



COLEGIO DE POSTGRADUADOS
INSTITUCION DE ENSEÑANZA E INVESTIGACION EN CIENCIAS AGRICOLAS

CAMPUS CAMPECHE

POSTGRADO EN BIOPROSPECCIÓN Y SUSTENTABILIDAD AGRÍCOLA EN EL
TRÓPICO

**IDENTIFICACIÓN Y ANÁLISIS DE LOS GENES
INVOLUCRADOS EN LAS RUTAS BIOSINTÉTICAS DE
CELULOSA Y LIGNINA EN *Agave tequilana* Weber**

LUIS FERNANDO MACEDA LOPEZ

T E S I S

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS

SIHOCHAC, CHAMPOTÓN, CAMPECHE

2020



COLEGIO DE POSTGRADUADOS

INSTITUCION DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

MÉXICO-PUEBLA-SAN LUIS POTOSI-TABASCO-VERACRUZ-CÓRDOBA-CAMPECHE

CARTA DE CONSENTIMIENTO DE USO DE LOS DERECHOS DE AUTOR Y DE LAS REGALIAS COMERCIALES DE PRODUCTOS DE INVESTIGACION

En adición al beneficio ético, moral y académico que he obtenido durante mis estudios en el Colegio de Postgraduados, el(la) que suscribe Luis Fernando Maceda Lopez, Alumno(a) de esta institución, estoy de acuerdo en ser partícipe de las regalías económicas y/o académicas, de procedencia nacional e internacional, que se deriven del trabajo de investigación que realicé en esta institución, bajo la dirección del Profesor Dr. Fulgencio Alatorre Cobos, por lo que otorgo los derechos de autor de mi tesis IDENTIFICACIÓN Y ANÁLISIS DE LOS GENES INVOLUCRADOS EN LAS RUTAS BIOSINTÉTICAS DE CELULOSA Y LIGNINA EN *Agave tequilana* Weber y de los productos de dicha investigación al Colegio de Postgraduados. Las patentes y secretos que se pueden derivar serán registrados a nombre del Colegio de Postgraduados y las regalías económicas que se deriven serán distribuidas entre la Institución, El Consejero o Director de Tesis y El que suscribe, de acuerdo a las negociaciones entre las tres partes, por ello me comprometo a no realizar ninguna acción que dañe el proceso de explotación comercial de dichos productos a favor de esta institución.

Sihochac, Champotón, Campeche a 07 de febrero de 2020.

I.Bt. Luis Fernando Maceda López

Vo.Bo.

Dr. Fulgencio Alatorre Cobos Profesor Consejero

La presente tesis, titulada: Identificación y análisis de los genes involucrados en las rutas biosintéticas de celulosa y lignina en *Agave tequilana* Weber. realizada por el alumno: **Luis Fernando Maceda López**, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS
BIOPROSPECCIÓN Y SUSTENTABILIDAD AGRÍCOLA EN EL
TRÓPICO CONSEJO PARTICULAR

CONSEJERO:


DR. FULGENCIO ALATORRE COBOS

DIRECTORA:


DRA. ELSA BEATRIZ GÓNGORA CASTILLO

ASESORA:


DRA. JUNE KILPATRICK SIMPSON WILLIAMSON

ASESORA:


DRA. ITZEL LÓPEZ ROSAS

Sihochac, Champotón, Campeche a 14 de febrero de 2020.

IDENTIFICACIÓN Y ANÁLISIS DE LOS GENES INVOLUCRADOS EN LAS RUTAS BIOSINTÉTICAS DE CELULOSA Y LIGNINA EN *Agave tequilana* Weber

Luis Fernando Maceda López, M.C.

Colegio de postgraduados, 2020

RESUMEN

Las rutas biosintéticas y de regulación génica para los compuestos estructurales de la pared celular primaria y secundaria han sido principalmente caracterizadas en plantas dicotiledóneas, especialmente para la planta herbácea *Arabidopsis thaliana*. Muy pocos elementos y las interacciones de tales rutas se han reportado para plantas monocotiledóneas. El objetivo de este estudio fue identificar y caracterizar los genes ortólogos involucrados en las rutas de biosíntesis de celulosa y lignina, dos compuestos principales que conforman la pared celular en *Agave tequilana* Weber y *Agave deserti*. Para esto, se utilizó un enfoque de minería de datos de transcriptomas de *A. tequilana* Weber y análisis de expresión para algunos genes seleccionados. Se utilizaron datos previamente publicados y sin reportar (J. Simpson, CINVESTAV Irapuato), de órganos y/o tejidos con pared celular primaria y/o secundaria. La expresión de algunos de los genes seleccionados fue validada experimentalmente. Los resultados de ortología permitieron identificar genes que codifican enzimas de cada uno de los pasos de la biosíntesis de celulosa y lignina. Los análisis filogenéticos y de motivos conservados para genes clave de ambas rutas mostraron un alto grado de conservación evolutiva respecto a otras monocotiledóneas como maíz y arroz, lo cual sugiere además un rol funcional en la conformación de la pared celular en agaves. Los ensayos de expresión tipo RT-PCR punto final mostraron expresión enriquecida en tejidos adyacentes a fibras para genes involucrados en la formación de celulosa y lignina, similar a lo que se ha reportado para otros tejidos de soporte como xilema.

Palabras claves: transcriptoma, genes ortólogos, celulosa, lignina.

ANALYSIS OF GENES INVOLVING IN CELLULOSE AND LIGNIN PATHWAYS IN *Agave tequilana* Weber

Luis Fernando Maceda López, M.C.
Colegio de Postgraduados, 2019

ABSTRACT

The biosynthetic and genetic regulation pathways for the structural compounds of the primary and secondary cell wall have been mainly characterized in dicotyledonous plants, especially for the *Arabidopsis thaliana* herbaceous plant. Very few elements and interactions of such routes have been reported for monocotyledonous plants. The objective of this study was to identify and characterize the orthologous genes involved in the biosynthesis pathways of cellulose and lignin, two main compounds that make up the cell wall in *Agave tequilana* Weber and *Agave deserti*. For this, a mining approach to transcriptome data from *A. tequilana* Weber and expression analysis was used for some selected genes. Previously published and unreported data (J. Simpson, CINVESTAV Irapuato), of organs and / or tissues with primary and / or secondary cell wall were used. The expression of some of the selected genes was experimentally validated. The orthology results allowed to identify genes that encode enzymes of each steps of the biosynthesis of cellulose and lignin. Phylogenetic and conserved motive analyzes for key genes of both routes showed a high degree of evolutionary conservation compared to other monocots such as corn and rice, which also suggests a functional role in the conformation of the cell wall in agaves. End-point RT-PCR type expression assays showed rich expression in fiber-adherent tissues for genes involved in cellulose and lignin formation, similar to what has been reported for other supportive tissues such as xylem.

Keywords: transcriptome, orthologous genes, cellulose, lignin.

DEDICATORIA

Dedico esta tesis a mi familia por su incondicional apoyo, amor, comprensión, voluntad y paciencia.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca concedida (número 485607) y por el financiamiento al proyecto de fronteras de la ciencia otorgado al Dr. Fulgencio Alatorre Cobos y titulado: "Study of molecular mechanisms underlying the biosynthesis and deposition of structural carbohydrates during the cell wall development in Agave sp: implications for the use of waste agave leaves for biofuel production" (clave 1049).

A mi familia por haberme acompañado de manera continua, en las buenas y en las malas hasta el día de hoy, a pesar de la distancia.

Al Colegio de postgraduados Campus Campeche, por todas las facilidades brindadas para realizar mis estudios de posgrado.

De igual forma se agradece al Instituto de Ecología A.C. (INECOL) y al Centro de Investigaciones Científicas de Yucatan (CICY), por todas las facilidades brindadas para el uso del recurso de computación de alto rendimiento, proporcionado para la realización de esta investigación.

Al M. en C. Emanuel Villafán De La Torre y al Dr. Jose Luis Villalpando Aguilar por su apoyo prestado para la realización de este proyecto.

Gracias al comité tutorial, conformado por: el Dr. Fulgencio Alatorre Cobos, la Dra. Elsa Beatriz Gongora Castillo, la Dra. June Kilpatrick Simpson Willianson y la Dra. Itzel Lopez Rosas, por las críticas y recomendaciones dadas.

A mis amigos y compañeros que han hecho de mi estancia en la institución, divertida, agradable y amena.

CONTENIDO

RESUMEN	III
ABSTRACT.....	IV
DEDICATORIA	V
AGRADECIMIENTOS.....	VI
CONTENIDO.....	VII
LISTA DE CUADROS.....	10
LISTA DE FIGURAS.....	12
I. INTRODUCCIÓN.....	14
II. REVISIÓN DE LITERATURA	16
2.1. Biocombustibles	16
2.1.1 Biocombustibles de primera generación	17
2.1.2 Biocombustibles de segunda generación	17
2.1.3 Biocombustibles de tercera generación	18
2.2. Proceso de obtención de los biocombustibles	18
2.3. Bioetanol	19
2.4.1 Agave como fuente de fibras	25
2.5. Pared celular como una fuente de celulosa y lignina	27
2.6 Biosíntesis de celulosa en <i>Arabidopsis thaliana</i> Lynch	29
2.6.1 Genes involucrados en la biosíntesis de celulosa en <i>Arabidopsis thaliana</i> y su función biológica en el desarrollo vegetal.....	30
2.7 Biosíntesis de lignina en <i>Arabidopsis thaliana</i> Lynch	34
2.8. Regulación transcripcional de las rutas de formación de la pared celular en plantas	38
2.8.1. Regulación transcripcional en <i>Arabidopsis thaliana</i>	38
2.8.2. Regulación transcripcional en plantas monocotiledóneas	39
2.9. Importancia de las ciencias ómicas en el estudio de <i>Agave tequilana</i> 42	
2.9.1. Aplicaciones de las ciencias ómicas en <i>Agave spp.</i>	42
2.10. Clasificación de los genes homólogos	43
III. OBJETIVO.....	45

3.1	Objetivo general	45
3.2	Objetivos particulares	45
IV.	Hipótesis	45
4.1	Hipótesis general.....	45
V.	MATERIALES Y MÉTODOS	46
5.1	Metodología general usada en este trabajo	46
5.2.	Descarga de los reads de los transcriptomas de <i>Agave tequilana</i> y <i>Agave deserti</i>	49
5.3	Filtrado y preensamblado de secuencias	50
5.4	Ensamblado de los transcriptomas.....	51
5.5	Identificación de las regiones codificantes	51
5.6	Identificación de genes ortólogos en <i>Agave tequilana</i> y <i>Agave deserti</i>	52
5.7	Análisis de similitud e identidad en los genes ortólogos identificados en <i>A. tequilana</i> y <i>A. deserti</i>	52
5.8	Alineamiento de las secuencias de aminoácidos y análisis filogenético de CESA y CAD en <i>A. tequilana</i> y <i>A. deserti</i>	53
5.9	Generación de la matriz de los perfiles de expresión.....	53
5.10	Análisis de expresión de genes involucrados en pared celular en raíz, piña, meristemo y hoja de <i>A. tequilana</i>	54
5.10.1	Colecta del material	54
5.10.2	Extracción RNA total.....	57
5.10.3	Electroforesis	57
5.11	Análisis de expresión de genes involucrados en pared celular en tejido proximal a fibras.....	59
5.11.1	Colecta del material	59
5.11.2	Extracción de RNA total y electroforesis.....	59
5.12	Diseño de oligonucleótidos para ensayos tipo RT-PCR	61
5.13	Análisis de expresión por RT-PCR punto final.....	62
5.13.1	Síntesis de DNA complementario (cDNA)	62
5.13.2	Reacciones de la Polimerasa en Cadena (PCRs)	63
VI.	RESULTADOS Y DISCUSIÓN.....	64
6.1	Identificación <i>in silico</i> de los genes ortólogos involucrados en las rutas de biosíntesis de celulosa y lignina que están relacionadas con la formación de la pared celular en <i>A. tequilana</i> Weber y <i>A. deserti</i>	64
6.1.1	Ensamblado <i>de novo</i> : Anotación y categorización funcional del	

transcriptoma de <i>Agave tequilana</i> y <i>Agave deserti</i>	64
6.1.2 Genes ortólogos en <i>A. tequilana</i> y <i>A. deserti</i> involucrados en la biosíntesis de celulosa.....	66
6.1.3 Motivos funcionales y estructurales de los genes <i>CESA</i> en <i>A. tequilana</i> Weber y <i>A. deserti</i>	73
6.1.4 Análisis filogenético de los genes <i>CESA</i> de <i>Agave</i>	76
6.1.5 Genes ortólogos en <i>A. tequilana</i> y <i>A. deserti</i> involucrados en la biosíntesis de lignina.....	80
6.1.6 Conservación del gen <i>CAD</i> en <i>Agave</i> respecto a plantas monocotiledóneas y dicotiledóneas	83
6.1.7 Motivos funcionales y estructurales de los genes <i>CAD</i> en <i>A. tequilana</i> Weber y <i>A. deserti</i>	84
6.1.8 Análisis filogenético de los genes <i>CAD</i> de <i>Agave</i>	85
6.2. Análisis del perfil transcriptómico <i>in silico</i> de los genes identificados en los diferentes órganos de <i>A. tequilana</i> Weber y <i>A. deserti</i>.....	88
6.2.2 Perfil transcriptómico de los genes involucrados en la biosíntesis de lignina en <i>A. tequilana</i> Weber y <i>A. deserti</i>	91
6.3. Modificaciones menores al protocolo estándar de la extracción de RNA total	95
6.4 Análisis de expresión de los genes <i>SUS2</i>, <i>CESA1</i>, <i>CESA7</i>, <i>CCR2</i> y <i>CAD5</i> en raíz, piña, meristemo y hoja de <i>A. tequilana</i>.....	99
6.4.1 Identificación de genes endógenos como controles de carga	99
6.4.2 Expresión de <i>SUS2</i> , <i>CESA1</i> y <i>CESA7</i>	99
6.4.3 Expresión de <i>CCR2</i> y <i>CAD5</i>	100
6.5 Análisis de expresión de los genes <i>SUS2</i>, <i>CESA1</i>, <i>CESA7</i>, <i>CCR2</i> y <i>CAD5</i> en tejido adyacente a las fibras	103
VII. CONCLUSIONES Y PERSPECTIVAS.....	105
7.1 Conclusiones.....	105
7.2 Perspectivas.....	106
VIII. LITERATURA CITADA	107
IX. ANEXOS.....	116

LISTA DE CUADROS

Cuadro 1. Comparación del contenido de celulosa y lignina (Elaborada con datos de Hongjia <i>et al.</i> , 2012 y Corbin <i>et al.</i> , 2015).	24
Cuadro 2. Mutantes disponibles y fenotipos para el gen CELULOSA SINTASA (CESA) en <i>Arabidopsis thaliana</i>	31
Cuadro 3. Datos de los transcriptomas utilizados en este estudio (Gross <i>et al.</i> , 2013).	50
Cuadro 4. Características de las plantas de <i>A. tequilana</i> colectadas en el Rancho Agropecuaria Santa Genoveva, Campeche.	55
Cuadro 5. Oligonucleótidos utilizados para los análisis RT-PCR. F: Forward; R: Reverse.	62
Cuadro 6. Comparación del número de genes por familias involucradas en la biosíntesis de celulosa en <i>A. tequilana</i> , <i>A. deserti</i> y en la planta modelo <i>A. thaliana</i>	67
Cuadro 7. Genes ortólogos identificados en la ruta de biosíntesis de celulosa en los transcriptomas <i>A. tequilana</i> y <i>A. deserti</i>	69
Cuadro 8. Homología entre genes <i>CESA</i> identificadas en el análisis de ortología en <i>A. tequilana</i> y <i>A. deserti</i> y los genes <i>CESA</i> de <i>A. tequilana</i> reportados por Huang <i>et al.</i> , 2019. Los valores muestran el porcentaje de identidades (número izquierdo) y similitudes (número derecho) de aminoácidos.	71
Cuadro 9. Secuencias homologas entre las secuencias de aminoácidos ortólogas de AqCESA y AdCESA y los CESAs de otras plantas. Los valores muestran el porcentaje de identidades (número izquierdo) y similitudes (número derecho) de aminoácidos. (Ad) <i>A. deserti</i> , (Aq) <i>A. tequilana</i> , (Ao) <i>A. officinalis</i> , (Zm) <i>Z. mays</i> y (At) <i>A. thaliana</i>	73
Cuadro 10. Genes AqCesA y AdCESA y sus ortólogos de <i>Arabidopsis</i> , trigo, maíz y arroz involucrados en la formación de la pared celular primaria (PCW) y la pared celular secundaria (SCW).	79
Cuadro 11. Comparación del número de genes por familias involucradas en la biosíntesis de lignina en <i>A. tequilana</i> , <i>A. deserti</i> y en la planta modelo <i>A. thaliana</i>	

.....	80
Cuadro 12. Genes ortólogos identificados en la ruta de biosíntesis de lignina en los transcriptomas <i>A. tequilana</i> y <i>A. deserti</i>	82
Cuadro 13. Secuencias homologas entre las secuencias de aminoácidos ortólogas de AqCAD y AdCAD y los CADs de otras plantas. Los valores muestran el porcentaje de identidades (número izquierdo) y similitudes (número derecho) de aminoácidos.	84
Cuadro 14. Rendimiento y pureza del RNA total extraído sin purificar y purificado de raíz de <i>A. tequilana</i>	96
Cuadro 15. Rendimiento y pureza del RNA total extraído de raíz, piña, meristemo y hoja (Análisis 1) y hojas y fibras (Análisis 2) de <i>A. tequilana</i>	97

LISTA DE FIGURAS

Figura 1. Proceso de obtención de biocombustibles (Modificada de Ballesteros, 2019).....	19
Figura 2. Producción mundial de bioetanol en millones de litros durante el año 2018, (Modificado de Riazi <i>et al.</i> ,2017).	20
Figura 3. Tipos de pretratamientos utilizados en la producción de bioetanol (Modificada de Tu <i>et al.</i> , 2019).....	22
Figura 4. Diversidad de Agave en México.	26
Figura 5. Pared celular. (A) Pared celular primaria. (B) Pared celular secundaria. (Modificada de Nakano <i>et al.</i> , 2015).....	28
Figura 6. Biosíntesis de celulosa (Babb <i>et al.</i> , 2001).....	30
Figura 7. Esquema de los distintos fenotipos de AtCESAs en la elongación y división celular, integridad de la pared celular y producción de biomasa vegetal (Modificado de Hu <i>et al.</i> , 2018).	33
Figura 8. Modelos estructurales de los monómeros de lignina.....	34
Figura 9. Biosíntesis de lignina (Sattler <i>et al.</i> , 2013).	37
Figura 10. Ruta de regulación genética para la biosíntesis de los principales componentes de pared celular en <i>Arabidopsis thaliana</i> (A) y monocotiledóneas (B)..	41
Figura 11. Árbol de los genes de insulina ortólogos y parálogos en ratón, rata y humano.	44
Figura 12. Metodología general usada en este trabajo.	48
Figura 13. Colecta del material de <i>A. tequilana</i> para la extracción de RNA total. A) Planta de aproximadamente 5 años de edad. B) Raíces. C) Piña y meristemo. D) Secciones utilizadas de las hojas.....	56
Figura 14. Protocolo de extracción de RNA total de raíces de <i>A. tequilana</i>	58
Figura 15. Colecta del material de <i>A. tequilana</i> para la extracción de RNA total.	60
Figura 16. Mapa de calor de los niveles de expresión de los genes control utilizados para los análisis de RT-PCR punto final.....	61
Figura 17. Relaciones filogenéticas de las especies utilizadas para el proceso de	

anotación funcional en <i>A. tequilana</i> y <i>A. deserti</i>	65
Figura 18. Porcentaje de unigenes de <i>A. tequilana</i> y <i>A. deserti</i> anotados en función de la homología de los mismos con proteínas en el genoma de diferentes especies de plantas angiospermas.	66
Figura 19. Alineamiento de las secuencias de los motivos que diferencian los genes ortólogos para celulosa sintasa en la pared celular primaria (PCW) y la pared celular secundaria (SCW)	75
Figura 20. Árbol filogenético no enraizado de los genes <i>CESA</i> de <i>Agave tequilana</i> (AqCESA), <i>Agave deserti</i> (AdCESA) y de otras especies de plantas en las cuales se han reportado los genes <i>CESA</i>	78
Figura 21. Secuencias de los motivos identificados en los genes ortólogos cinamil alcohol deshidrogenasa (CAD) en (Ad) <i>A. deserti</i> , (Aq) <i>A. tequilana</i> , (Zm) <i>Z. mays</i> y (At) <i>A. thaliana</i>	85
Figura 22. Árbol filogenético no enraizado de los genes CAD de <i>Agave tequilana</i> (AqCAD), <i>Agave deserti</i> (AdCAD), <i>Oryza sativa</i> (OsCAD), <i>Zea mays</i> (ZmCAD) y <i>Arabidopsis thaliana</i> (AtCAD).....	87
Figura 23. Mapa de calor de los niveles de expresión de los genes identificados en la biosíntesis de celulosa en A) <i>A. tequilana</i> y B) <i>A. deserti</i>	90
Figura 24. Mapa de calor de los niveles de expresión de los genes identificados en la biosíntesis de lignina en A) <i>A. tequilana</i> y B) <i>A. deserti</i>	94
Figura 25. Gel Electroforesis de RNA total de raíz roja y raíz blanca de <i>A. tequilana</i> en gel de agarosa al 1%..	96
Figura 26. Gel de agarosa teñido con bromuro de etidio en el que se observan los RNAs totales de raíz, piña, meristemo y hoja (Análisis 1) y hojas y fibras (Análisis 2) de <i>A. tequilana</i>	98
Figura 27. RT-PCR de <i>SUS2</i> , <i>CESA1</i> , <i>CESA7</i> , <i>CCR2</i> y <i>CAD5</i> en raíz, piña meristemo y hoja en <i>A. tequilana</i>	102
Figura 28. RT-PCR de <i>SUS2</i> , <i>CESA1</i> , <i>CESA7</i> , <i>CCR2</i> y <i>CAD5</i> en hoja y fibras de <i>A. tequilana</i>	104

I. INTRODUCCIÓN

Actualmente, la disminución del consumo de los combustibles fósiles como medida para disminuir la contaminación ambiental, ha tomado gran importancia a nivel mundial debido a que se han buscado nuevas alternativas para la producción de energía (Owusu *et al.*, 2016; Ediger *et al.*, 2019). Es por eso que se ha optado por el aprovechamiento de la materia orgánica para la producción de bioenergía (Demirbaş *et al.*, 2005).

