on 4 May 2021)] [47]. The signal sequence was predicted using the cNLSMapper program [http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi (accessed on 5 May 2021)] [23].

Multiple sequence alignment was carried out using ClustalW [48] and visualized with Jalview [49].

The construction of phylogenetic trees without roots was carried out using the MEGA version 7 program with the UPGMA method and 1000 bootstrap replicas [50]. GLY sequences from other organisms containing the N-terminal lactoylglutathione lyase domain were analyzed, such as *Vitis vinifera* (VvGLY), *Oryza sativa* (OsGLY), *Arabidopsis thaliana* (AtGLY), *Glicine max* (GmGLY), *Medicago truncatula* (MtGLY), *Brassica juncea* (BjGLY), and human (GloI). The same procedure was performed for GLYII proteins with the metallo- β -lactamase domains and the C-terminal Hydroxyacylglutathione hydrolase domain; *Vitis vinifera* (VvGLYII), *Oryza sativa* (OsGLYII), *Arabidopsis thaliana* (AtGLYII), *Glicine max* (GmGLYII), *Medicago truncatula* (MtGLYII), *Brassica juncea* (BjGLYII), and *Sorghum bicolor*

(SorBiGLYII). For DJ-1 proteins, construction was carried out with GlyIII proteins with the DJ-1 / Pfpl domain; *Homo sapiens* (HsDJ-1), *Oryza sativa* (OsDJ-1), *Arabidopsis thaliana* (AtDJ-1 and YLS5DJ-1).

The prediction of the 3D structure modeling of the putative protein sequences was performed with the program AlphaFold Colab [<https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb> (accessed on 4 January 2022)] (Jumper et al. 2021, Varadi et al. 2021). The protein–protein interaction analysis was carried out using the STRING platform (<https://string-db.org/> (accessed on 26 December 2021)) [51].

4.5. Expression Analysis

From the coding sequences for GLYI, GLYII, and GLYIII in the transcriptome, oligonucleotides were designed (Supplementary Material Table S1). The 25S ribosomal subunit was used as a reference gene with the primers [52].

Seedlings of *M. bombycina* 4 cm tall and one year old were used, testing the following treatments in triplicate: (1) in vitro seedlings (control), (2) ex vitro seedlings kept in a greenhouse, (3) in vitro seedlings dehydrated for 6 h, and (4) in vitro seedlings dehydrated for 24 h. For each treatment, nine seedlings were used in each one of them. Total RNA was extracted using the TRIzol kit (Invitrogen) following the manufacturer's recommendations. The cDNA synthesis was carried out with the Radiant 1-step kit.

Real-time expression analysis was performed using the SYBER Green kit (Applied Biosystem, Carlsbad, CA, USA). The reaction mixture was 5 μ L of SYBER Green reagent, 0.3 μ M of each oligonucleotide, 100 ng of cDNA, and 3 μ L of sterile distilled water for a final reaction of 10 μ L. Statistical analysis was carried out using the $2^{-\Delta\Delta CT}$ method, followed by a one-way ANOVA ($p < 0.05$) and a Tukey test in the Statistic 10.0 program [53].

4.6. Sequence Data

These sequence data have been submitted to the SRA/GenBank databases under accession number: PRJNA764261.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/plants11030399/s1>, Table S1: Primers used in this study, Table S2: Information obtained from the sequencing performed on the Next Seq 500 equipment, in which a 2X75 cycle configuration was used, Table S3: Comparative table of the results obtained from the sequencer and those obtained after trimming and filtering sequences with the program, FastP, Table S4: Table of the percentages retained from clipping sequences with the FastP program, Figure S1: Distribution of hits in the *M. bombycina* transcriptome Nr. database, Figure S2: Alignment of MbGLYI-1, AtGLYI-2, and OsGLYI-8, Figure S3: Multiple alignment and architecture of the MbGLYI, MbGLYII, and MbdJ-1 domains.

Author Contributions: Conceptualization, C.E.-G., G.O. and J.F.M.-D.; methodology, C.E.-G., C.G.-P., O.A.C.-H., S.E.-A., A.L.-M., G.O., E.P.-M.B. and J.F.M.-D.; software, C.E.-G., S.E.-A. and O.A.C.-H.; validation, C.E.-G. and A.L.-M.; formal analysis, C.E.-G., C.G.-P. and J.F.M.-D.; investigation, C.E.-G. and J.F.M.-D.; resources, G.O. and J.F.M.-D.; data curation, C.E.-G.; writing—original draft preparation, C.E.-G. and J.F.M.-D.; writing—review and editing, C.E.-G., C.G.-P., O.A.C.-H., S.E.-A., A.L.-M., G.O., E.P.-M.B. and J.F.M.-D.; visualization, J.F.M.-D.; supervision, J.F.M.-D.; project administration, J.F.M.-D.; funding acquisition, G.O. and J.F.M.-D. All authors have read and agreed to the published version of the manuscript.

Funding: This work obtained financial support from projects CB-2015-01-25604 of CONACYT and PIB17-7N of the Universidad Autónoma de Aguascalientes.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful to CONACYT for the scholarship with the number 766920.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meza-Rangel, E.; Tafuya, F.; Lindig-Cisneros, R.; Sigala-Rodríguez, J.J.; Pérez-Molphe-Balch, E. Present and potential distribution of the cacti *Ferocactus histrix*, *Mammillaria bombycina* and *M. perezdelarosa* in the state of Aguascalientes, Mexico. *Acta Botánica Mex.* **2014**, *108*, 67–80. [\[CrossRef\]](#)
2. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Instituto del Medio Ambiente del Estado de Aguascalientes (IMAE), Universidad Autónoma de Aguascalientes (UAA). *La Biodiversidad en Aguascalientes: Estudio de Estado*; Biodiversity Heritage Library: Aguascalientes, México, 2008; 385p, ISBN 978-970-9000-45-0. [\[CrossRef\]](#)
3. Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT). *Norma Oficial Mexicana NOM-059-SEMARNAT-2010, Protección Ambiental—Especies Nativas de México de Flora y Fauna Silvestres—Categorías de Riesgo y Especificaciones Para su Inclusión, Exclusión o Cambio—Lista de Especies en Riesgo*; Diario Oficial de la Federación: Mexico city, México, 2010.
4. Lustre-Sánchez, H.; Manzanero-Medina, G.I. Germinación y latencia comparativa de especies del género *Mammillaria* (Cactaceae) del Valle de Tehuacán-Cuicatlán, Puebla, México. *Cactáceas Suculentas Mex.* **2012**, *57*, 4–15.
5. Pérez-Molphe-Balch, E.; Santos-Díaz, M.D.S.; Ramírez-Malagón, R.; Ochoa-Alejo, N. Tissue culture of ornamental cacti. *Sci. Agric.* **2015**, *72*, 540–561. [\[CrossRef\]](#)
6. Hernández-Camacho, S.; Pérez-Molphe-Balch, E.; Alpuche-Solis, A.G.; Morales-Domínguez, J.F. Identification and evolutionary relationships of partial gene sequences from dehydrin group in three species of cacti. *Phyton-Int. J. Exp. Bot.* **2017**, *86*, 151–162. [\[CrossRef\]](#)
7. Garza-Padrón, R.A.; Verde-Star, M.J.; Morales-Rubio, M.E.; Oranday-Cárdenas, A.; Rivas-Morales, C.; Núñez-González, M.A.; Barrón-González, M.P. Amoebicid and antioxidant activity and phytochemical profile of methanolic extracts of callus and stem of the cactus *Astrophytum myriostigma*. *Polibotánica* **2010**, *30*, 111–121.
8. Loza-Comejo, S.; Aparicio-Fernández, X.; Torres-Bernal, M.G.; de Jesús Velázquez-Placencia, N. Chemical and morphological characterization of *Mammillaria uncinata* (Cactaceae) fruits. *J. Prof. Assoc. Cactus Dev.* **2013**, *15*, 32–41. [\[CrossRef\]](#)
9. Nahar, K.; Hasanuzzaman, M.; Alam, M.M.; Fujita, M. Glutathione-induced drought stress tolerance in mung bean: Coordinated roles of the antioxidant defence and methylglyoxal detoxification systems. *AoB Plants* **2015**, *7*, plv069. [\[CrossRef\]](#)
10. Kaur, C.; Ghosh, A.; Pareek, A.; Sopory, S.K.; Singla-Pareek, S.L. Glyoxalases and stress tolerance in plants. *Biochem. Soc. Trans.* **2014**, *42*, 485–490. [\[CrossRef\]](#)
11. Bhowal, B.; Singla-Pareek, S.L.; Sopory, S.K.; Kaur, C. From methylglyoxal to pyruvate: A genome-wide study for the identification of glyoxalases and D-lactate dehydrogenases in *Sorghum bicolor*. *BMC Genom.* **2020**, *21*, 145. [\[CrossRef\]](#)