Uno de los biocombustibles que ha tomado importancia a nivel mundial es el bioetanol (Proskurina *et al.*, 2019). Este compuesto se obtiene a partir de la materia lignocelulósica presente en diversas plantas como la caña de azúcar y el maíz además de residuos orgánicos maderables (Ummalyma *et al.*, 2019). A pesar de que el bioetanol es renovable, rentable y amigable con el medioambiente (Saini *et al.*, 2015), tiene una gran desventaja que es poco atractiva para la industria y es su alto costo de producción debido a la recalcitrancia entre la celulosa y lignina (Li *et al.*, 2016).

Debido a esto, la industria ha buscado nuevas alternativas de materia lignocelulósica que presenten un alto contenido de celulosa y menor contenido de lignina para solucionar la recalcitrancia de la celulosa (Zhao *et al.*, 2012). Una especie que ha sido atractiva es *A. tequilana* por su bajo contenido de lignina en el bagazo y hojas y además porque son plantas que crecen en suelos con poca disponibilidad de agua (Hongjia *et al.*, 2012; Corbin *et al.*, 2015).

En este trabajo se buscó contribuir al entendimiento de los genes involucrados en la formación de la pared celular primaria y secundaria en *A. tequilana* y *A. deserti*. En la primera parte del trabajo, se identificaron los genes involucrados en la biosíntesis de celulosa y lignina en ambas especies de agave mediante el uso de minería de datos en los transcriptomas disponibles para dichas especies.

La segunda parte, de esta investigación consistió en analizar experimentalmente los niveles de expresión de al menos dos genes involucrados en la biosíntesis de celulosa y lignina en tejidos de raíz, piña, meristemo, hojas y fibras de *A. tequilana*. Los resultados muestran que hay genes de la familia *CESA* y la familia

CAD que son específicos del tipo de pared celular y que además las rutas de biosíntesis de celulosa y lignina están conservadas evolutivamente con otras plantas monocotiledóneas y dicotiledóneas.

II. REVISIÓN DE LITERATURA

El incremento mundial de la demanda de combustibles fósiles acompañada con su agotamiento a lo largo de los últimos años, así como la contaminación que provoca su uso, ha generado la búsqueda de nuevas alternativas para la producción de energía, que provengan de fuentes renovables como la biomasa lignocelulósica (Demirbaş *et al.*, 2005; Hamelinck *et al.*, 2005). Se ha demostrado que la lignocelulosa presenta un gran potencial para la producción de biocombustibles y otros bioproductos como biofibras y enzimas (Millati *et al.*, 2011). Los biocombustibles obtenidos de lignocelulosa son renovables, rentables y amigables con el medioambiente. Sin embargo, su producción se realiza con un alto costo debido a la recalcitrancia entre la celulosa y lignina.

2.1. Biocombustibles

Los biocombustibles son combustibles no fósiles líquidos o gaseosos obtenidos a partir de materia orgánica derivada de algún proceso biológico o industrial. Los biocombustibles se dividen en primarios y secundarios. Los biocombustibles primarios son aquellos que no han sido procesados industrialmente como la leña y son utilizados principalmente para la calefacción o cocina. Los biocombustibles secundarios, por su parte, son aquellos que han sido procesados como el bioetanol, el biodiesel y el biogas. (Demirbas, 2007). Esta clase de combustibles son producto de los avances en genética, biotecnología y en procesos químicos que han permitido convertir la biomasa en biocombustibles (Ragauskas *et al.*, 2006). Dependiendo de las fuentes de materia prima utilizada en los biocombustibles secundarios, se pueden clasificar en combustibles de primera, segunda o tercera generación (Singh *et al.*, 2011).

2.1.1 Biocombustibles de primera generación

Los biocombustibles de primera generación son generados a partir de cultivos que son destinados para el consumo humano, como la caña de azúcar en Brasil, el maíz en Estados Unidos de América y el trigo en Europa. Así mismo, el biodiesel es generado a partir de aceites de plantas, por ejemplo, el aceite de palma utilizado en América central, Malasia, Tailandia y África (Sharma *et al.*, 2017). Estados Unidos de América y Brasil son los países que tienen la mayor producción de bioetanol a nivel mundial, abarcando el 85 % de dicha producción, mientras que el resto del mundo produce solo el 15 %. El amplio uso de granos para bioetanol ha generado algunos conflictos dado que la materia prima está incluida en la cadena alimentaria (Sharma *et al.*, 2017).

2.1.2 Biocombustibles de segunda generación

Por otra parte, los biocombustibles de segunda generación se obtienen de residuos agrícolas y forestales ricos en celulosa y cuyo uso no está destinado para el consumo humano. Según un estudio de la Agencia de Protección Medioambiental de Estados Unidos de América (<https://www.epa.gov/>) muestra que en dicho país se producen 2.18 millones de galones de bioetanol obtenido a partir de residuos agrícolas y forestales. Otro país que produce biocombustibles de segunda generación es China, con el uso de un 66 % de residuos agrícolas y 34 % de residuos forestales, haciendo un total de 12,693 toneladas de biomasa disponibles para la producción de bioenergía (Gao *et al.*, 2016).

No obstante, el uso de este tipo de residuos agrícolas y forestales también ha generado un gran número de obstáculos para las industrias de los biocombustibles; y es debido a la naturaleza cristalina de la celulosa, por el grado de polimerización que presenta, el tamaño de la partícula de la biomasa y la recalcitrancia debido a la cubierta protectora de lignina ocasionando un área superficial menor para la reacción de hidrolisis enzimática (Horn *et al.*, 2012; Zavrel *et al.*, 2009).

2.1.3 Biocombustibles de tercera generación

Finalmente, los combustibles de tercera generación son el bioetanol y biodiesel obtenidos de algas y malezas marinas. Son de bajo costo y provienen de recursos completamente renovables. Los biocombustibles y bioproductos obtenidos de algas presentan un gran potencial para reemplazar los combustibles fósiles, siendo así una fuente de energía prometedora para la producción de energía verde sostenible (Sharma *et al.*, 2017). Se ha estimado que las microalgas cultivadas en fotobiorreactores optimizados pueden producir hasta 57,000 L de aceite por acre por año (Demirbas, 2010). Este tipo de organismos pueden crecer en áreas inadecuadas para los cultivos de primera y segunda generación, por ejemplo, las aguas residuales y el agua salada. Sin embargo, presentan un alto costo de operación y por lo tanto pueden ser poco atractivos para las industrias de los biocombustibles.

2.2. Proceso de obtención de los biocombustibles

Uno de los procesos más simples para la producción de bioetanol es a partir del aprovechamiento de materias primas ricas en sacarosa como la caña de azúcar y el sorgo (Perdices, 2018). La sacarosa se transforma en etanol mediante levaduras fermentativas sin ninguna etapa de hidrólisis previa. Otro proceso de obtención de etanol es a partir de granos de cereales ricos en almidón. El almidón se hidroliza enzimáticamente en glucosa y entonces estos subproductos se fermentan en etanol (Figura 1).

El biodiesel se produce a través de una reacción química llamada transesterificación entre los aceites vegetales, las grasas y el alcohol, todos estos compuestos en presencia de un catalizador ácido, alcalino o enzimático. En el proceso de transesterificación las moléculas de triglicéridos son transformadas en ésteres monoalquílicos (Keera *et al.*, 2011).

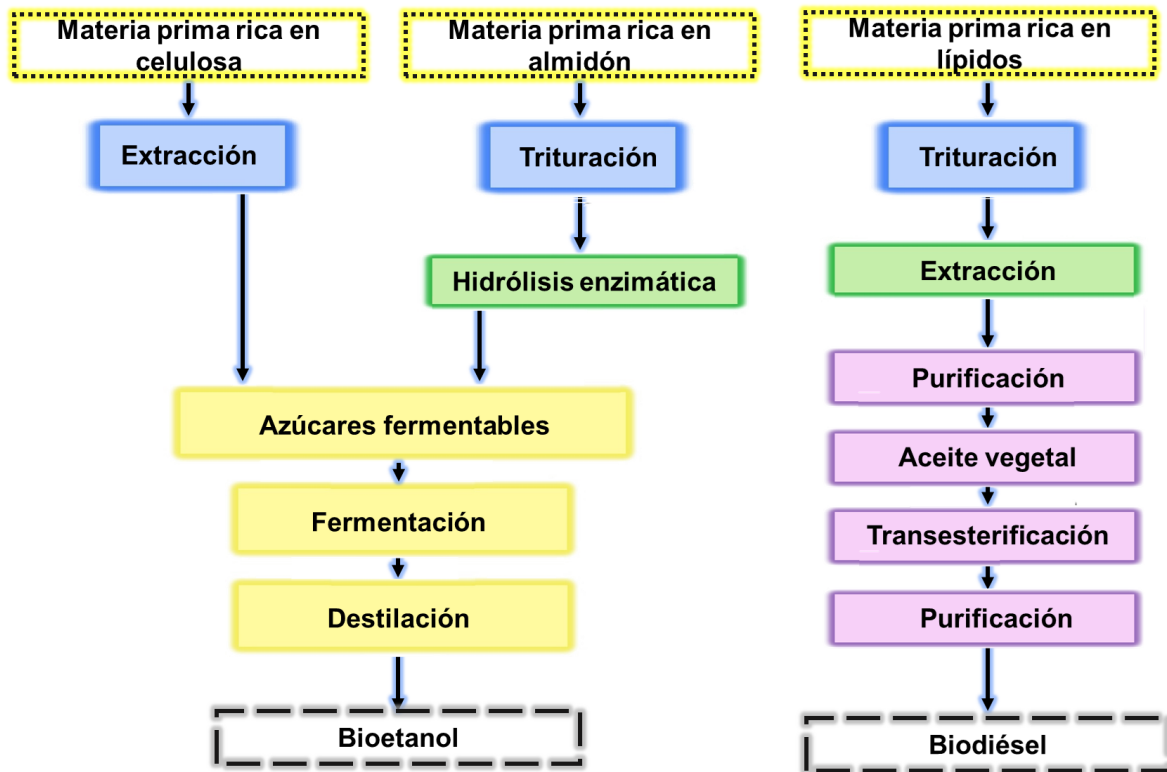


Figura 1. Proceso de obtención de biocombustibles (Modificada de Ballesteros, 2019).

2.3. Bioetanol

El bioetanol es un tipo de alcohol inflamable que se obtiene a partir de la fermentación de los azúcares que están presentes en la materia orgánica, tales como cereales, caña de azúcar y sorgo. Los principales países que lideran actualmente la producción mundial de bioetanol son Estados Unidos y Brasil (Figura 2). Estados Unidos produce bioetanol a partir de el almidón presente en maíz y trigo, mientras que Brasil produce bioetanol a partir de la glucosa contenida en la caña de azúcar (Riazi *et al.*, 2017; Bušić *et al.*, 2018).

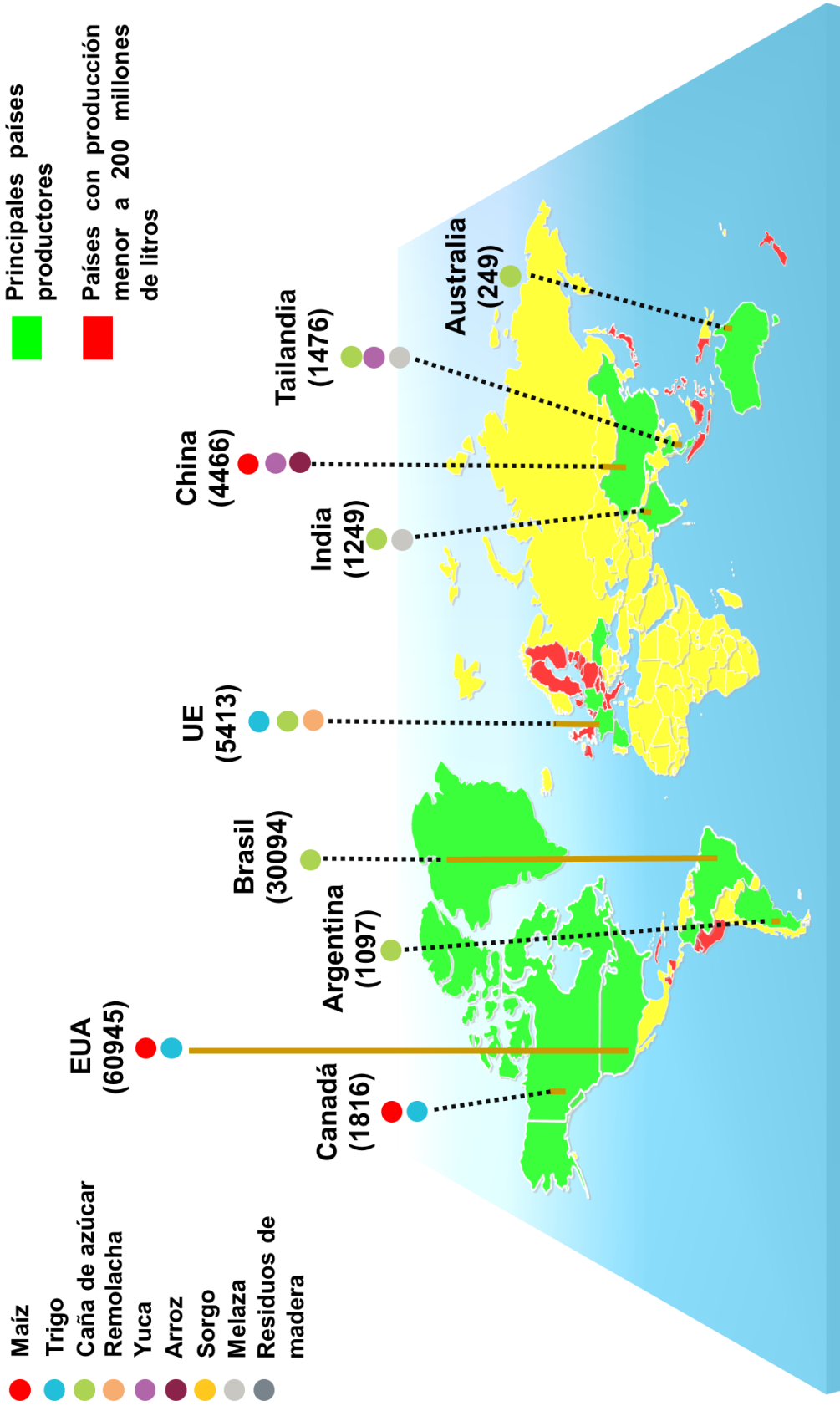


Figura 2. Producción mundial de bioetanol en millones de litros durante el año 2018, (Modificado de Riazi *et al.*, 2017).

Aunque la industria actual produce bioetanol a partir de materia orgánica rica en glucosa y almidón, se ha comenzado a utilizar otros residuos agrícolas que presentan un alto contenido de celulosa. Entre las principales plantas o residuos orgánicos que presentan un alto contenido de celulosa están las especies maderables (Brostow *et al.*, 2010; Wasiak, 2019). La madera está compuesta básicamente de celulosa, el cual es un homopolímero estructural de la pared celular compuesta por moléculas de β -glucosa, formando cadenas lineales de celulosa. Estas cadenas de celulosa se unen fuertemente por puentes de hidrogeno. Así mismo, la celulosa se encuentra intercalada con otros compuestos como la lignina y la hemicelulosa en la pared celular. Todo este conjunto de biomasa lignocelulósica es un sistema recalcitrante debido a la fuerte unión que presentan en la pared celular (Guarnizo *et al.*, 2009).

Dada este hecho de recalcitrancia de la pared celular, la industria que produce bioetanol a partir de celulosa tiene problemas con sus rendimientos de sacarificación. Y es que para romper la recalcitrancia, específicamente durante la fermentación, se requiere a la celulosa libre de lignina y hemicelulosa. Para lograr esto, se han desarrollado una serie de pretratamientos para retirar la lignina y la hemicelulosa de los residuos orgánicos, elevando los costos de producción del bioetanol (Figura 3) (Carrillo *et al.*, 2019; Tu *et al.*, 2019; Mahmood *et al.*, 2019).

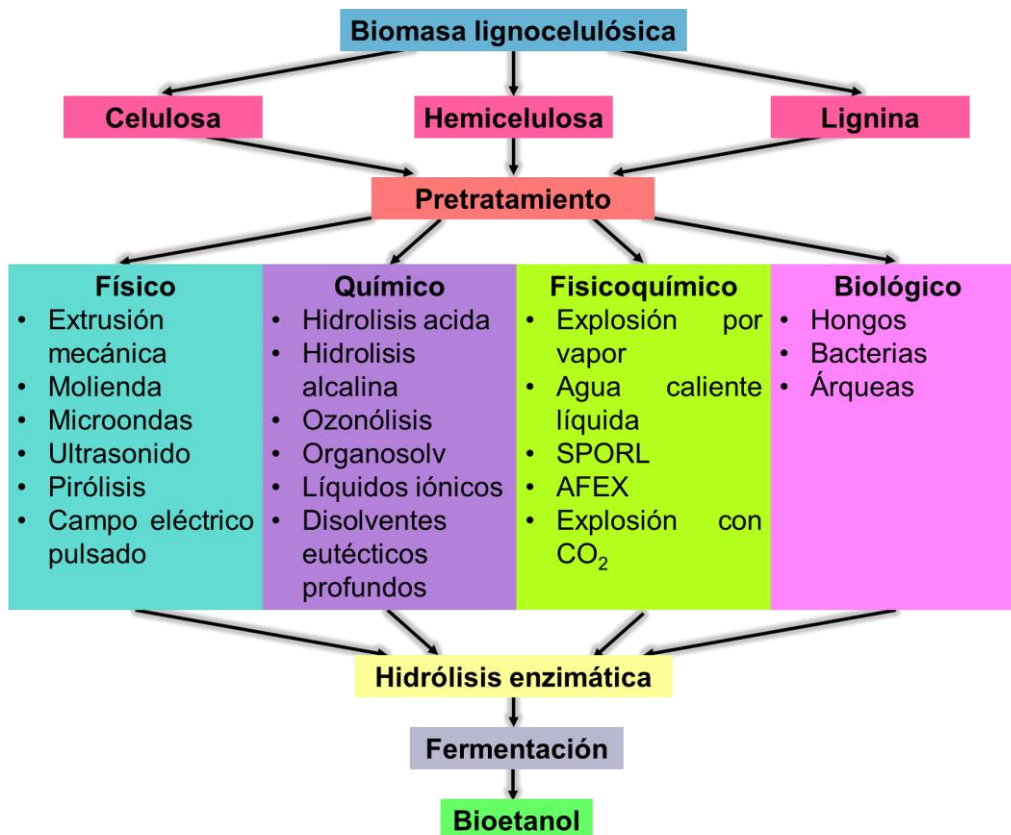


Figura 3. Tipos de pretratamientos utilizados en la producción de bioetanol (Modificada de Tu *et al.*, 2019).

2.4. *Agave tequilana* Weber como fuente potencial de biomasa para la producción de biocombustibles

Con el objetivo de solucionar la recalcitrancia de la celulosa, la industria busca nuevas especies vegetales que presenten un alto contenido de celulosa y menor contenido de lignina. En 2015, Corbin y colaboradores reportaron que el bagazo de caña de azúcar presenta entre un 30 y 45 % de contenido de celulosa y entre un 20 y 30 % en el contenido de lignina mientras que el bagazo de *Agave tequilana* presenta aproximadamente un 43 % de contenido de celulosa y un 15 % de contenido de lignina. Esto es sumamente atractivo para la industria del bioetanol porque representa un 65 % menos contenido de lignina, que puede traducirse en una disminución en los costos de producción (Cuadro 1). A partir de datos previamente publicados, el cuadro 1 muestra que además de *A. tequilana*

existen otras especies de agaves donde el contenido de lignina es de aproximadamente 5 %. Es interesante el caso de agaves fibrosos como *A. lechuguilla* y *A. fourcroydes* donde el contenido de celulosa oscila entre 77 y 79 %, con contenidos de lignina aún por debajo de los cultivos bioenergéticos típicos como maíz y caña de azúcar (Cuadro 1). Junto con su bajo contenido de lignina en el bagazo y hojas, los agaves poseen otras ventajas relacionadas con su cultivo y uso de agua respecto a las especies que tradicionalmente se usan para biocombustibles. En México particularmente, esta planta prospera en más del 75 % del territorio nacional, en zonas con clima árido o semiárido donde la disponibilidad de agua es baja o muy baja durante la mayor parte del año (Figura 3) (García-Mendoza, 2002).

Cuadro 1. Comparación del contenido de celulosa y lignina (Elaborada con datos de Hongjia *et al.*, 2012 y Corbin *et al.*, 2015).

Especies	Tejido	Celulosa (% p/p)	Lignina (% p/p)	Referencia
<i>Zea mays</i>	Rastrojo sin mazorcas	31-38	17-21	Corbin <i>et al.</i> , 2015
<i>Triticum aestivum</i>	Planta completa	33	17	Corbin <i>et al.</i> , 2015
<i>Sorghum bicolor</i>	Planta completa	23	11	Corbin <i>et al.</i> , 2015
<i>Saccharum</i> spp.	Bagazo	32-43	23-28	Corbin <i>et al.</i> , 2015
<i>Panicum virgatum</i>	Planta completa	30-35	17-20	Corbin <i>et al.</i> , 2015
<i>Populus</i> spp.	Árbol completo sin hojas	41-43	24-28	Corbin <i>et al.</i> , 2015
<i>Agave tequilana</i>	Residuos enteros de la elaboración de tequila	31	17	Corbin <i>et al.</i> , 2015
<i>Agave tequilana</i>	Bagazo	43	15	Cedeno-Cruz y Alvares-Jacobs, 1999
<i>Agave americana</i>	Fibras de hoja	68.4	4.9	Mylsamy y Rajendran, 2010
<i>Agave salmiana</i>	Bagazo	47.3	10.1	García & Rangel, 2009
<i>Agave lechuguilla</i>	Fibras de hoja	79.8	15.3	Vieira <i>et al.</i> , 2002
<i>Agave fourcroydes</i>	Fibras de hoja	77.6	13.1	Vieira <i>et al.</i> , 2002
<i>Agave sisalana</i>	Fibras de hoja	43	15	McDougall <i>et al.</i> , 1993

De acuerdo con la Cámara Nacional de la Industria Tequilera (CNIT), en el año 2013 para producir un litro de tequila se desechan aproximadamente entre 10 y 12 L de vinazas y de 3.5 a 7 Kg de bagazo de agave. Si anualmente se producen 309.1 millones de litros de tequila (CRT 2018), aproximadamente se estarían generando 700 mil toneladas de hojas de agave, 1.700 mil litros de vinazas y 270 mil toneladas de bagazo. Dada esta información existe una gran cantidad de residuos que representan un gran desafío para su aprovechamiento y así evitar

riesgos de contaminación ambiental. Así mismo, las hojas de los agaves no son biomasa destinada para consumo humano como sucede con el maíz y la caña de azúcar.

2.4.1 Agave como fuente de fibras

En México existen 159 especies de Agave que representan el 75 % de las 210 especies reportadas (García-Mendoza, 2012), 119 de estas especies son endémicas para nuestro país (Figura 4) donde se les da un uso diverso. Los agaves con menor contenido de fibra se han usado principalmente para la producción de bebidas alcohólicas, como es el caso de *Agave tequilana* que se emplea para la elaboración de tequila (Martínez-Ramírez *et al.*, 2013). En tanto, agaves como *Agave angustifolia*, *Agave fourcroydes*, *Agave americana* y *Agave lechuguilla* son preferidos como fuentes de fibras naturales (Mielenz *et al.*, 2015).

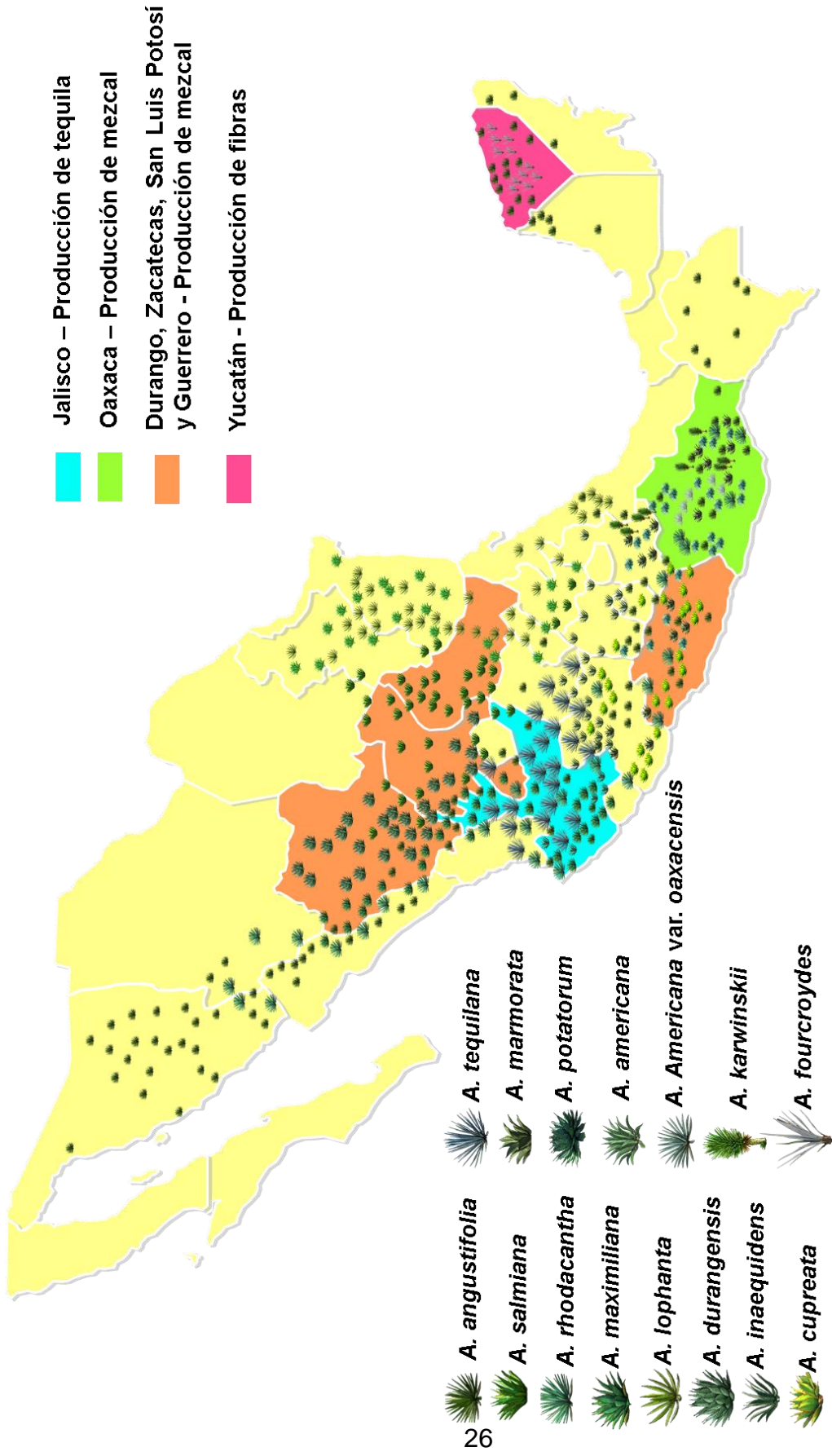


Figura 4. Diversidad de Agave en México y sus principales regiones productoras.

2.5. Pared celular como una fuente de celulosa y lignina

Hasta la fecha, el mejor modelo en plantas para estudiar a la pared celular ha sido *Arabidopsis thaliana*. De ella conocemos que la pared celular es una capa protectora rígida y semipermeable con una estructura compleja, que rodea la membrana plasmática y proporciona protección, estructura y soporte (Figura 5). Su composición puede variar según el organismo, edad y respuesta al medio ambiente. Tal es el caso en la pared celular de las fibras en las hojas de agave, donde la pared es una estructura rígida de múltiples capas compuesta por una lámina central, una pared primaria y una pared secundaria. La celulosa es el componente principal de ambas paredes, y comprende aproximadamente el 95% en peso de las fibras naturales; la lignina, por su parte se distribuye principalmente en la pared secundaria. (Han y Choi, 2010).

La microscopía electrónica de barrido se ha utilizado para conocer la estructura en las fibras de agaves. Las micrografías de la figura 5 (paneles 5C y 5D) muestran secciones transversales y longitudinales, respectivamente, de la estructura celular de las de fibras de *A. angustifolia*. La fibra tiene una ranura tipo C a lo largo del eje y está agrupando de 190 a 210 fibras (Figura 5C), las fibras son ovalada y poligonales. En la figura 5D se observa la disposición longitudinal de las fibras individuales, así como la estructura interna de las microfibrillas.

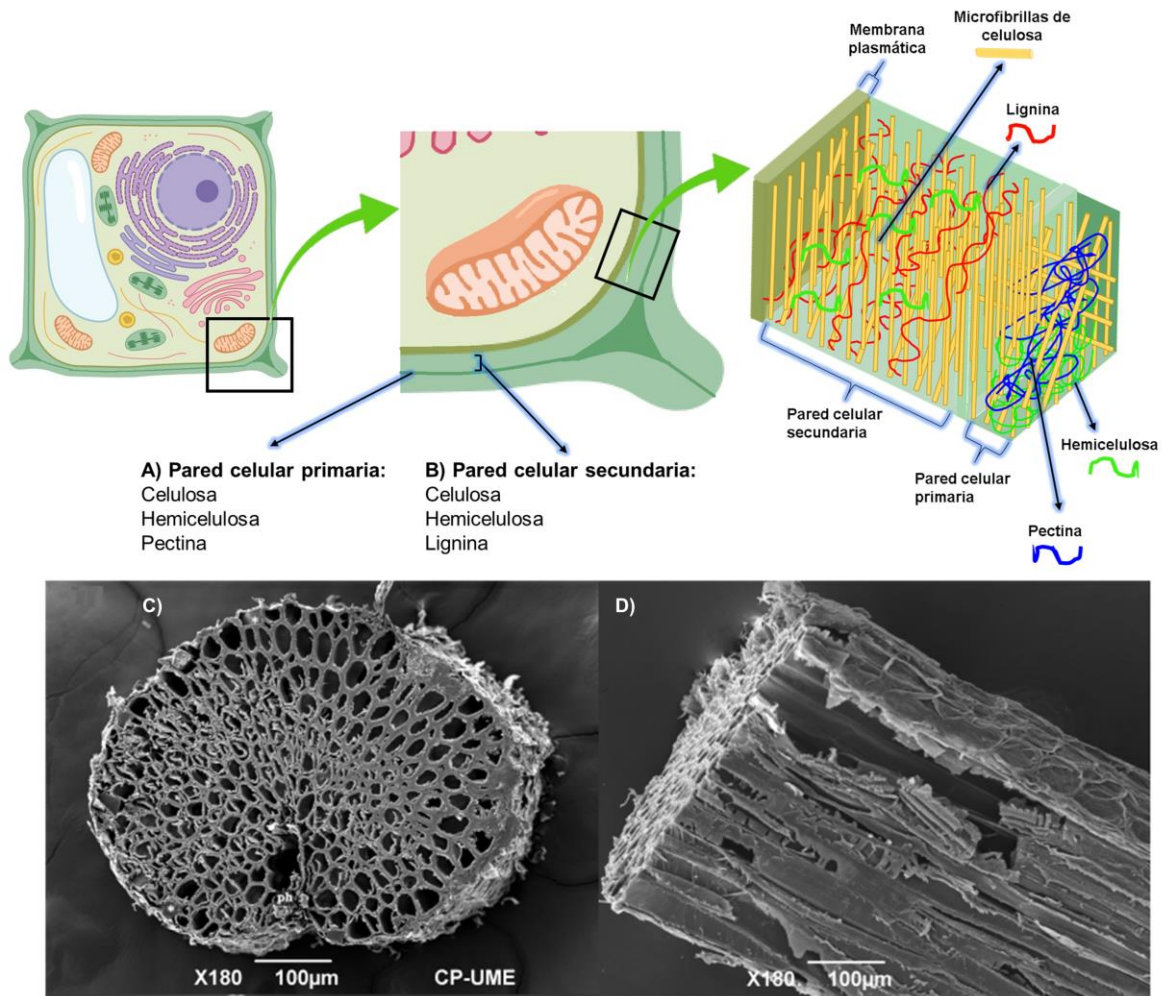


Figura 5. Pared celular. (A) Pared celular primaria. Las microfibrillas de celulosa en la pared celular primaria son cortas y delgadas, en comparación con las de la pared celular secundaria. (B) Pared celular secundaria. Se deposita entre la pared celular primaria y la membrana plasmática. La pared celular secundaria contiene principalmente microfibrillas de celulosa y lignina (Modificada de Nakano *et al.*, 2015). (C) Micrografía electrónica de barrido de la sección transversal y (D) Sección longitudinal de una fibra de *A. angustifolia*.

2.6 Biosíntesis de celulosa en *Arabidopsis thaliana* Lynch

La celulosa es el material estructural en la pared celular primaria y secundaria. Está compuesta por subunidades D-Glucosa unidas por enlaces β -1,4 glucósidos (Pérez *et al.*, 2002), que conforman polímeros lineales, sin ramificaciones y están organizadas de forma paralela para la formación de microfibrillas (Figura 6). Las microfibrillas de celulosa presentan enlaces de hidrogeno inter e intramoleculares generando una estructura cristalina rígida y amorfa (O'Sullivan *et al.*, 1997).

En *A. thaliana*, la biosíntesis de celulosa ocurre en el citoplasma y en la membrana plasmática. La ruta tiene como sustrato inicial glucosa y concluye con la obtención de celulosa. Esta ruta implica una serie de reacciones enzimáticas en las que la glucosa se convierte primero en glucosa-6-fosfato por la enzima glucoquinasa. Posteriormente, glucosa 6-fosfato se convierte en glucosa-1-fosfato por la enzima fosfoglucomutasa. En el siguiente paso, la glucosa-1-fosfato se convierte en UDP-glucosa en presencia de UTP y la enzima UDPG pirofosforilasa. La UDP-glucosa producida se usa finalmente como sustrato por la enzima celulosa sintasa (Figura 6).

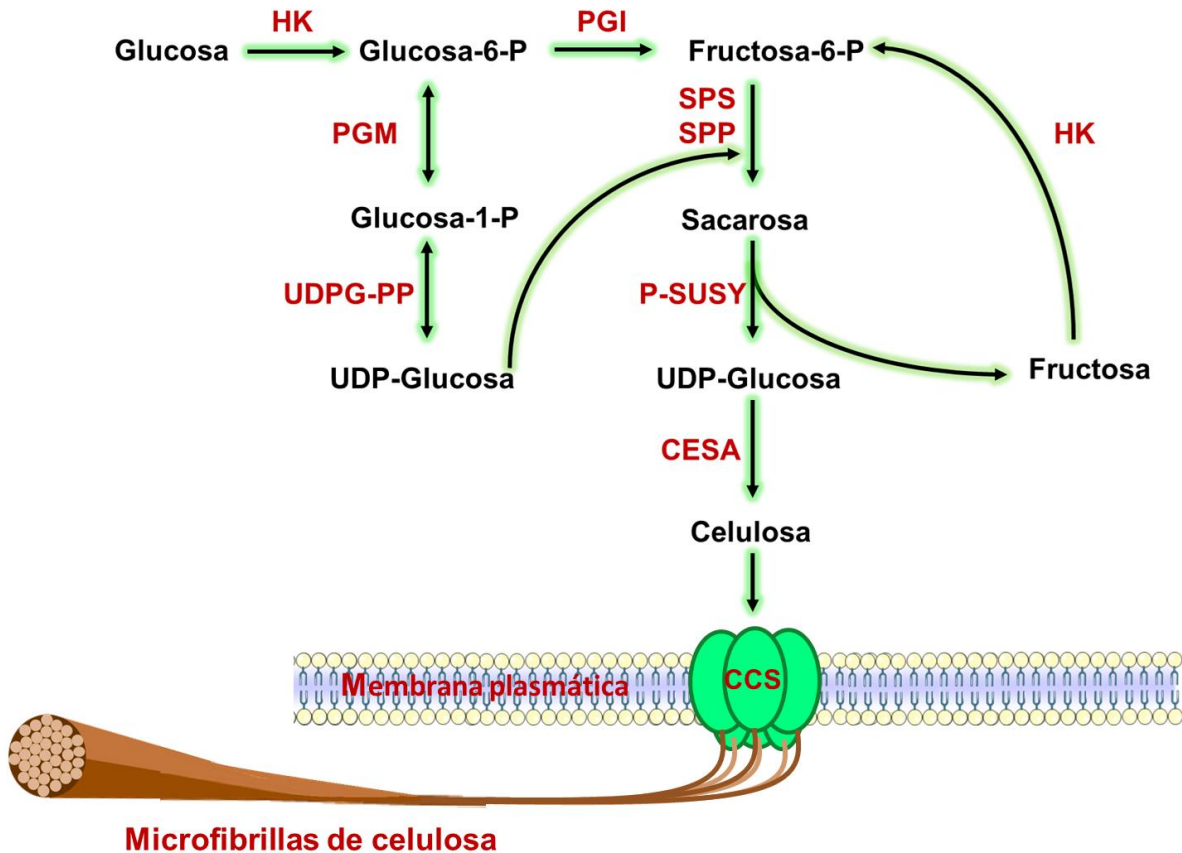


Figura 6. Biosíntesis de celulosa. Hexocinasa (HXK), Fosfoglucoisomerasa (PGI), Fosfoglucomutasa (PGM), Celulosa sintasa (CESA), Fructokinasa (FK), Sucrosa sintasa (SUS), (Babb *et al.*, 2001).

2.6.1 Genes involucrados en la biosíntesis de celulosa en *Arabidopsis thaliana* y su función biológica en el desarrollo vegetal

Se han identificado los genes que codifican para las enzimas involucradas en la biosíntesis de celulosa en el genoma de *Arabidopsis*. Se han reportado 10 genes que codifican para enzimas celulosa sintasa (CESA) en esta planta modelo (Pear *et al.*, 1996).

Las enzimas celulosa sintasa son las encargadas de realizar la polimerización de la glucosa, formando la celulosa. Hasta ahora, las enzimas CESA1, CESA3, CESA6, CESA2, CESA5 y CESA9 (Persson *et al.*, 2007) se han asociado con la biosíntesis de celulosa en la formación de pared primaria, mientras que las

enzimas CESA4, CESA7 y CESA8 se han reportado que están presentes en la biosíntesis de pared secundaria (Turner y Somerville, 1997; Taylor *et al.*, 2000, 2003; Nawaz *et al.*, 2018). El cuadro 2 muestra los fenotipos asociados en las diferentes mutantes alélicas para cada uno de los 10 genes *CESA* en *Arabidopsis*.

Cuadro 2. Mutantes disponibles y fenotipos para el gen *CELULOSA SINTASA* (*CESA*) en *Arabidopsis thaliana*.