12. Yadav, S.K.; Singla-Pareek, S.L.; Ray, M.; Reddy, M.; Sopory, S. Methylglyoxal levels in plants under salinity stress are dependent on glyoxalase I and glutathione. *Biochem. Biophys. Res. Commun.* **2005**, *337*, 61–67. [\[CrossRef\]](#)
13. Thornalley, P. Glyoxalase I—structure, function and a critical role in the enzymatic defence against glycation. *Biochem. Soc. Trans.* **2003**, *31*, 1343–1348. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Ghosh, A.; Kushwaha, H.R.; Hasan, M.R.; Pareek, A.; Sopory, S.K.; Singla-Pareek, S.L. Presence of unique glyoxalase III proteins in plants indicates the existence of shorter route for methylglyoxal detoxification. *Sci. Rep.* **2016**, *6*, 18358. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Hölzer, M. A decade of de novo transcriptome assembly: Are we there yet? *Mol. Ecol. Resour.* **2020**, *21*, 11–13. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Sarwar, M.B.; Ahmad, Z.; Rashid, B.; Hassan, S.; Gregersen, P.L.; Leyva, M.D.L.O.; Nagy, I.; Asp, T.; Husnain, T. De novo assembly of *Agave sisalana* transcriptome in response to drought stress provides insight into the tolerance mechanisms. *Sci. Rep.* **2019**, *9*, 396. [\[CrossRef\]](#)
17. Hölzer, M.; Marz, M. De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *GigaScience* **2019**, *8*, giz039. [\[CrossRef\]](#)
18. Ibarra-Laclette, E.; Zamudio-Hernández, F.; Pérez-Torres, C.A.; Albert, V.A.; Ramírez-Chávez, E.; Molina-Torres, J.; Fernández-Cortés, A.; Calderón-Vázquez, C.; Olivares-Romero, J.L.; Herrera-Estrella, A.; et al. De novo sequencing and analysis of *Lophophora williamsii* transcriptome, and searching for putative genes involved in mescaline biosynthesis. *BMC Genom.* **2015**, *16*, 657. [\[CrossRef\]](#)
19. Huang, X.; Xiao, M.; Xi, J.; He, C.; Zheng, J.; Chen, H.; Gao, J.; Zhang, S.; Wu, W.; Liang, Y.; et al. De Novo Transcriptome Assembly of *Agave H11648* by Illumina Sequencing and Identification of Cellulose Synthase Genes in *Agave* Species. *Genes* **2019**, *10*, 103. [\[CrossRef\]](#)
20. Xiong, R.; Liu, C.; Xu, M.; Wei, S.-S.; Huang, J.-Q.; Tang, H. Transcriptomic analysis of flower induction for long-day pitaya by supplementary lighting in short-day winter season. *BMC Genom.* **2020**, *21*, 329. [\[CrossRef\]](#)
21. Gross, S.M.; Martin, J.A.; Simpson, J.; Abraham-Juarez, M.J.; Wang, Z.; Visel, A. De novo transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genom.* **2013**, *14*, 563. [\[CrossRef\]](#)
22. Rodríguez-Alonso, G.; Matvienko, M.; López-Valle, M.L.; Lázaro-Mixteco, P.E.; Napsuctaly-Mendivil, S.; Dubrovsky, J.G.; Shishkova, S. Transcriptomics insights into the genetic regulation of root apical meristem exhaustion and determinate primary root growth in *Pachycereus pringlei* (Cactaceae). *Sci. Rep.* **2018**, *8*, 8529. [\[CrossRef\]](#)
23. Kosugi, S.; Hasebe, M.; Tomita, M.; Yanagawa, H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10171–10176. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Lai, Z.; Lin, Y. Analysis of the global transcriptome of longan (*Dimocarpus longan* Lour.) embryogenic callus using Illumina paired-end sequencing. *BMC Genom.* **2013**, *14*, 561. [\[CrossRef\]](#) [\[PubMed\]](#)