Proteína/ Id gen	Alelo	Mutación	Fenotipo	Pared celular	Referencia
CESA1 Atg32410	rsw1-1	A549V	Inflamación de raíz, crecimiento atrofiado (sensible a temperaturas altas) Deficiencia en contenido de celulosa y número de TCs	Primaria	Arioli T <i>et al.</i> , 1998
	rsw1-2	G631S	Letal		Gillmor <i>et al.</i> , Fagard <i>et al.</i> ,
	rsw1-10	T-DNA	Raíces enrolladas, crecimiento detenido		Beeckman <i>et al.</i> ,
	rsw1-20	D780N	Crecimiento anormal en la oscuridad		2002
	rsw1-45	E779K	Crec. deducido de raíz e hipocótilo		
	cesa1	T-DNA	Letalidad del gametofito, Granos de polen estériles y deformados		Persson <i>et al.</i> , 2007
	Aegeus1-2	A903V	Resistencia a quinoxifeno, cristalinidad de la celulosa modificada		Harris <i>et al.</i> , 2012
any1	D604N	Reducción en cristalinidad de celulosa	Fujita <i>et al.</i> , 2013		
CESA2	cesa2	Ds	Orientación MT, expansión abnormal en la célula	Primaria	Chu <i>et al.</i> , 2007
	cesa2	T-DNA	Hipocótilo corto etiolado		Desprez <i>et al.</i> , 2007 Persson <i>et al.</i> , 2007
CESA3 At5g44030	ixr1-1	G998D	Resistencia a Isoxabeno, fenotipo de crecimiento leve	Primaria	Scheible <i>et al.</i> , 2001
	ixr1-2	T942I			
	cev1	G617E	Raíz robusta, crecimiento, inducción de las respuestas de defensa		Elis <i>et al.</i> , 2002
	eli1-1	S301F	Raíz e hipocótilo hinchado, crecimiento atrofiado, inducción de la defensa		Caño-Delgado <i>et al.</i> , 2003
	eli1-2	A522V			
	rsw5	P1056S	Crecimiento atrofiado de la raíz		Wang <i>et al.</i> , 2006
	Tan	P578S	Crecimiento atrofiado de la raíz		Daras <i>et al.</i> , 2009
	repp3	P578L	Raíz corta e hipocótilo etiolado en PIN2:PIN1-HA: pin2		Feraru <i>et al.</i> , 2011
cesa3	T-DNA	Letalidad gametofítica, granos de polen estériles deformados	Persson <i>et al.</i> , 2007		
CESA4 At5g44030	irx5-1	Ds	xilema irregular, biosíntesis de celulosa defectuosa, plantas enanas	Secundaria	Taylor <i>et al.</i> , 2003
irx5-2	W995Stop				
irx5-3	Q263Stop				

	exi2	Y939Stop	Defecto vascular, defecto de expansión celular, xilema colapsado, hojas de roseta pequeñas de, reducción en la expansión celular		Rubio-Díaz <i>et al.</i> , 2012
CESA5	cesa5	T-DNA	Reducción en la deposición de mucilago en la capa de la semilla	Primaria	DeBolt unpublished
CESA6 At5g64740	prc-19 prc1-4/5 prc1-9	Y275Stop W777Stop K7222Stop	Hipocótilo y raíces deformadas, pared celular incompleta	Primaria	Fagard <i>et al.</i> , 2000
	prc1-1/3	Q720Stop	Raíces hinchadas y aturdidas e hipocótilos de crecimiento oscuro, alelos muchos, fenotipo del pelo de la raíz		Fagard <i>et al.</i> , 2000 Singh <i>et al.</i> , 2008
	ixr2 cesa6	R1064W T-DNA	Resistencia semidominante al isoxabeno Fenotipos sutiles		Desprez <i>et al.</i> , 2002 Persson <i>et al.</i> ,
CESA7 At5g17420	irx3	W859Stop	Xilema irregular y biosíntesis de celulosa defectuosa Xilema colapsado, vástago débil		Taylor <i>et al.</i> , 1999 Taylor <i>et al.</i> , 2000
	fra5	P557T	Pared celular de las fibras reducida en contenido de celulosa		Zhong <i>et al.</i> , 2003
	mur10-1 mur10-2	W444Stop H734Y	Reducción en el crecimiento y coloración verde-oscuro de las partes aéreas No formación de pared celular	Secundaria	Bosca <i>et al.</i> , 2006
	exi5	W954Stop	Defecto vascular, xilema colapsado, hojas pequeñas de roseta, reduciendo la expansión celular		Rubio-Díaz <i>et al.</i> , 2012
CESA8 At4g18780	irx1-1 irx1-2	D683N S679L	Vasos de xilema colapsados		Taylor <i>et al.</i> , 2000
	fra6	R362K	Grosor reducido en las fibras, Alelo recesivo		Zhong <i>et al.</i> , 2003
	lew2-1 lew2-2 exi1-1	W217Stop L792F splicing	Tolerante a estrés osmótico, más severo que <i>irx1</i>	Secundaria	Chen <i>et al.</i> , 2005
	exi1-2	G508E	Defecto vascular, defecto de expansión celular, defecto de xilema colapsado, hojas pequeñas de roseta		Rubio-Díaz <i>et al.</i> , 2012
CESA9 At2g21770	cesa9	T-DNA	Testa de las semillas empobrecida de células secundarias	Primaria Secundaria	Stork <i>et al.</i> , 2010
CESA10 At2g25540	--	---	Ningún fenotipo reportado		

La figura 7 ejemplifica además los fenotipos de tres tipos principales de mutantes *AtCESA*: a) *rsw1*, una mutante de *AtCesA1*, b) *prc1-1* y *cesa6*, ambas mutantes nulas de *AtCesA6*; y c) *irx3*, una mutante de *AtCesA7*. La mutante *rsw1*, del acrónimo del inglés *radially swollen1*, presenta disminución en el alargamiento y en la división celular de plántulas jóvenes, lo cual resulta letal a medida que va transcurre el ciclo de vida de la planta (Arioli *et al.*, 1998). El fenotipo de la mutante *irx3*, acrónimo del *irregular xylem3*, presenta un impacto menor en el crecimiento celular y de plántulas; tiene una pared celular secundaria reducida y xilemas

colapsados, y dado que la mutación no es letal se logra formar una pequeña planta madura (Taylor *et al.*, 1999). Las dos mutantes nulas de *AtCESA6*, *prc1-1* y *cesa6*, presentan pocos defectos en la división celular, aunque el alargamiento celular se reduce afectando la integridad de la pared celular y el contenido de biomasa en plantas maduras (Figura 7). Los análisis fenotípicos de mutantes de genes para pared primaria *AtCESA2*, *AtCESA3*, *AtCESA5* y *AtCESA9* mostraron que dichos genes pueden también estar involucrados en la restauración parcial de la pared celular. Sin embargo, la sobreexpresión de *AtCESA2* y *AtCESA5* en mutantes nulas de *AtCesA6* incrementa la división celular y restaura completamente la integridad de la pared celular, produciendo un incremento en el grosor de la pared celular secundaria y en la biomasa en plantas maduras (Hu *et al.*, 2018).

Tipo	Mutante CesaA y planta transgénica	Plántula joven	PCW		SCW Fibras xilares	Planta madura
			Raíz L9	Hipocótilo D9		
I	<i>rsw1</i> (Mutante <i>cesa1</i>)					
II	<i>IRX3</i> (Mutante <i>cesa7</i>)					
III	<i>cesa6</i> <i>prc1-1</i>					
	<i>A2/cesa6</i> <i>A5/cesa6</i>					
	<i>A2/prc1-1</i> <i>A5/prc1-1</i>					
	WT <i>A6/cesa6</i> <i>A6/prc1-1</i>					
IV	<i>A(2+5)/cesa6</i> <i>A(2+5)/prc1-1</i>					

Figura 7. Esquema de los distintos fenotipos de *AtCESAs* en la elongación y división celular, integridad de la pared celular y producción de biomasa vegetal en tres tipos principales de mutantes *AtCesA* (*rsw1*, *cesa6/prc1-1* e *IRX3*) y los mutantes transgénicos que sobreexpresan *AtCesA2* y *AtCesA5* en mutantes nulas de *AtCesA6* (Modificado de Hu *et al.*, 2018).

2.7 Biosíntesis de lignina en *Arabidopsis thaliana* Lynch

Lignina es un hetero-biopolímero, amorfo, insoluble en agua formado por unidades de fenilpropanoides unidas por enlaces de carbono-carbono y aril-éter. La lignina es el segundo biopolímero más abundante en la naturaleza, solo después de la celulosa. Es un componente estructural de la pared celular secundaria (Figura 5B), y entre sus principales funciones está facilitar el transporte de agua y actuar como barrera física contra los patógenos (Pérez *et al.*, 2002; Cesarino *et al.*, 2012).

El polímero de lignina es producido a partir de la ruta de fenilpropanoides; está compuesto de unidades de p-hidroxifenil (H), guaiacil (G) y siringil (S) (Figura 8), derivados de la polimerización de los alcoholes de hidroxicinamilo, p-coumaril, coniferil y alcohol sinapil (Vanholme *et al.*, 2019).

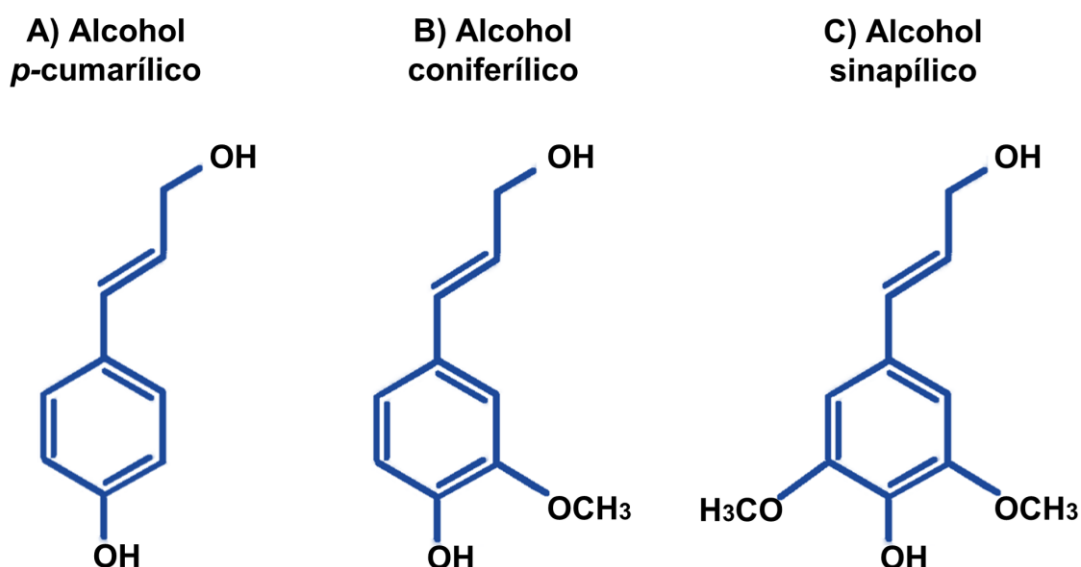


Figura 8. Modelos estructurales de los monómeros de lignina.

Los monolignoles se sintetizan en el citosol a través de la vía fenilpropanoide (Bonawitz y Chapple, 2010) (Figura 9). Este proceso es iniciado por una combinación de enzimas oxidasas y peroxidasas. La polimerización inicia con radicales libres produciendo una reacción de deshidrogenación enzimática y la macromolécula final se extiende por adición de nuevos monómeros (Doherty, 2011).

Entre las enzimas importantes para la biosíntesis de monómeros de lignina están fenilalanina amonio liasa (PAL), hidroxicinamoil CoA: shikimato transferasa (HCT), ácido cafeico O-metiltransferasa (COMT) y alcohol cinámico deshidrogenasa (CAD) (Sattler *et al.*, 2013). La fenilalanina amonio liasa (PAL) es una enzima que está involucrada en la síntesis de compuestos fenólicos; ella cataliza el primer paso comprometido en la biosíntesis de lignina y las vías de los fenilpropanoides. Se ha demostrado que alterar la expresión del gen *PAL* afecta las interacciones planta-patógeno en los sistemas modelo debido a que *PAL* está involucrado en la síntesis de compuestos fenólicos implicados en la defensa como el ácido clorogénico y los compuestos flavonoides (Sattler *et al.*, 2013). Otra enzima clave es Hidroxicinamoil CoA: shikimato transferasa (HCT), la cual constituye un punto clave en la ruta pues en el paso que ella cataliza, la ruta puede desviarse hacia la síntesis de ácido cafeico, o continuar hacia la formación de lignina (Figura 9). La supresión del gen *HCT* en *Arabidopsis* da como resultado reducciones en el contenido de lignina (Li *et al.*, 2010).

Después de la acción de HCT, la ruta continua con CCR quien puede usar tanto p-cumaril CoA o feruloil CoA para avanzar hacia la formación de los monómeros de lignina. La enzima CCR convierte feruloil CoA en coniferaldehído (Leple *et al.*, 2007). El gen *CCR* en *A. thaliana* se expresa en las flores y hojas además se encuentra altamente expresado en tejidos como los tallos, acorde con su función biológica en la síntesis de lignina y el rol de este compuesto en la dureza de la pared celular en órganos de soporte (Lauvergeat *et al.*, 2001). En el penúltimo paso de la biosíntesis de lignina, se ha reportado que la enzima ácido caféico O-metiltransferasa (COMT) participa en la biosíntesis de siringil lignina (Eudes *et al.*, 2017). Sorprendentemente la mutación de este gen en *Arabidopsis* incrementa la resistencia de plantas a patógenos (Maury *et al.*, 2010). Finalmente, la última enzima de la ruta es CAD la cual es la encargada de sintetizar los tres monómeros de lignina (Figura 9). Se ha reportado que el silenciamiento del gen *CAD* aumenta la susceptibilidad de las plantas a patógenos (Tronchet *et al.*, 2010). Así mismo, en investigaciones recientes se ha reportado que el silenciamiento del gen *CAD*,

es una estrategia prometedora para disminuir el contenido de lignina en la biomasa lignocelulósica (Acker *et al.*, 2017).

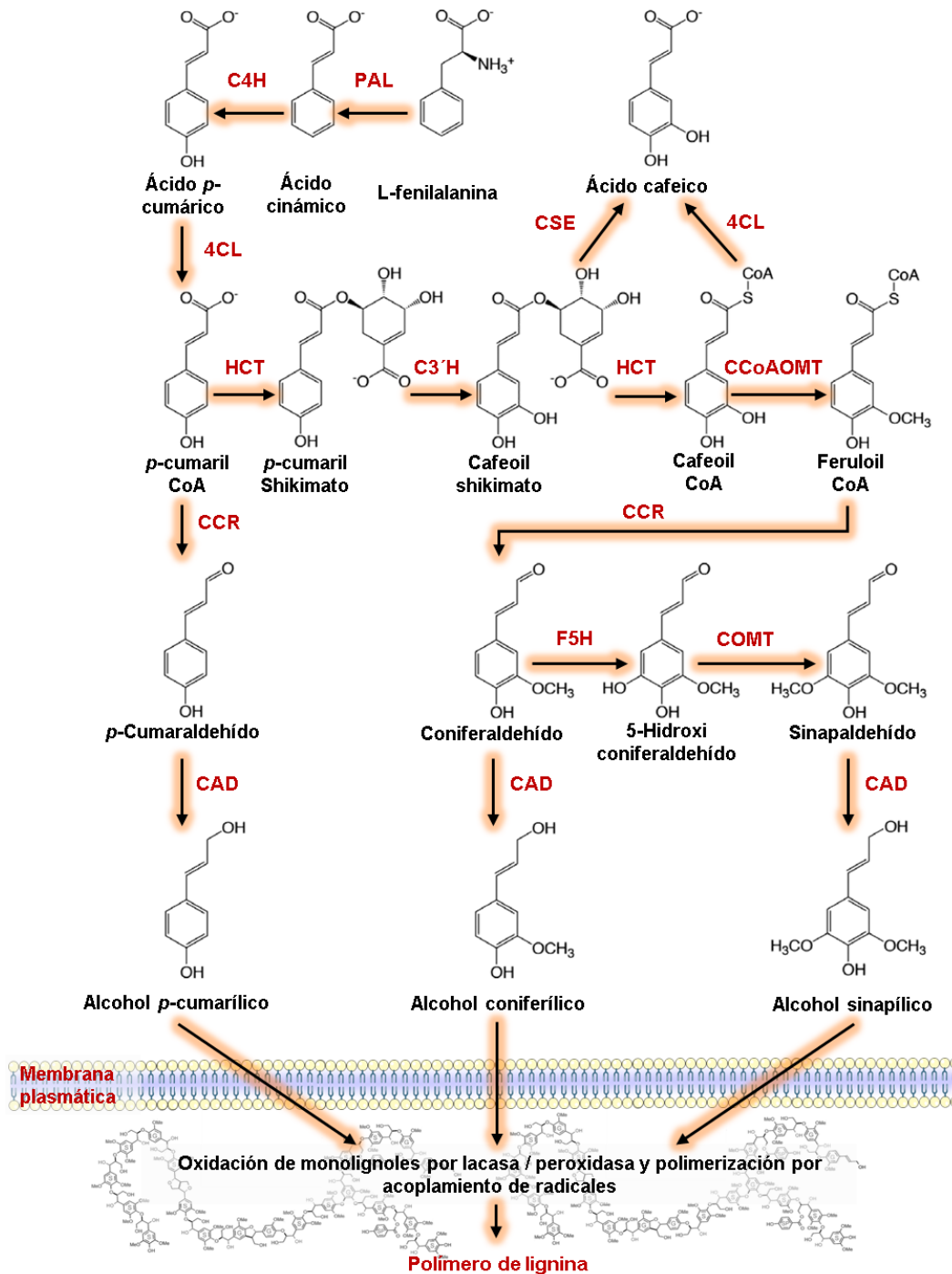


Figura 9. Biosíntesis de lignina. Hay 10 etapas enzimáticas que conducen a alcoholes hidroxicinámicos que se polimerizan en lignina; fenilalanina amonio liasa (PAL), cinamato 4-hidroxilasa (C4H), 4-cumarato-CoA ligasa (4CL), hidroxicinamoil CoA: shikimato transferasa (HCT), p-cumarato 3-hidroxilasa (C3H), cafeoil CoA O - metiltransferasa (CCoAOMT), cinamil CoA reductasa (CCR), ferula 5-hidroxilasa (F5H), ácido cafeico O- metiltransferasa (COMT) y alcohol cinámico deshidrogenasa (CAD), (Sattler *et al.*, 2013).

2.8. Regulación transcripcional de las rutas de formación de la pared celular en plantas

La regulación genética de las rutas de biogénesis y desarrollo de pared celular es conocida para la pared celular primaria y secundaria en dicotiledóneas, especialmente en *A. thaliana*. En contraste, para plantas monocotiledóneas los trabajos sobre regulación transcripcional son aún escasos. Los que se han reportado para algunas especies usan la genómica comparativa respecto a lo conocido en *Arabidopsis*, por ejemplo: *Zea mays* (Xiao *et al.*, 2017), *Brachypodium distachyon* (Valdivia *et al.*, 2013) y *Oryza sativa* (Yoshida *et al.*, 2013).

2.8.1. Regulación transcripcional en *Arabidopsis thaliana*

En *Arabidopsis*, se conoce que la biosíntesis de los componentes de pared celular secundaria está regulada por factores de transcripción tipo MYB y NAC. Se ha reportado que el sistema de regulación transcripcional MYB-NAC está implicado en la modificación de la biosíntesis de la pared celular de tejidos para conducir agua (vasos de xilema) y proporcionar soporte (fibras) (Ohtani *et al.*, 2019).

La familia de genes *NAC* está conformada por más de 100 miembros en *A. thaliana* (Ooka *et al.*, 2003; Nuruzzaman *et al.*, 2010). Demura en el año 2002 identificó el dominio clave en siete factores de transcripción *NAC*. Estas proteínas fueron llamadas VND1 a VND7 (del acrónimo del inglés VASCULAR-RELATED NAC DOMAIN) (Figura 10A). En línea con esto, todos los genes *VND* se expresan en tejidos vasculares. Un claro ejemplo es VND7 cuya expresión regula la formación de los vasos de xilema en raíces y brotes. Las proteínas VND actúan por tanto como reguladores maestros de la diferenciación celular en los vasos de xilema (Yamaguchi *et al.*, 2010).

Otras proteínas relacionadas con VND como NST1 a NST3 (del acrónimo del inglés NAC SECONDARY WALL THICKENING PROMOTING FACTOR), SND1 (SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN 1) regulan también la diferenciación celular en la pared celular secundaria, en *Arabidopsis* (Figura 10A). De hecho, la sobreexpresión de *VND* y *NST* aumenta la expresión de los

genes involucrados en la biosíntesis de celulosa, hemicelulosa y lignina. Además, otros estudios han identificado los reguladores negativos de los genes *VND*, incluyendo los factores de transcripción tipo WRKY, WRKY12 (Figura 10A). La proteína WRKY12 regula negativamente la formación de la pared celular secundaria inhibiendo directamente la expresión de *NST2* en células de parénquima (Wang *et al.*, 2010).

Los factores de transcripción MYB en plantas tienen sólo dos dominios y son llamadas específicamente proteínas R2R3-MYB, y están codificadas por 126 genes en *Arabidopsis*. Los genes de la familia R2R3-MYB están involucrados en varios procesos de desarrollo, respuesta al estrés y metabolismo. Uno de los principales procesos en los cuales diversos estudios en MYBs se han enfocado ha sido en la biosíntesis de lignina (Stracke *et al.*, 2001; Dubos *et al.*, 2010). (Figura 10A). Los genes MYB pueden regular además la biosíntesis de lignina por medio de los elementos AC; por ejemplo, los análisis de los promotores de los genes *PAL* y *4CL* revelan la importancia de los elementos AC, también conocidos como motivo-C1, caja-PAL o caja-H las cuales son ricas en la secuencia adenina-citosina (AC) (Ohtani *et al.*, 2019).

2.8.2. Regulación transcripcional en plantas monocotiledóneas

En plantas monocotiledóneas también se han caracterizado los genes involucrados en la biosíntesis de la pared celular. Sin embargo, la información respecto a cómo se están regulando estos genes en la pared celular en este tipo de plantas es escasa todavía (Figura 10B).

La subfamilia de proteínas NAC, que incluye *VND*, *NST*, *SMB* y *BRN* de *Arabidopsis* se ha denominado como la familia VNS (Ohtani *et al.*, 2011; Xu *et al.*, 2014). Las proteínas VNS en plantas monocotiledóneas han sido estudiadas respecto a sus funciones moleculares y niveles de expresión (Valdivia *et al.*, 2013). Los miembros de los grupos *VND* y *NST* en *Oryza sativa* y *Zea mays* fueron llamados SWN (del acrónimo del inglés SECONDARY WALL-

ASSOCIATED NAC). Estas proteínas están presentes en la formación de la pared celular secundaria, en vasos de xilema y en fibras corticales (Nakano *et al.*, 2015).

De los pocos factores de transcripción que se han estudiado en monocotiledóneas destacan ZmMYB31 y ZmMYB42 en maíz. Ambos genes están asociados con la desregulación o represión en la biosíntesis de lignina y por consiguiente una reducción en el contenido de lignina (Fornalé *et al.*, 2010). Así mismo, la sobreexpresión de ZmMYB42 causa una reducción en el tamaño de la hoja (Sonbol *et al.*, 2009). En la planta *Panicum virgatum* se ha reportado que el elemento AC está implicado en el sitio de enlace de PvMYB4 disminuyendo la acumulación de lignina. Además, los ortólogos de *A. thaliana* en maíz ZmSWNs y arroz OsSWNs, fueron demostrados que activan la biosíntesis de la pared celular secundaria cuando se sobreexpresan en *A. thaliana* (Zhong y Ye, 2012) (Figura 10B).

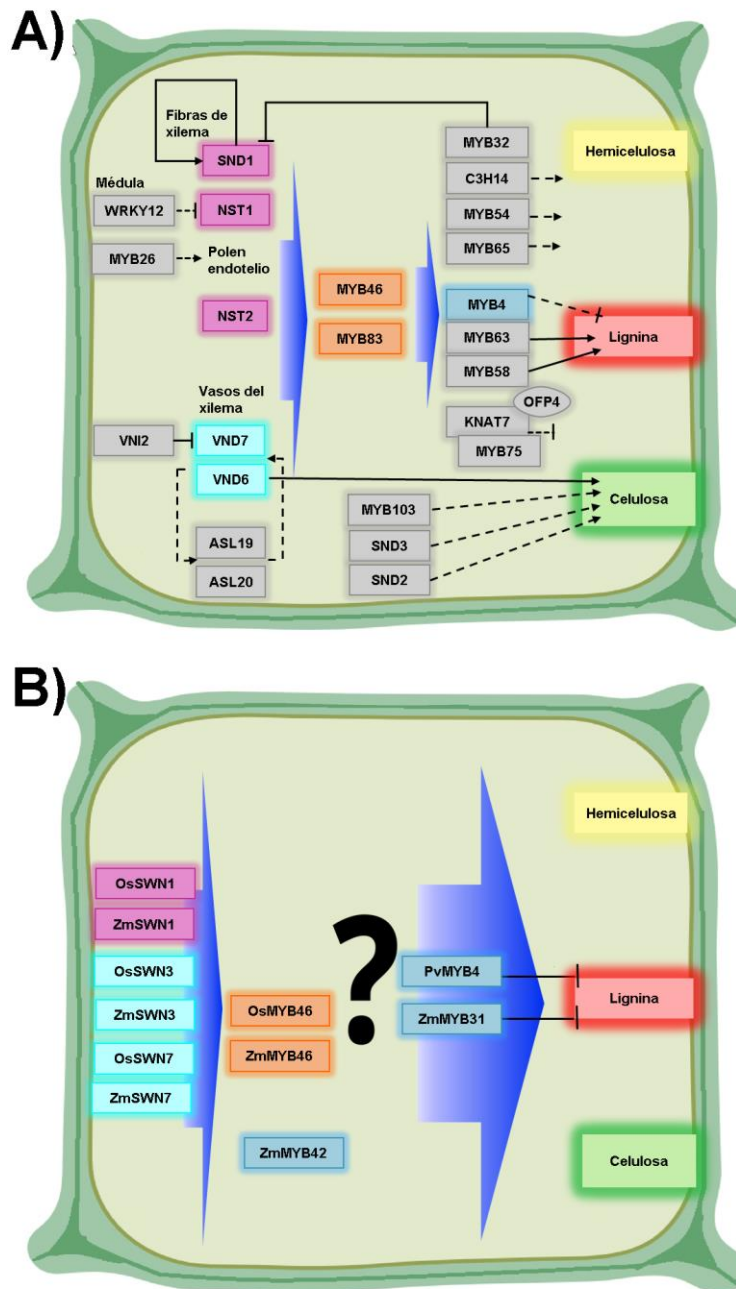


Figura 10. Ruta de regulación genética para la biosíntesis de los principales componentes de pared celular en *Arabidopsis thaliana* (A) y monocotiledóneas (B). Los rectángulos representan factores de transcripción. El óvalo indica la interacción con una proteína. Las flechas sólidas significan las interacciones directas. La flecha discontinua indica no interacción directa. Los colores similares indican la ortología entre *A. thaliana* y las monocotiledóneas (Handakumbura *et al.*, 2012; McCahill *et al.*, 2019).

2.9. Importancia de las ciencias ómicas en el estudio de *Agave tequilana*

El estudio de los sistemas biológicos requiere de la unión de diversas ciencias, que tienen por objetivo describir con precisión los mecanismos que ocurren a nivel de fenotipo y genotipo y cómo es que han cambiado a lo largo del tiempo. Algunos mecanismos pueden ser las interacciones de las plantas con las características ambientales, como la temperatura y la disponibilidad de nutrientes o las interacciones con otros organismos. En la actualidad existen técnicas que permiten describir los sistemas biológicos a nivel de genotipo, proteínas y metabolitos. El uso de estas técnicas ómicas ha permitido comprender las respuestas de las plantas a ciertos estímulos y como estos afectan su fenotipo (Labate *et al.*, 2019).

2.9.1. Aplicaciones de las ciencias ómicas en *Agave spp.*

El uso de las ciencias ómicas como genómica, transcriptómica, proteómica y metabolómica están permitiendo tener un conocimiento amplio de los mecanismos a nivel celular. Desde la identificación de los genes, cómo se transcriben, las reacciones de las proteínas y el potencial de los metabolitos que están presentes.

El uso de estudios ómicos en *A. tequilana* son de gran importancia para entender los procesos y funciones biológicas de esta planta durante su desarrollo y su interacción con el ambiente. Usando la plataforma de secuenciación de Illumina, ahora se conocen genes e isoformas alternativas expresados, y perfiles de expresión en los diferentes tejidos y órganos, en diferentes tipos de condiciones. En el año 2013 fueron publicados por primera vez los transcriptomas de raíz, tallo y hoja de *A. tequilana* y *A. deserti*, realizados por Gross y colaboradores. Para la secuenciación utilizaron Illumina HiSeq 2000 con una profundidad de 978 millones de reads para *A. tequilana* y 615 millones de reads para *A. deserti*. Así mismo como resultado final de ensamblado, obtuvieron 34870 genes codificantes para *A. tequilana* y 35086 para *A. deserti*. Además, los análisis transcriptómicos en hojas de *A. deserti* muestran que la expresión de la biosíntesis de celulosa y lignina es

gradual y que va disminuyendo desde la parte basal de la hoja hasta la parte apical (Gross *et al.*, 2013).

Por otra parte, en un estudio más reciente realizado por Huang *et al.*, 2019, se reportó el ensamblado *de novo* del transcriptoma de *Agave* H11648 ((*Agave angustifolia* Haw. x *Agave amaniensis* Trel. & W. Nowell) x *Agave angustifolia*), el cual es utilizado para la producción de fibras. Los tejidos colectados para este estudio fueron hojas de plantas de *A.* H11648, *A. americana*, *A. deserti* y *A. tequilana* de aproximadamente tres años de edad. Para la secuenciación de *A.* H11648 utilizaron la plataforma Illumina HiSeq obteniendo una profundidad de 60 millones de reads crudos. Como resultado final de ensamblado obtuvieron 148046 unigenes. Así mismo el objetivo de este estudio era identificar los genes de celulosa sintasa en las diferentes especies de agave. Su estudio incluyó la búsqueda de genes *CESA*. Los genes que encontraron fueron *CESA1*, *CESA3*, *CESA4*, *CESA5* y *CESA7* en las especies de *A.* H11648, *A. americana*, *A. deserti* y *A. tequilana*, cuyos análisis filogenéticos indicaron patrones de evolución conservada (Huang *et al.*, 2019).

2.10. Clasificación de los genes homólogos

La identificación de los genes homólogos ha sido de gran interés debido a que permite conocer lo que está conservado y lo que ha cambiado durante la evolución. Además, evaluar la relación evolutiva de los genes proporciona una forma para definir la estructura y función de las proteínas que son codificadas entre las diferentes especies (Eisen, 1998). Los homólogos se clasifican en dos grupos: a) Ortólogos (mismo gen realizando la misma función en diferentes especies) y b) Parálogos (son copias de un gen que eventualmente pueden diferir en función a través de la evolución o perder función). Los homólogos son secuencialmente similares y comparten similitud estructural y funcional. Sin embargo, la similitud no necesariamente significa homología (Adrain y Freeman, 2012).

En la figura 11 se muestra el ejemplo de los genes de la insulina en humanos, ratones y ratas. Todos descienden de un gen ancestral común en el último ancestro común de todos los mamíferos. El ratón tiene dos genes de insulina *Ins1* e *Ins2*, que se duplican dentro del linaje roedor (Shiao *et al.*, 2008). El humano tiene una copia, *INS*. Por lo tanto, *Ins1* es ortólogo a *INS*, *Ins2* es ortólogo a *INS*, pero *Ins1* no es ortólogo sino parálogo a *Ins2*.

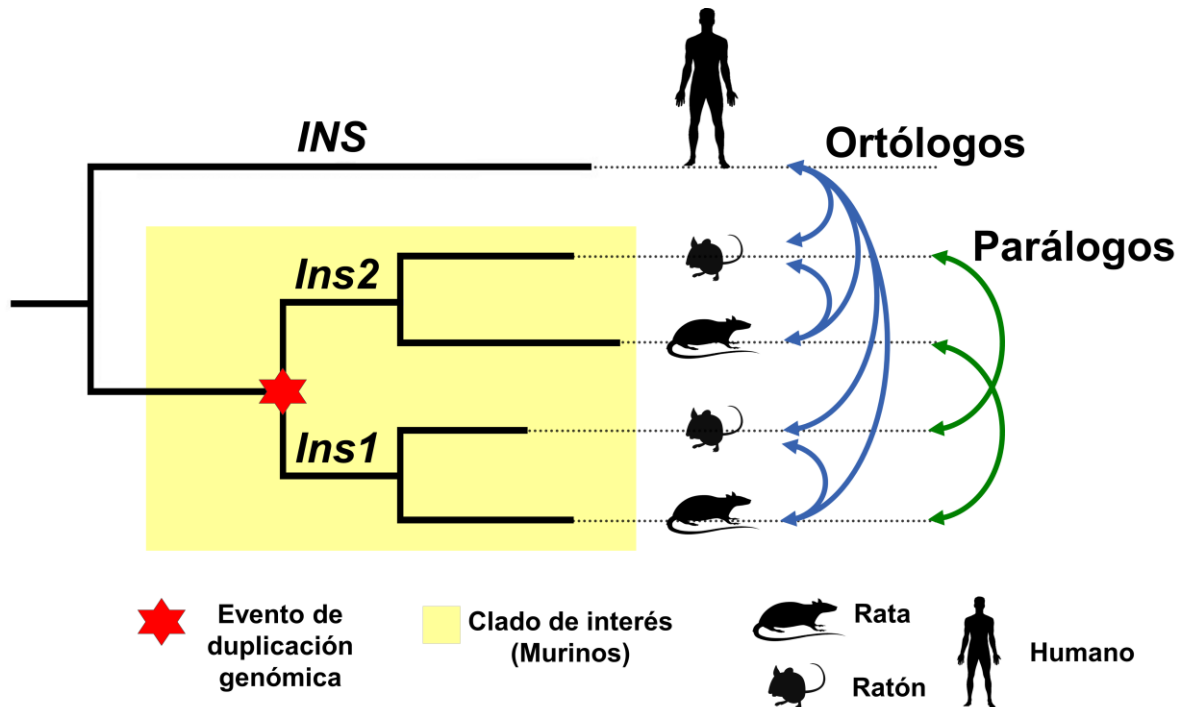


Figura 11. Árbol de los genes de insulina ortólogos y parálogos en ratón, rata y humano (Modificado de Fernández *et al.*, 2019).

III. OBJETIVO

3.1 Objetivo general

- Identificar y caracterizar los genes involucrados en las rutas biosintéticas de celulosa y lignina de *Agave tequilana*, y medir su expresión génica en diferentes tejidos.

3.2 Objetivos particulares

- Identificar *in silico* los genes ortólogos involucrados en las rutas de biosíntesis de celulosa y lignina que están relacionadas en la formación de la pared celular en *A. tequilana* Weber.
- Analizar el perfil transcriptómico *in silico* de los genes identificados en los diferentes tejidos de *A. tequilana* Weber.
- Analizar experimentalmente la expresión de genes candidatos por medio de ensayos RT-PCR en *A. tequilana* Weber.

IV. Hipótesis

4.1 Hipótesis general

Las rutas biosintéticas de lignina y celulosa se encuentran conservadas en plantas, lo que permitirá la identificación de los genes ortólogos involucrados en estas rutas en *Agave tequilana* Weber para realizar los estudios de expresión génica.

V. MATERIALES Y MÉTODOS

5.1 Metodología general usada en este trabajo

Las actividades realizadas en este trabajo se agrupan en 4 diferentes etapas. La metodología general se ilustra en la figura 12. A continuación se describen brevemente cada una de las etapas:

- a) *Entrenamiento bioinformático y análisis preliminar de identificación de genes ortólogos involucrados en la biosíntesis de celulosa y lignina*: Para esta primera parte, se realizó una estancia bajo la asesoría de la Dra. Elsa Góngora Castillo, en el Centro de Investigaciones Científicas de Yucatán (CICY) donde el objetivo fue familiarizarse con herramientas bioinformáticas y realizar una primera identificación de los genes ortólogos involucrados en la biosíntesis de celulosa y lignina en los transcriptomas *A. tequilana*. Se utilizaron datos previamente publicados y sin reportar (J. Simpson, CINVESTAV Irapuato), de órganos y/o tejidos con pared celular primaria y/o secundaria. Esta búsqueda fue realizada utilizando el software OrthoFinder versión 2.2.7 (Emms *et al.*, 2015) y para la generación de los perfiles de expresión se utilizó la librería gplots del paquete estadístico R versión 3.5.1 (<https://cran.r-project.org/bin/windows/base/old/3.5.1/>).
- b) *Análisis complementario de búsqueda de genes ortólogos involucrados en la biosíntesis de celulosa y lignina*. La primera identificación de genes ortólogos fue complementada realizando un segundo análisis ahora ensamblando de novo transcriptomas previamente reportados (Gross *et al.*, 2013) de *A. tequilana* y *A. deserti* cuyos datos están disponibles a todo el público. Para esto, se realizó una estancia bajo la supervisión del Dr. Enrique Ibarra Laclette, en el Instituto de Ecología A.C. (INECOL). La metodología utilizada se describe de la sección 5.2 a la 5.8 de este documento. Los genes ortólogos hallados con esta metodología fueron finalmente cotejados con aquellos previamente obtenidos durante la estancia en CICY; para algunas familias génicas como *CELULOSA SINTASA (CESA)*, el número de genes hallados fueron comparados con

listas recién publicadas para el agave híbrido *A. H11648*, *A. deserti* y *A. tequilana* (Huang *et al.*, 2019)

- c) *Perfiles transcriptómicos de los genes ortólogos*: Para conocer los perfiles transcriptómicos de los genes ortólogos relacionados con la síntesis de celulosa y lignina, se analizó una base local de datos transcriptómicos de *A. tequilana* (8 transcriptomas) y *A. deserti* (4 transcriptomas) (Laboratorio de Genética Molecular del Desarrollo Sexual y Asexual, Dra. June Simpson) (ver metodología en el apartado 5.9 de este escrito).
- d) *Validación experimental de la expresión de genes involucrados con la síntesis de celulosa y lignina*. La última parte de este trabajo correspondió a la validación experimental de los niveles de expresión para algunos genes ortólogos hallados. Brevemente, para esto se realizaron ensayos RT-PCR para determinar la expresión de los genes involucrados en la biosíntesis de celulosa, *SUS2*, *CESA1* y *CESA7*, y 2 genes para lignina, *CCR2* y *CAD5*, en raíz, piña, meristemo y hoja de plantas de *A. tequilana* de aproximadamente 5 años de edad. En un segundo experimento, se analizaron la expresión de los genes utilizados en hojas de plantas de aproximadamente 1.5 años de edad de *A. tequilana*, y en el tejido adyacente a las fibras (ver Sección 5.10 a 5.13).

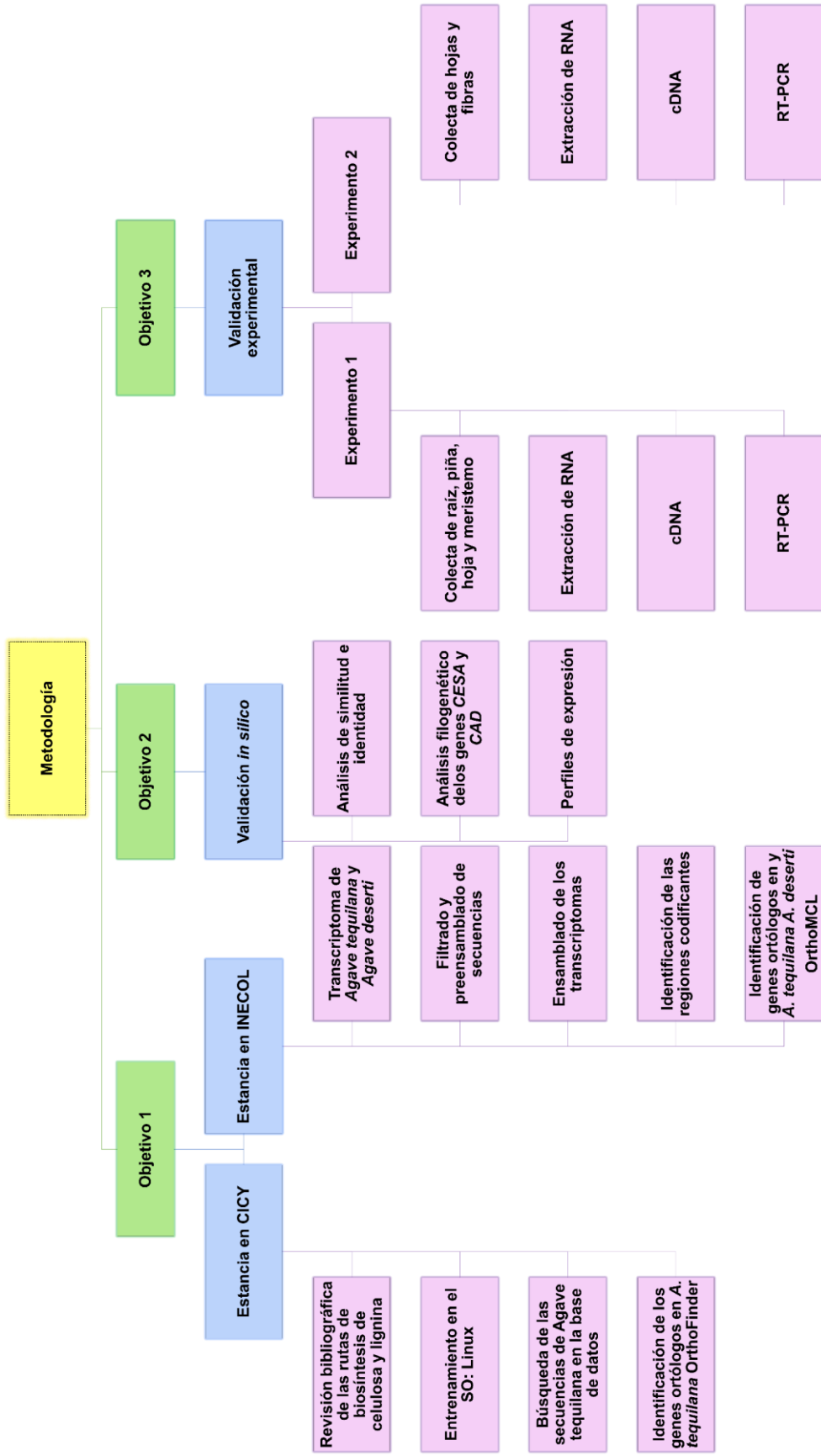


Figura 12. Metodología general usada en este trabajo.