25. Hirayama, T.; Shinozaki, K. Research on plant abiotic stress responses in the post-genome era: Past, present and future. *Plant J.* **2010**, *61*, 1041–1052. [[CrossRef](#)] [[PubMed](#)]
26. Kaur, C.; Kushwaha, H.R.; Mustafiz, A.; Pareek, A.; Sopory, S.K.; Singla-Pareek, S.L. Analysis of global gene expression profile of rice in response to methylglyoxal indicates its possible role as a stress signal molecule. *Front. Plant Sci.* **2015**, *6*, 682. [[CrossRef](#)]
27. Kwon, K.; Choi, D.; Hyun, J.K.; Jung, H.S.; Baek, K.; Park, C. Novel glyoxalases from *Arabidopsis thaliana*. *FEBS J.* **2013**, *280*, 3328–3339. [[CrossRef](#)]
28. Mustafiz, A.; Singh, A.K.; Pareek, A.; Sopory, S.K.; Singla-Pareek, S.L. Genome-wide analysis of rice and *Arabidopsis* identifies two glyoxalase genes that are highly expressed in abiotic stresses. *Funct. Integr. Genom.* **2011**, *11*, 293–305. [[CrossRef](#)]
29. Lee, T.-H.; Tang, H.; Wang, X.; Paterson, A.H. PGDD: A database of gene and genome duplication in plants. *Nucleic Acids Res.* **2013**, *41*, D1152–D1158. [[CrossRef](#)]
30. Quigley, P.M.; Korotkov, K.; Baneyx, F.; Hol, W.G.J. The 1.6-Å crystal structure of the class of chaperones represented by *Escherichia coli* Hsp31 reveals a putative catalytic triad. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3137–3142. [[CrossRef](#)]
31. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
32. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2021**, *50*, D439–D444. [[CrossRef](#)]
33. Li, T.; Cheng, X.; Wang, Y.; Yin, X.; Li, Z.; Liu, R.; Liu, G.; Wang, Y.; Xu, Y. Genome-wide analysis of glyoxalase-like gene families in grape (*Vitis vinifera* L.) and their expression profiling in response to downy mildew infection. *BMC Genom.* **2019**, *20*, 362. [[CrossRef](#)] [[PubMed](#)]
34. Ghosh, A. Genome-Wide Identification of Glyoxalase Genes in *Medicago truncatula* and Their Expression Profiling in Response to Various Developmental and Environmental Stimuli. *Front. Plant Sci.* **2017**, *8*, 836. [[CrossRef](#)] [[PubMed](#)]
35. Murashige, T.; Skoog, F. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiol. Plant.* **1962**, *15*, 473–497. [[CrossRef](#)]
36. Grabherr, M.; Haas, B.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)]
37. Bienert, S.; Waterhouse, A.; de Beer, T.A.P.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository—New features and functionality. *Nucleic Acids Res.* **2017**, *45*, D313–D319. [[CrossRef](#)] [[PubMed](#)]
38. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2011**, *40*, D109–D114. [[CrossRef](#)]
39. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [[CrossRef](#)]
40. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L.L. Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [[CrossRef](#)]
41. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786. [[CrossRef](#)]
42. Young, M.D.; Wakefield, M.J.; Smyth, G.K.; Oshlack, A. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* **2010**, *11*, R14. [[CrossRef](#)]
43. Powell, S.; Szklarczyk, D.; Trachana, K.; Roth, A.; Kuhn, M.; Muller, J.; Arnold, R.; Rattei, T.; Letunic, I.; Doerks, T.; et al. eggNOG v3.0: Orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* **2011**, *40*, D284–D289. [[CrossRef](#)] [[PubMed](#)]
44. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Springer Protocols Handbooks; Humana Press: Totowa, NJ, USA, 2005; pp. 571–607. [[CrossRef](#)]
45. Horton, P.; Park, K.-J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **2007**, *35*, W585–W587. [[CrossRef](#)] [[PubMed](#)]
46. Yu, C.S.; Chen, Y.C.; Lu, C.H.; Hwang, J.K. Predicción de la localización subcelular de proteínas. *Proteínas Estruct. Función Bioinformática* **2006**, *64*, 643–651. [[CrossRef](#)] [[PubMed](#)]
47. Sperschneider, J.; Catanzariti, A.-M.; DeBoer, K.; Petre, B.; Gardiner, D.M.; Singh, K.B.; Dodds, P.N.; Taylor, J.M. LOCALIZER: Subcellular localization prediction of both plant and effector proteins in the plant cell. *Sci. Rep.* **2017**, *7*, 44598. [[CrossRef](#)] [[PubMed](#)]
48. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigar, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)]
49. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [[CrossRef](#)]
50. Núñez-Colín, C.A.; Escobedo-López, D. Uso correcto del análisis clúster en la caracterización de germoplasma vegetal. *Agronomía. Mesoam.* **2011**, *22*, 415–427. [[CrossRef](#)]

51. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [[CrossRef](#)]
52. Martínez-Vázquez, J.P.; Loera-Muro, A.; Gómez-Aguirre, Y.A.; Morales-Domínguez, J.F. Identification and characterization of the EXPA7, EXPA18 and EXT10 genes in *Turbinicarpus lophophoroides* (Werderm.) Buxb. & Backeb; and their expression analysis in the root under abiotic stress. *Mol. Biol. Rep.* **2021**, *48*, 1633–1644. [[CrossRef](#)]
53. Livak, K.J.; Schmittgen, T.D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)]