5.2. Descarga de los reads de los transcriptomas de *Agave tequilana* y *Agave deserti*

Los datos utilizados en este estudio fueron previamente reportados por Gross *et al.*, 2013 y están disponibles en el NCBI a través de archivos que contienen las secuencias de los reads (NCBI Sequence Read Archive: SRA) con el número de accesión para *A. tequilana* SRP019885, y para *A. deserti* SRP019506 (Gross *et al.*, 2013). Los reads de estos transcriptomas provienen de la secuenciación por Illumina HiSeq 2000 correspondientes a plantas de *A. tequilana* de tejidos o estadios juveniles, meristemo, raíces y hojas adultas. En *A. deserti*, los tejidos secuenciados fueron hojas plegadas y meristemo, hijuelos, raíces, y hojas divididas en cuatro secciones (Sección 1 basal, sección 2 proximal media, sección 3 media distal y sección 4 distal) y la hoja proximal sección 1 y 2 (Cuadro 3).

Cuadro 3. Datos de los transcriptomas utilizados en este estudio (Gross *et al.*, 2013).

Especie	Plataforma	Estrategia	Fuente	Selección	Diseño	Experimento	Nombre de la librería
<i>Agave tequilana</i> [GenBank: SRP019885]	Illumina	RNA-Seq	Transcriptoma	cDNA	PAIRED	SRX253025	Juveniles 500bp
						SRX253024	Juveniles 250bp
						SRX253023	Raíces 500bp
						SRX253022	Raíces 250bp
						SRX253021	Meristemo 500bp
						SRX253020	Meristemo 250bp
						SRX253019	Hojas adultas 500bp
						SRX253018	Hojas adultas 250bp
<i>Agave deserti</i> [GenBank: SRP019506]	Illumina	RNA-Seq	Transcriptoma	cDNA	PAIRED	SRX252521	Hojas sección 1 y 2 500bp
						SRX252520	Hojas sección 1 y 2 250bp
						SRX252519	Hijuelos 500bp
						SRX252518	Hijuelos 250bp
						SRX252517	Hojas sección 4 500bp
						SRX252516	Hojas sección 4 250bp
						SRX252515	Hojas sección 3 500bp
						SRX252514	Hojas sección 3 250bp
						SRX252459	Hojas sección 2 500bp
						SRX252458	Hojas sección 1 500bp
						SRX252457	Hojas sección 2 250bp
						SRX252456	Hojas sección 1 500bp
						SRX252455	Meristemo 500bp
						SRX252454	Meristemo 250bp
SRX252453	Raíces 500bp						
SRX252452	Raíces 250bp						

5.3 Filtrado y preensamblado de secuencias

Las lecturas pareadas fueron filtradas para eliminar las secuencias de baja calidad. Un “script” de python disponible en GitHub (<https://github.com/Czh3/NGSTools/blob/master/qualityControl.py>) fue utilizado para este fin. Los parámetros utilizados para obtener las lecturas pareadas de alta calidad fueron los siguientes: -q 20 (valor mínimo de 20 para la calidad permitida en formato Phred), -p 90 (el porcentaje mínimo de bases en la secuencia que deben tener una calidad q) y -a 25 (límite de la calidad promedio en formato Phred

estimada a lo largo de la secuencia). Una vez obtenidas aquellas secuencias de alta calidad, los pares de secuencias (R1 y R2) fueron unidas para identificar una región de solapamiento en sus porciones terminales. El programa utilizado para esto fue SeqPrep (<https://github.com/jstjohn/SeqPrep>) utilizando las opciones -L 75 (longitud mínima de lectura) y -o 25 (superposición mínima de bases para unir dos lecturas). El proceso de preensamblado (unir secuencias R1 y R2) se realizó con el objetivo de mejorar la calidad del ensamblado final al incluir secuencias de mayor longitud, así como optimizar los requerimientos computacionales. Las secuencias apareadas (R1 y R2) que no se unieron a ninguna otra fueron descartadas por no cumplir con los parámetros de calidad.

5.4 Ensamblado de los transcriptomas

El ensamblador Trinity (Grabherr *et al.*, 2011) fue utilizado para realizar el ensamblado *de novo* de las secuencias apareadas que cumplieron con los filtros de calidad, incluyendo las secuencias (R1 y R2) que mostraron solapamiento. El programa Trinity se ejecutó utilizando los parámetros por default o preestablecidos. Los contigs resultantes fueron procesados con el programa SeqClean (<https://sourceforge.net/projects/seqclean/>) para eliminar y recortar las secuencias de las regiones terminales de baja complejidad, ricas en bases indeterminadas o colas de poli A/T. Posteriormente se utilizó el programa BLASTClust (<https://www.ncbi.nlm.nih.gov/Web/Newsltr/Spring04/blastlab.html>) con el fin de crear un conjunto de secuencias no redundantes. Para identificar y eliminar las secuencias contaminantes, se utilizó el programa DeconSeq versión 0.4.3 (<http://deconseq.sourceforge.net/manual.html>). Este programa utiliza una base de datos proporcionada por el usuario en la cual están depositadas las secuencias de varios organismos pertenecientes a diferentes reinos.

5.5 Identificación de las regiones codificantes

Para obtener el conjunto de unigenes no redundantes representativos de los transcriptomas de *A. tequilana* y *A. deserti* se identificaron las regiones codificantes. Los marcos de lectura se corrigieron realizando el alineamiento de

los unigenes con las secuencias de genes homólogos identificados en una base de datos que contiene las secuencias codificantes de los genes de plantas cuyo genoma ha sido completamente secuenciado. Debido a esto se utilizó AlignWise (Evans & Loose, 2015), un algoritmo que maneja diferentes programas como BLAST (Basic Local Alignment Search Tool; Altschul *et al.*, 1990), para la identificación de secuencias homologas con alto porcentaje de similitud; Muscle (Edgarr, 2004), para liderar los alineamientos y las secuencias y GeneWise (Birney *et al.*, 2004) para identificar regiones codificantes y corregir los marcos de lectura producto de inserciones y deleciones erróneas.

5.6 Identificación de genes ortólogos en *Agave tequilana* y *Agave deserti*

Los genes ortólogos y parálogos fueron identificados utilizando el algoritmo de agrupamiento de Marcov (MCL del inglés Marcov Clustering; Van Dongen *et al.*, 2000) utilizando un valor de inflación de 1.5 (Enright *et al.*, 2002). El software utilizado para cumplir con este propósito fue OrthoMCL versión 2.0.9 (Li *et al.*, 2003). Para la formación de los grupos ortólogos se consideró una longitud mínima de 30 aminoácidos en todas las proteínas incluidas en el análisis.

5.7 Análisis de similitud e identidad en los genes ortólogos identificados en *A. tequilana* y *A. deserti*

Para los análisis de similitud e identidad se escogieron solamente las familias más importantes en las rutas de biosíntesis de celulosa y lignina. Para la biosíntesis de celulosa se seleccionó la familia de genes *CESA* la cual está involucrada en la formación de pared primaria y secundaria. Para la biosíntesis de lignina se seleccionó la familia de genes *CAD* la cual está involucrada en la formación de pared secundaria. Para estos análisis se utilizó el programa Sequence Manipulation Suite: Ident and Sim (https://www.bioinformatics.org/sms2/ident_sim.html; Stothard, 2000).

La identidad de una secuencia es la cantidad de caracteres que coinciden exactamente entre dos secuencias diferentes. La similitud se calcula considerando

todas las sustituciones idénticas y favorables entre dos secuencias. Tanto la identidad y similitud se expresan en porcentaje (%). Ambas se utilizan para deducir la homología. Sin embargo, la homología tiene una definición específica la cual es tener un ancestro evolutivo común (Nair *et al.*, 2002; Campanella *et al.*, 2003).

5.8 Alineamiento de las secuencias de aminoácidos y análisis filogenético de CESA y CAD en *A. tequilana* y *A. deserti*

Las secuencias de aminoácidos de las proteínas AqCESA, AdCESA, AqCAD, AdCAD fueron obtenidas a partir de los transcriptomas ensamblados de *A. tequilana* y *A. deserti* en este trabajo y las proteínas CESA y CAD de otras plantas fueron obtenidas de la base de datos NCBI (National Center for Biotechnological Information. <https://www.ncbi.nlm.nih.gov/>). Con todas las secuencias obtenidas se realizó un alineamiento múltiple de secuencias de aminoácidos con Clustal-W (Thompson *et al.*, 1994), y el análisis filogenético fue realizado con MEGA versión 6.06, usando el método de Neighbor-joining. En el anexo A1 y A2 se enlistan las proteínas usadas para la familia CESA y CAD.

5.9 Generación de la matriz de los perfiles de expresión

Los perfiles de expresión *in silico* de los diferentes órganos de *A. tequilana* y *A. deserti* que presentan tejidos con pared primaria (meristemo, anteras, ovarios, tépalos y pistilos) y aquellos con pared primaria y potencialmente enriquecidos con pared secundaria (raíz, piña y hojas) fueron obtenidos a partir de los transcriptomas reportados y sin reportar (J. Simpson, CINVESTAV Irapuato; Ávila de Dios *et al.*, 2015; Ávila de Dios *et al.*, 2019). Dichos perfiles de expresión de los tejidos de *A. tequilana* y *A. deserti* se obtuvieron utilizando el software RSEM (del acrónimo en inglés de “RNA-seq by Expectation Maximization”; Li y Dewey, 2011). RSEM utiliza un algoritmo de maximización de expectativas, el cual produce una matriz con los perfiles de expresión en transcritos por millón” TPM”. Los mapas de calor fueron creados con base en los niveles de expresión de los genes involucrados en la biosíntesis de celulosa y lignina en *A. tequilana* y *A. deserti*,

usando la función `heatmap.2` de la librería `gplots` del paquete estadístico R versión 3.6.1 (<https://www.r-project.org>).

5.10 Análisis de expresión de genes involucrados en pared celular en raíz, piña, meristemo y hoja de *A. tequilana*

5.10.1 Colecta del material

Para este primer análisis, los órganos y tejidos fueron obtenidos a partir de dos plantas de *A. tequilana* de aproximadamente 5 años de edad, localizadas en el Rancho Agropecuaria Santa Genoveva, en Campeche (Cuadro 4 y Figura 13A). Para la selección de la hoja a colectar, se eligió la cuarta hoja a partir de la hoja 1 enrollada. Esta hoja fue asperjada con etanol al 70 %, limpiada con hojas de papel interdobladadas y entonces seccionada en 3 partes: parte proximal de la hoja, parte media de la hoja y parte apical de la hoja (Figura 13D). De cada sección, se tomó la parte central, se cortó en cubos pequeños, los cuales se congelaron inmediatamente en nitrógeno líquido (N₂). Para el caso del meristemo, la piña se jimió para acceder a este tejido, posterior la piña se cortó transversalmente en su parte media, y se tomaron muestras y se congelaron también en N₂ en ambos casos (Figura 13B y 13C).

Las raíces en crecimiento de *A. tequilana* presentan un compuesto de color rojo en su epidermis, el cual afecta la extracción de RNA, según trabajos previos en nuestro laboratorio. Por esto, optamos por retirar el compuesto mediante etanol absoluto y papel; la raíz limpia se seccionó en fragmentos de 4-6 cm, y se colectó como previamente se describió para los otros órganos.

Cuadro 4. Características de las plantas de *A. tequilana* colectadas en el Rancho Agropecuaria Santa Genoveva, Campeche.

	Planta 1	Planta 2
Alto	1.68 m	1.71 m
Ancho	1.95 m	1.74 m
Numero de hojas	62	54
Longitud de la hoja seleccionada	1.40 m	1.43 m

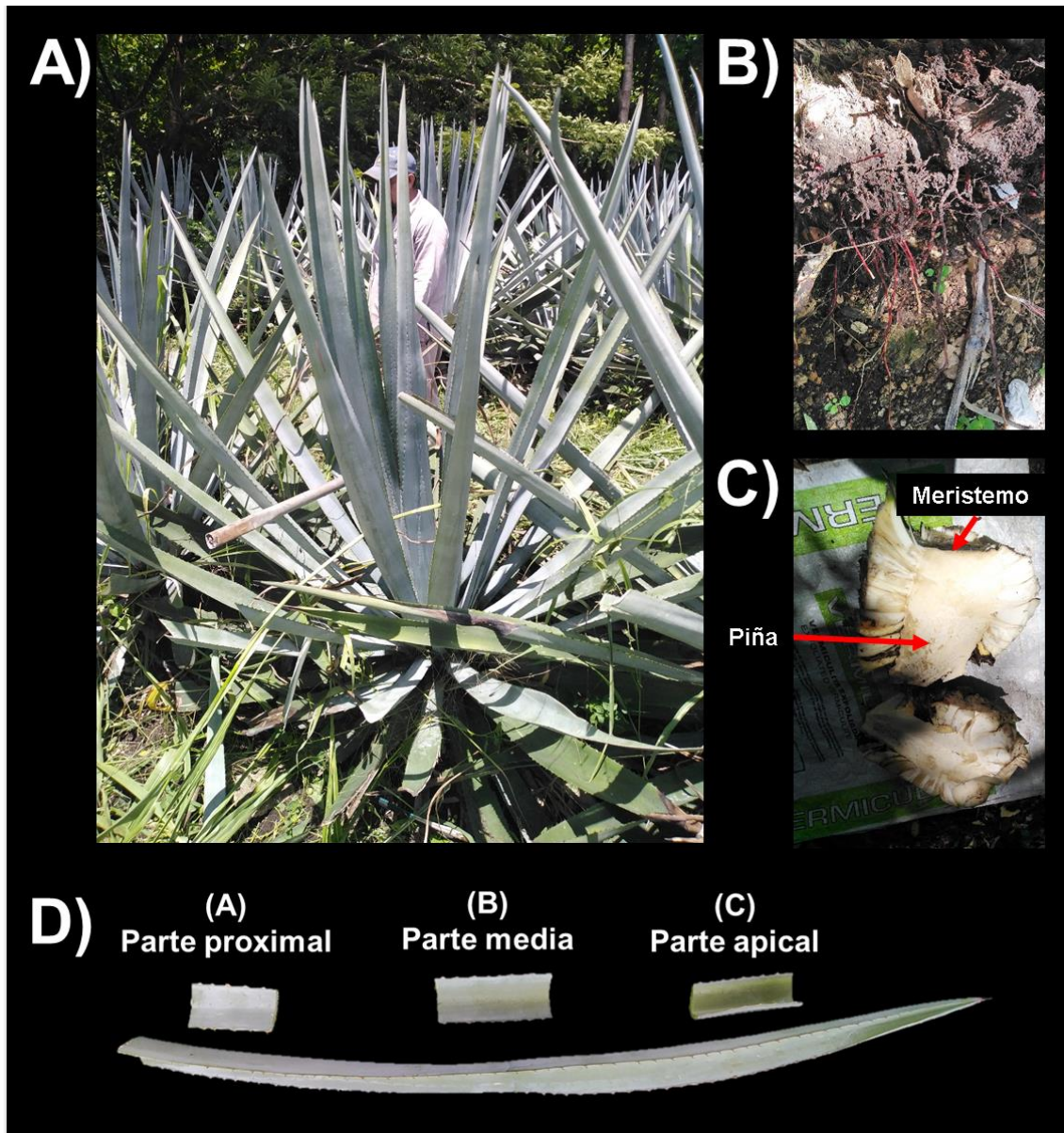


Figura 13. Colecta del material de *A. tequilana* para la extracción de RNA total. A) Planta de aproximadamente 5 años de edad. B) Raíces. C) Piña y meristemo. D) Secciones utilizadas de las hojas.

5.10.2 Extracción RNA total

Cada tejido colectado de Agave fue pulverizado haciendo uso de mortero y pistilo con nitrógeno líquido. Aproximadamente 100 mg del homogeneizado fueron empleados para la extracción de RNA total usando el reactivo de lisis QIAzol (QIAzol Lysis Reagent, Quiagen), con algunas modificaciones al protocolo que indica el fabricante. Las modificaciones al protocolo se ilustran con el tejido de raíz, en la figura 14. Los RNAs extraídos fueron purificados utilizando el kit RNeasy Mini Kit (Quiagen). Para determinar el grado de pureza del RNA obtenido, se consideraron los valores de absorbancia a longitudes de onda específicas (260 y 280 nm) obtenidos con un espectrofotómetro NanoDrop™ Lite de Thermo Scientific™. Las relaciones de las absorbancias 260/280 fueron consideradas como indicadores del grado de pureza de las muestras debido a que a 280 nm y 260 nm absorben principalmente proteínas y ácidos nucleicos, respectivamente (Sambrook y Rusell, 2001; Kasem *et al.*, 2008; Jayasena *et al.*, 2014; Xiao *et al.*, 2013).

5.10.3 Electroforesis

La integridad de los RNAs purificados fue evaluada mediante electroforesis en geles de agarosa (1.5 %) no desnaturalizantes que contenían bromuro de etidio (10 mg/ml). Los geles fueron visualizados utilizando el equipo BIO-RAD ChemiDoc™ MP Imaging System, con el software Image Lab Version 5.1 build 8. Los parámetros evaluados en el gel fueron la ausencia de DNA genómico y la integridad del RNA obtenido analizando bandas bien definidas correspondientes a las subunidades grande y pequeña del RNA ribosomal 28S rRNA y 18S rRNA, respectivamente.

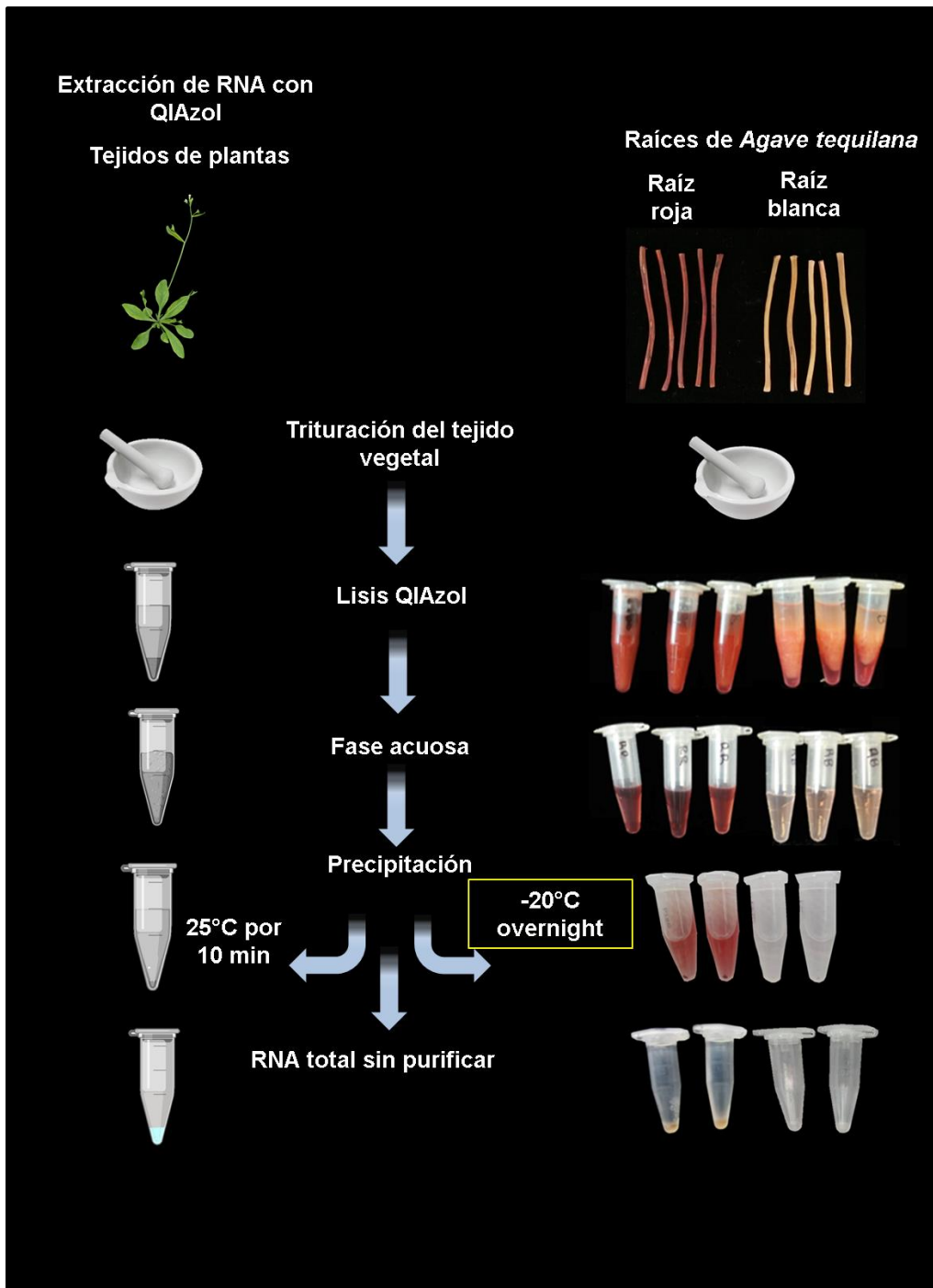


Figura 14. Protocolo de extracción de RNA total de raíces de *A. tequilana*. Se hace una comparación del protocolo sugerido por el fabricante (izquierda) del reactivo de lisis QIAzol (QIAzol Lysis Reagent, Quiagen) y la modificación señalada en color amarillo desarrollada en este estudio (derecha).

5.11 Análisis de expresión de genes involucrados en pared celular en tejido proximal a fibras

5.11.1 Colecta del material

Para un segundo análisis, se colectaron hojas de dos plantas de aproximadamente 1-1.5 años de edad de *A. tequilana* var. Azul, mantenidas en macetas en el Colpos Campus Campeche, que provenían del Rancho Santa Genóveva, Campeche (Figura 15A). En esta ocasión se colectó la primera hoja expandida. De las hojas colectadas se tomaron 2 tipos de muestras: 1) muestras de la sección proximal (A), media (B) y distal de la hoja (C), y 2) fibras que contenían aún tejido adyacente (Figura 15B). Las fibras fueron extraídas de las hojas con la ayuda de una pinza para cultivo *in vitro*, e inmediatamente pulverizadas en un mortero con N₂. El tejido adyacente en las fibras colectadas fue verificado con una tinción con blanco de calcofluor (SIGMA) y observados mediante Microscopía Laser de Barrido Confocal con el microscopio ZEISS AXIO Imager.Z2, (ZEISS, Germany) y una cámara AxioCamMRm, sistema de iluminación de fluorescencia X-Cite Series 120PCQ, fuente de alimentación ZEISS EPLAX VP232-2 (Figura 15C).

5.11.2 Extracción de RNA total y electroforesis

La extracción de RNA total y electroforesis se realizó de la misma manera que se indicó para el análisis anterior de expresión por órganos en la sección 5.10.2

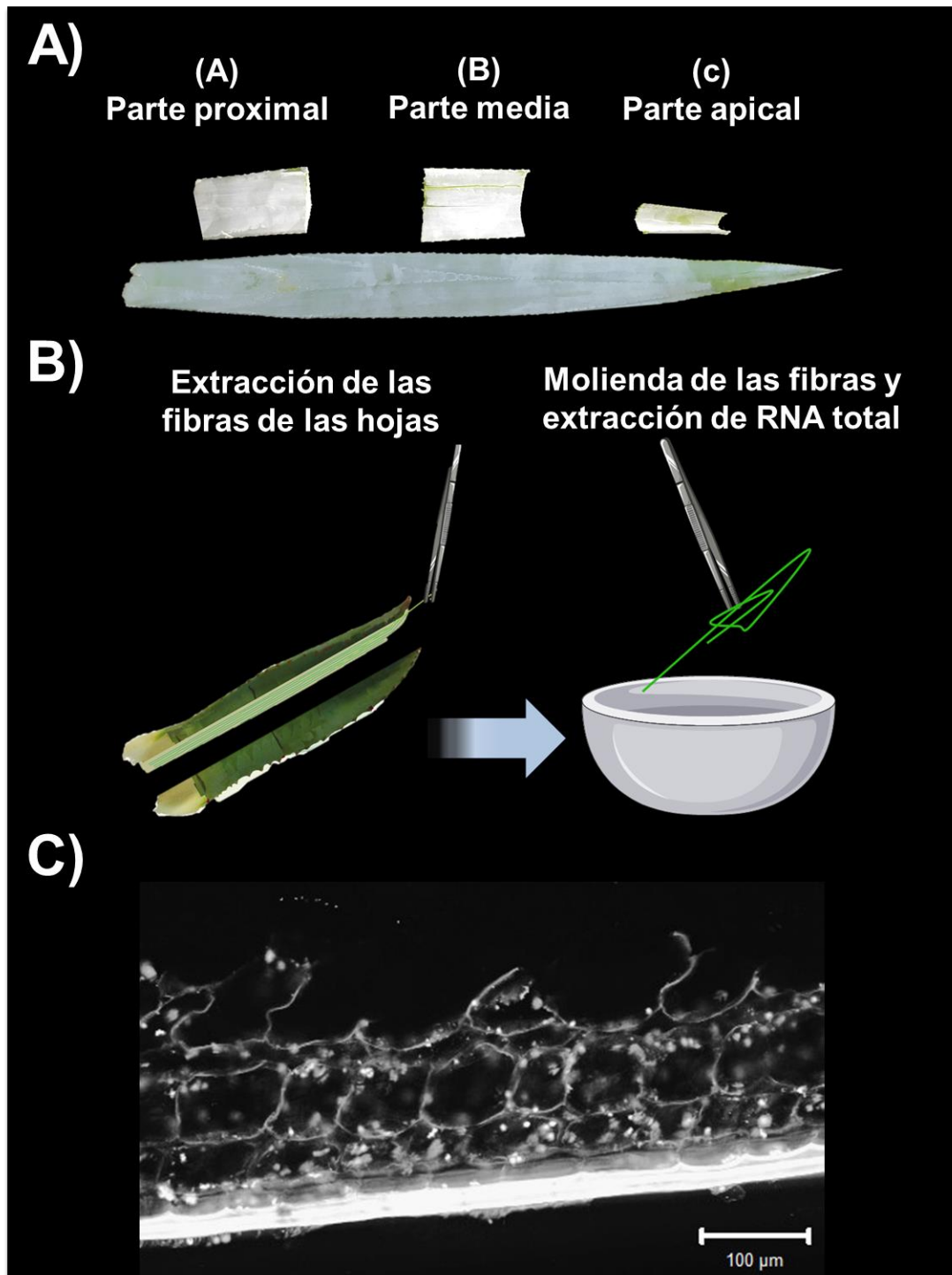
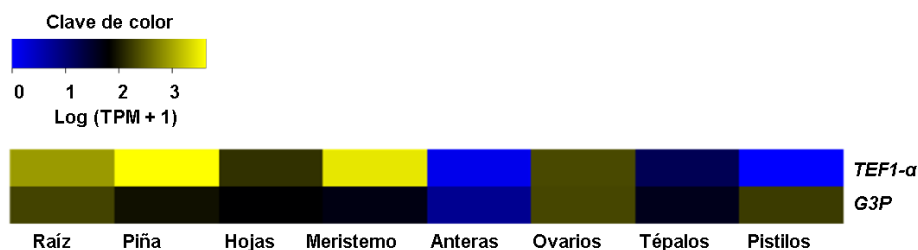


Figura 15. Colecta del material de *A. tequilana* para la extracción de RNA total. A) Secciones utilizadas de las hojas de plantas de aproximadamente 1.5 años de edad. B) Extracción de fibras con una pinza de cultivo in vitro. C) Estructura de la fibra visualizada con Microscopía confocal barra=100 µm.

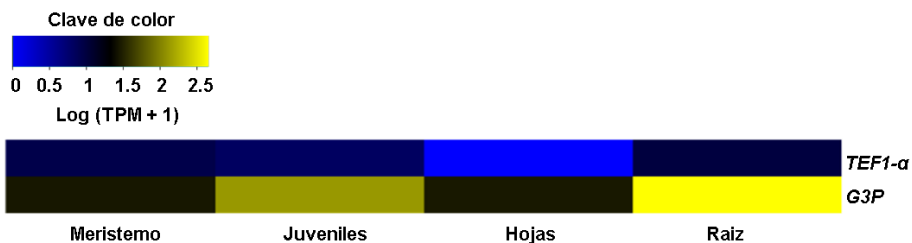
5.12 Diseño de oligonucleótidos para ensayos tipo RT-PCR

Para los análisis de expresión mediante RT-PCR punto final se seleccionaron 3 genes que participan en la biosíntesis de celulosa, *SUS2*, *CESA1* y *CESA7*, y 2 genes para lignina, *CCR2* y *CAD5*. Esta selección se basó en la importancia de estos genes en sus rutas de biosíntesis respectivas. Para los genes de carga (controles endógenos) se eligieron *G3P* y *TEF1- α* , de acuerdo a lo previamente reportado (Souza *et al.*, 2017) y lo obtenido en los mapas de calor para estos genes (Figura 16).

A) Genes control *Agave tequilana* Ávila de Dios *et al.*, 2019



B) Genes control *Agave tequilana* Gross *et al.*, 2013



C) Genes control *Agave deserti* Gross *et al.*, 2013

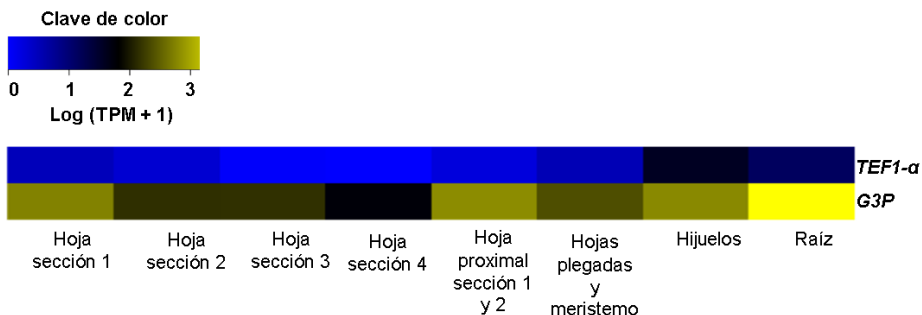


Figura 16. Mapa de calor de los niveles de expresión de los genes control utilizados para los análisis de RT-PCR punto final en los transcriptomas de A) *A. tequilana* (Ávila de Dios *et al.*, 2019), B) *A. tequilana* (Gross *et al.*, 2013) y C) *A. deserti* (Gross *et al.* 2013)

Para cada gen, se usaron las secuencias de sus respectivos Open Reading Frames (ORFs), y cada par de oligonucleótidos se diseñó usando el software Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>; Ye *et al.*, 2012). La secuencia de todos los oligonucleótidos usados se muestra en el cuadro 5.

Cuadro 5. Oligonucleótidos utilizados para los análisis RT-PCR.

Gen	Oligonucleótido	Amplicón	Tm (° C)
<i>G3P</i>	F: ACACTGTCCACGGTTCATGG R: TCCACAACGTA CT CAGCACC	148 pb	64
<i>TEF1-α</i>	F: CCTTGAGGCCCTTGACATGA R: AAGTTGGCAGCCTCCTTAGC	336 pb	62
<i>SUS2</i>	F: GAATGCTGCAACGCCATCAA R: CTTAGAGCTGCCATCTGCGA	269 pb	62
<i>CESA1</i>	F: GTGCCAGATCTGTGGTGACA R: CTGGTCAGACGTGGGATACG	345 pb	62
<i>CESA7</i>	F: AGAGTCGAAGGGGATGACGA R: TCCCGGGCTCAGATAATGGA	331 pb	62
<i>CCR2</i>	F: AAGAACGATCACCTCAGGGC R: CTCGTTGCTGCTTGTCTG	381 pb	62
<i>CAD5</i>	F: TGTGTTCCATCCCTTGGAGC R: CGAGATTGCTTCCAGCAACG	308 pb	62

F: Forward; R: Reverse.

5.13 Análisis de expresión por RT-PCR punto final

5.13.1 Síntesis de DNA complementario (cDNA)

Para llevar a cabo la síntesis de DNA complementario (cDNA) se utilizaron 800 ng de RNA purificado de cada muestra de raíz, piña meristemo y hoja; y 300 ng de RNA de hoja y fibras. El kit comercial utilizado fue SuperScript™ III Reverse Transcriptase de invitrogen, siguiendo las instrucciones del proveedor. Primero se realizó la mezcla con 1 μ L de oligo dT, 1 μ L de la mezcla de dNTP's 10 mM y X μ L de agua DEPC; esta mezcla fue incubada a 65°C durante 5 minutos. Una vez terminada la incubación se colocó inmediatamente en hielo. A esta mezcla se le agregó 4 μ L de 5X First-Strand Buffer (250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM MgCl₂), 1 μ L de 0.1 M DTT, 1 μ L de agua DEPC y 1 μ L de la enzima SuperScript™ III RT (200 unidades/ μ L,) para tener un volumen final de 20 μ L. Esta

mezcla fue incubada a 50°C durante 60 minutos y finalmente a 70°C durante 15 minutos y el cDNA resultante se guardó a -20°C.

5.13.2 Reacciones de la Polimerasa en Cadena (PCRs)

Después de la síntesis de cDNA de todas las muestras de ambos experimentos, se realizaron los PCRs respectivos, de acuerdo con las instrucciones del fabricante. El volumen final de la reacción en todos los casos fue de 20 µL, conteniendo 5.8 µL de agua DEPC, 4 µL de 5X SuperFi™ Buffer, 2 µL de la mezcla de dNTP's 2 mM, 2 µL de la mezcla de oligonucleótidos, 2 µL de cDNA, 4 µL de 5X SuperFi™ GC Enhancer, y 0.2 µL Platinum™ SuperFi™ DNA Polimerasa (2 U/µL). La desnaturalización inicial fue a 98°C por 30 segundos, y 25 ciclos para genes control, o 30 y 32 ciclos para el resto de genes: Desnaturalización 98°C por 30 segundos, alineamiento (ver cuadro 5 para la temperatura de cada gen) por 30 segundos, extensión 72°C por 20 segundos, y un ciclo de extensión final de 72°C por 5 minutos.

Los productos de PCR fueron evaluados mediante electroforesis estándar en un gel de agarosa no desnaturalizante al 1.5% (70 Volts durante 90 min), teñido con Bromuro de etidio (10 mg/ml) y visualizados en el equipo BIO-RAD ChemiDoc™ MP Imaging System, con el software Image Lab Version 5.1 build 8. Como marcador de masa molecular se usó 1 Kb Plus DNA (Invitrogen™).

VI. RESULTADOS Y DISCUSIÓN

6.1 Identificación *in silico* de los genes ortólogos involucrados en las rutas de biosíntesis de celulosa y lignina que están relacionadas con la formación de la pared celular en *A. tequilana* Weber y *A. deserti*

En el presente estudio se identificaron los genes ortólogos involucrados en la biosíntesis de celulosa y lignina en transcriptomas de dos especies de Agave: *A. tequilana*, utilizada para la elaboración de tequila, y *A. deserti*, la cual es usada para la producción de fibras. Como resultado de la minería de datos se obtuvo un total de 25 genes involucrados en la biosíntesis de celulosa y 19 genes involucrados en la biosíntesis de lignina. Cabe resaltar que en ambas especies se encontró la misma cantidad de genes para cada ruta de biosíntesis.

6.1.1 Ensamblado *de novo*: Anotación y categorización funcional del transcriptoma de *Agave tequilana* y *Agave deserti*

El proceso de anotación para aquellos unigenes no redundantes en Agave, en los que se identificó y corrigió el correspondiente marco de lectura abierto, incluyó una clasificación funcional y una búsqueda de secuencias similares (homólogos) pertenecientes a los conjuntos de proteínas predichas para diferentes plantas eudicotiledóneas y monocotiledóneas y cuyos genomas han sido completamente secuenciados (Figura 17). Las especies seleccionadas fueron *Vitis vinifera*, *Amborella trichopoda*, *Solanum lycopersicum*, *Zea mays* y *Arabidopsis thaliana*. Cada una de las especies usadas para la anotación pertenece a un diferente orden taxonómico lo cual fue un factor clave en la selección de estos genomas para la anotación de los transcriptomas ensamblados de agave.

Del total de unigenes ensamblados (816,724 transcritos únicos en *A. tequilana* y 608,962 transcritos únicos en *A. deserti*), se identificó y corrigió el correspondiente marco de lectura dando un total de 812,328 unigenes no redundantes en *A. tequilana* y 605,545 unigenes no redundantes en *A. deserti*. En línea con lo esperado, una vez que se corrigieron los marcos de lectura, la mayoría de las

proteínas traducidas a partir de las regiones codificantes identificadas en los unigenes no redundantes (un total de 156,063 proteínas en *A. tequilana* y 136,804 en *A. deserti*), mostraron similitud con un valor significativo (valor-e < 10⁻⁵) con proteínas en al menos una de las cinco especies de plantas usadas para la anotación funcional. El porcentaje de unigenes anotados en función de cada una de las especies oscila entre el 81.94 y el 92.01% (Figura 18).

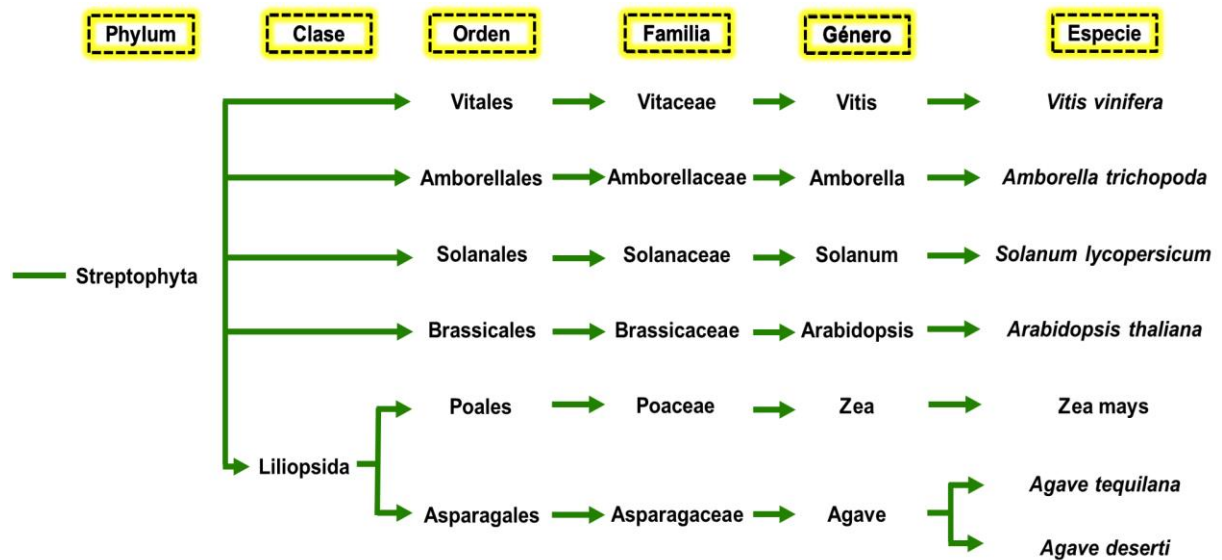


Figura 17. Relaciones filogenéticas de las especies utilizadas para el proceso de anotación funcional en *A. tequilana* y *A. deserti*.

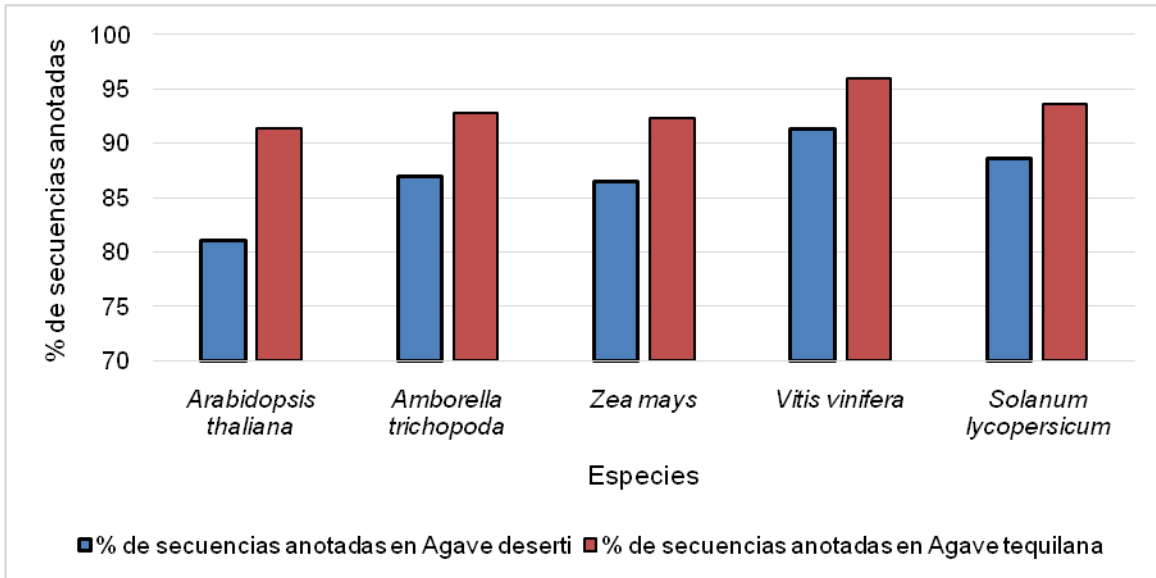


Figura 18. Porcentaje de unigenes de *A. tequilana* y *A. deserti* anotados en función de la homología de los mismos con proteínas en el genoma de diferentes especies de plantas angiospermas.

6.1.2 Genes ortólogos en *A. tequilana* y *A. deserti* involucrados en la biosíntesis de celulosa

El número total de genes ortólogos identificados *in silico* en la biosíntesis de celulosa en cada especie de *Agave* fue de 25. El cuadro 6 muestra la comparación del número de genes en cada familia génica involucrada en la biosíntesis de celulosa en la planta modelo *Arabidopsis thaliana* contra los genes identificados en *A. tequilana* y *A. deserti*.

Cuadro 6. Comparación del número de genes por familias involucradas en la biosíntesis de celulosa en *A. tequilana*, *A. deserti* y en la planta modelo *A. thaliana*.

Gen	<i>Arabidopsis thaliana</i>	<i>Agave tequilana</i>	<i>Agave deserti</i>
SPP	4	1	1
SPS	4	2	2
FK	7	5	5
PGI	1	1	1
PGM	3	2	2
SUS	6	3	3
H XK	3	2	2
UGP	3	3	3
CESA	10	6	6

La familia génica más importante en la ruta de celulosa es la enzima *celulosa sintasa* así que los siguientes análisis se centrarán solo en ellos. En *A. thaliana* se han reportado 10 genes *CESAs*, de las cuales a partir de los análisis de ortología se encontraron 6 genes ortólogos en ambas especies de agave. Los nombres para gen de Agave se asignaron siguiendo la nomenclatura de los genes homólogos de *A. thaliana*. En *A. tequilana* fueron: *AqCESA1* (UN017499), *AqCESA3* (UN003733), *AqCESA4* (UN045020), *AqCESA5* (UN052428), *AqCESA7* (UN068270) y *AqCESA8* (UN481687). En *A. deserti* fueron *AdCESA1* (UN005336), *AdCESA3* (UN024289), *AdCESA4* (UN031321), *AdCESA5* (UN085584), *AdCESA7* (UN100087) y *AdCESA8* (UN050033).

Recientemente se reportaron 5 genes *CESA* en el agave híbrido *A. H11648*, *A. americana*, *A. deserti* y *A. tequilana* (Huang *et al.*, 2019). El número de genes identificados en este estudio en *A. tequilana* y *A. deserti* es mayor que lo reportado por Huang en el 2019 para estas mismas dos especies. Suponemos que esta diferencia puede estar ser debida a que en nuestro estudio utilizamos mayor cantidad y diversidad de transcriptomas que los usados por Huang y colaboradores. El paisaje transcriptómico fue mayor por lo tanto, algunos genes *CESA* no reportados pudieron estar presentes y ser identificados en nuestro análisis de ortología. Muy probablemente la secuenciación del genoma de alguna

especie de agave reportará mayor número de genes *CESA*, considerando la ploidía que presentan los agaves, y otras especies del mismo orden taxonómico como *Asparagus officinalis* en la cual se han reportado 12 genes *CESA* (Huang *et al.*, 2019).

La longitud de los marcos de lectura abierto (ORF) de los *CESAs* de *A. tequilana* y *A. deserti* son de 2853 a 3270 nucleótidos y codifican polipéptidos relativamente grandes de 353 a 370 aminoácidos (Cuadro 7). Esto es similar comparado con el tamaño de los *CESAs* de *A. thaliana* donde la longitud es de 2958 a 3267 nucleótidos.

Nuestro estudio también identificó en agave genes ortólogos para el gen *CESA* de menor longitud en *A. thaliana*, el gen *CESA8* ([AT2G25540](#)); en ambas especies de agave, los ortólogos identificados tienen el mismo tamaño que el reportado para esa dicotiledónea. En *Arabidopsis*, los *CESAs* de mayor tamaño corresponden a *CESA9* ([AT2G21770](#)) con 3267 nucleótidos y *CESA5* ([AT5G09870](#)) con 3210 nucleótidos. En *A. tequilana* y *A. deserti*, el ortólogo *CESA5* tiene 3270 nucleótidos de longitud, ligeramente mayor al reportado en *Arabidopsis*. No se identificaron en los transcriptomas analizados genes ortólogos para *CESA9*, por lo que posible saber si existen genes codificantes para *CESA* de mayor tamaño en las especies de agaves analizadas.

Cuadro 7. Genes ortólogos identificados en la ruta de biosíntesis de celulosa en los transcriptomas *A. tequilana* y *A. deserti*.

Biosíntesis de celulosa						
Gen	<i>Agave tequilana</i>			<i>Agave deserti</i>		
	Identificador	Longitud nucleótidos	Longitud aminoácidos	Identificador	Longitud nucleótidos	Longitud aminoácidos
SPP	UN009406	1272	424	UN024948	1272	424
SPS1	UN001860	3165	1055	UN001860	3165	1055
SPS2	UN002490	3246	1082	UN002490	3246	1082
FK1	UN110091	648	216	UN045881	930	310
FK2	UN53179	933	311	UN274164	933	311
FK3	UN066338	930	310	UN041662	930	310
FK4	UN026769	939	313	UN014364	939	313
FK5	UN72180	1011	337	UN033414	1011	337
PGI	UN273756	1902	634	UN040814	1839	613
PGM1	UN022016	1869	623	UN016551	1869	623
PGM2	UN034431	1650	550	UN047077	1722	574
SUS1	UN010505	2520	840	UN006343	2520	840
SUS2	UN066280	2451	817	UN005890	2436	812
SUS3	UN110284	2415	805	UN074570	2415	805
HXK1	UN037472	1584	527	UN017047	1464	488
HXK2	UN028066	1350	450	UN015408	1485	495
UGP1	UN019706	1419	473	UN015809	1419	473
UGP2	UN036974	1422	474	UN085461	1563	521
UGP3	UN005147	2292	764	UN001173	2559	853
CESA1	UN017499	3261	1087	UN005336	3261	1087
CESA3	UN003733	3177	1059	UN024289	3177	1059
CESA4	UN045020	3171	1057	UN031321	3171	1057
CESA5	UN052428	3270	1090	UN085584	3267	1089
CESA7	UN068270	3174	1058	UN100087	3216	1072
CESA8	UN481687	3060	1020	UN050033	2853	951

Se realizó además un análisis de identidad y similitud de las secuencias de los genes *HAqCESA* (*A. tequilana*; Huang *et al.*, 2019) reportados y los genes *AqCESA* y *AdCESA* de este estudio (Cuadro 8). Para *AqCESA1* y *AdCESA1* la similitud es de 100 % con *HAqCESA1*. En *AqCESA3* y *AdCESA3* la similitud es de 99.81 % y 99.53 % con *HAqCESA3*.

Para el caso de *AqCESA4* tiene una similitud de 98.77 % con *HAqCESA4* y con *AdCESA4* tiene un 99.91 % de similitud, es interesante este resultado por que el gen que presenta mayor homología con *HAqCESA* es de *A. deserti*, este resultado puede ser debido a que la identificación de dichos genes fue realizada en diferentes transcriptomas de *A. tequilana* y no a que las secuencias estén incompletas debido a que todas las secuencias para *AqCESA4*, *AdCESA4* y *HAqCESA4* tienen 1057 aminoácidos de longitud.

Para *AqCESA7* y *AdCESA7* se encuentran los porcentajes más bajos de similitud respecto a *HAqCESA7* (Cuadro 8). Dichos porcentajes son 84.98 para *AqCESA7* y 85.05 de *AdCESA7*, respectivamente. En este caso el resultado coincide con *AqCESA4* y *AdCESA4* donde *A. deserti* presenta la mayor similitud respecto a *HAqCESA7*.

Cuadro 8. Homología entre genes *CESA* identificadas en el análisis de ortología en *A. tequilana* y *A. deserti* y los genes *CESA* de *A. tequilana* reportados por Huang *et al.*, 2019. Los valores muestran el porcentaje de identidades (número izquierdo) y similitudes (número derecho) de aminoácidos.

	HAqCESA1 (1087 aa)	HAqCESA3 (1059 aa)	HAqCESA4 (1057 aa)	HAqCESA5 (1090 aa)	HAqCESA7 (1050 aa)
AqCESA1 (1087 aa)	99.82/100.00	67.59/75.51	62.00/71.39	67.17/76.54	63.92/72.16
AqCESA3 (1059)	67.41/75.33	99.81/99.81	62.18/71.28	63.47/72.80	64.38/71.77
AqCESA4 (1057)	62.00/71.48	63.25/71.90	98.30/98.77	62.64/71.76	64.84/73.25
AqCESA5 (1090)	66.99/76.45	63.20/72.54	62.90/71.76	98.81/99.27	68.28/76.85
AqCESA7 (1058)	63.01/72.64	62.09/71.36	61.38/72.57	71.66/81.29	78.89/84.98
AqCESA8 (1020)	57.31/67.49	59.69/70.34	59.87/70.66	56.32/67.54	56.75/66.49
AdCESA1 (1087)	99.72/100	67.50/75.51	62.09/71.39	67.17/76.54	63.83/72.16
AdCESA3 (1059)	67.41/75.24	99.24/99.53	62.09/71.54	63.56/72.89	64.29/71.68
AdCESA4 (1057)	62.00/71.39	62.18/71.36	99.91/99.91	62.38/71.42	64.75/73.52
AdCESA5 (1089)	66.10/75.91	61.88/71.48	60.30/70.03	92.29/95.60	66.67/75.34
AdCESA7 (1072)	63.50/71.50	64.33/72.14	62.40/71.69	68.59/76.51	81.10/85.05
AdCESA8 (951)	56.90/65.83	58.29/67.59	59.33/69.22	56.27/66.40	56.66/65.97

Múltiples alineamientos de las secuencias de proteínas CESA revelaron que los AqCESAs y los AdCESAs tienen del 67 – 96 % de similitud con CESAs de otras especies de plantas (Cuadro 7). Además, en *A. tequilana* el alineamiento reveló que AqCESA1 es el que presenta la mayor homología, con 67-96 % de similitud y en *A. deserti* AdCESA 67-96 % de similitud con los CESAs de otras plantas.

Como se observa en el cuadro 9 AqCESA1 y AdCESA1 tienen el mayor porcentaje de identidad y similitud con AoCESA1 de la planta monocotiledónea *A. officinalis* respecto a otras especies monocotiledóneas como *Z. mays* y dicotiledóneas como *A. thaliana*. Así mismo AqCESA4 y AdCESA4 presentan mayor homología con AoCESA4. Además, AqCESA5 y AdCESA5 presentan un 95 % de similitud con AoCESA8.

De igual forma la homología de AqCESA8 y AdCESA8 es mayor con AoCESA2 teniendo un 90% de similitud. Otro gen que presenta mayor porcentaje de identidad y similitud con otra planta monocotiledónea es CESA7 de ambas especies de Agave la cual su identidad es de 84 %. Por último, CESA3 de ambas especies de Agave tiene 77 % de identidad y 84 % de similitud con AtCESA3 de la planta dicotiledónea *A. thaliana*. Todos estos análisis confirman que los genes *CESA* están altamente conservados en plantas, y especialmente entre monocotiledóneas. Ello también sugiere una función biológica conservada en la formación de pared celular en agaves.

Cuadro 9. Secuencias homologas entre las secuencias de aminoácidos ortólogas de AqCESA y AdCESA y los CESAs de otras plantas. Los valores muestran el porcentaje de identidades (número izquierdo) y similitudes (número derecho) de aminoácidos. (Ad) *A. deserti*, (Aq) *A. tequilana*, (Ao) *A. officinalis*, (Zm) *Z. mays* y (At) *A. thaliana*.

	AoCESA1	AoCESA2	AoCESA4	AoCESA8	ZmCESA1	ZmCESA3	ZmCESA12	AtCESA3	AtCESA5	AtCESA7
AqCESA1	93/96	58/67	60/71	67/76	87/91	86/91	63/72	69/78	62/72	62/71
AqCESA3	66/74	58/68	61/71	62/71	66/74	66/75	64/72	77/84	60/70	63/72
AqCESA4	61/71	62/72	91/94	64/73	60/71	61/71	64/73	61/69	62/73	64/73
AqCESA5	65/75	58/68	64/74	92/95	64/74	64/74	68/77	63/72	70/81	66/76
AqCESA7	61/72	56/67	62/72	68/78	60/71	61/72	75/84	61/72	60/75	73/83
AqCESA8	56/67	85/89	61/72	56/67	57/67	58/68	56/67	57/68	55/68	56/67
AdCESA1	93/96	58/67	60/71	67/76	87/91	86/91	63/72	69/78	62/72	62/71
AdCESA3	66/74	58/68	61/71	62/71	66/74	66/75	64/72	77/84	60/70	63/71
AdCESA4	61/71	62/72	93/95	63/73	61/71	61/71	64/73	60/69	61/72	63/73
AdCESA5	64/74	56/66	62/72	90/94	63/74	63/74	66/76	61/72	70/80	64/75
AdCESA7	63/72	58/67	62/73	69/78	64/72	64/72	75/83	63/72	66/78	74/82
AdCESA8	56/65	88/90	60/70	56/66	57/66	58/67	56/65	56/66	54/66	56/66

6.1.3 Motivos funcionales y estructurales de los genes *CESA* en *A. tequilana* Weber y *A. deserti*

Todas las proteínas CESA en *A. tequilana* y *A. deserti* poseen ocho dominios transmembranales: dos dominios en el extremo N-terminal y seis próximos al extremo C-terminal. Estas proteínas también presentan los motivos conservados D, DXD, D, QXXRW (Figura 19). Ambos tipos de dominios han sido previamente identificados en CESAs de *Arabidopsis* (Grenville-Briggs *et al.*, 2008).

Recientemente se reportó que las proteínas CESA que participan en la formación de pared celular primaria y pared celular secundaria se distinguen por presentar un motivo ZnF. En este trabajo, se realizó la identificación de tales motivos mediante alineamientos de aminoácidos de los ortólogos identificados para ambas especies de agave.

El motivo ha sido reportado para diferenciar las proteínas CESA de aquellos genes tipo CESA o también llamados CSL (CESA-like) (Hamann *et al.*, 2004; Heidari *et al.*, 2019). Este motivo estuvo presente en las CESA de *A. tequilana* (AqCESA1, AqCESA5, AqCESA7) y *A. deserti* (AdCESA1, AdCESA5). Además, las proteínas CESA de *A. tequilana* (AqCESA3, AqCESA4, AqCESA8) y *A. deserti* (AdCESA3, AdCESA4, AdCESA7, AdCESA8) poseen el motivo “CXXC”, donde los residuos intermedios son variables (Figura 19).

Otro motivo conservado en las proteínas CESA de agave es SVICEXWFA, hallado en el segundo dominio transmembranal. Este dominio está presente en las proteínas CESA específicas de pared celular primaria y secundaria en las dos especies de agave. Solo en algunas proteínas de *A. tequilana* (AqCESA1, AqCESA7) y *A. deserti* (AdCESA1, AdCESA8) el motivo presentó residuos variables dando lugar a “SXXCEXWFA”, donde XX pueden ser los aminoácidos VI e II.

Estos motivos, CQIC y SVICEXWFA, han sido hallados en plantas monocotiledóneas como *Zea mays*, *Oryza sativa*, *Triticum aestivum* y *Hordeum vulgare*, así como en las plantas dicotiledóneas reportadas como *Arabidopsis thaliana*, *Populus trichocarpa*, *Beta vulgaris*, *Eucalyptus grandis*, *Glycine max*, *Gossypium hirsutum* y *Solanum tuberosum* (Kaur *et al.*, 2016). La importancia de la identificación de estos motivos conservados ayudará a la identificación de aquellas CESAs que son específicas en la formación de pared primaria y secundaria. En agave, estos motivos serán útiles para estudios sobre la estructura, función y evolución de la familia CESA.

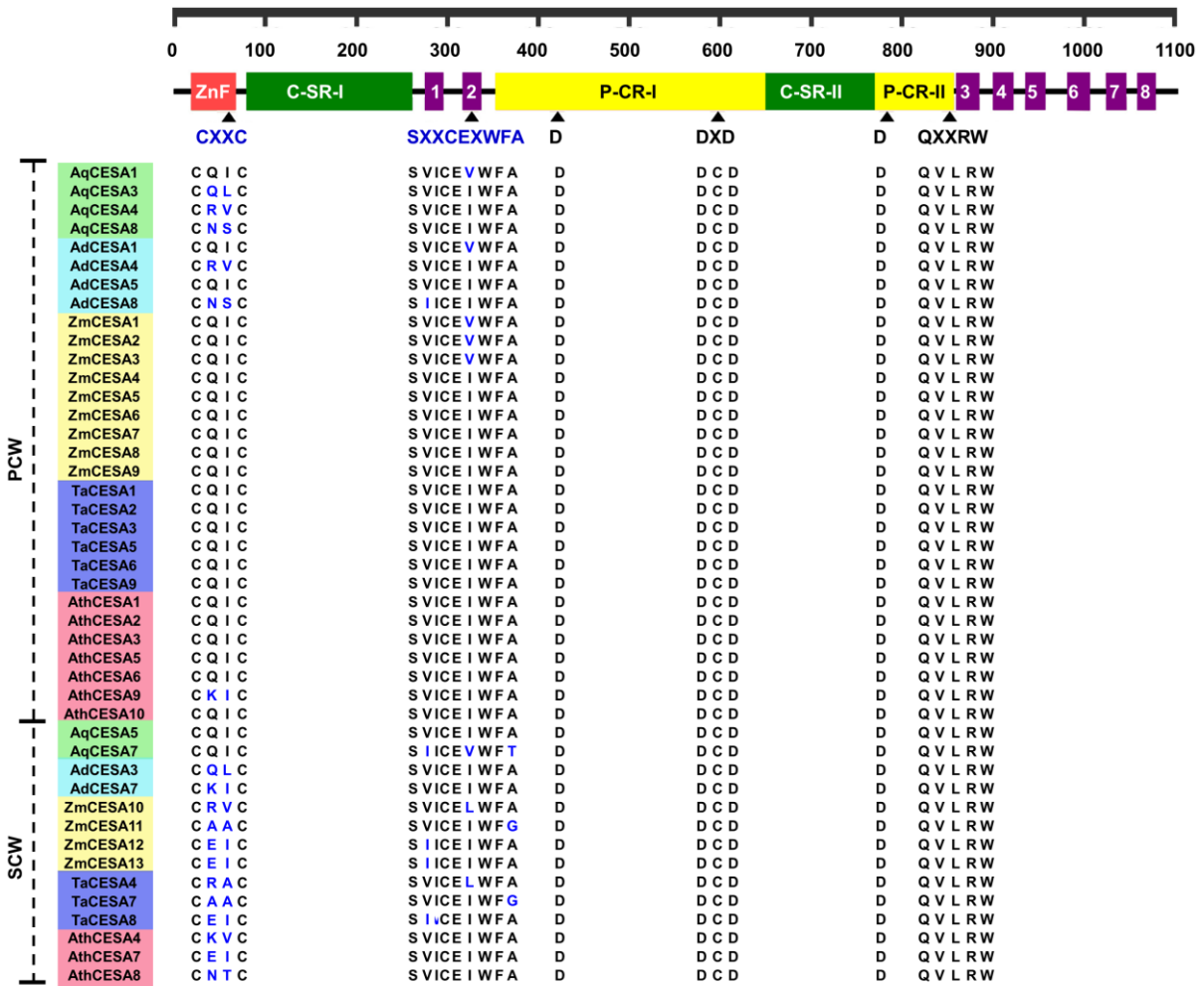


Figura 19. Alineamiento de las secuencias de los motivos que diferencian los genes ortólogos para celulosa sintasa en la pared celular primaria (PCW) y la pared celular secundaria (SCW) de diferentes especies. Los cambios de aminoácidos de los motivos se muestran en color azul. (Ad) *A. deserti*, (Aq) *A. tequilana*, (Ta) *T. aestivum*, (Zm) *Z. mays* y (Ath) *A. thaliana*.

6.1.4 Análisis filogenético de los genes *CESA* de *Agave*

Todas las proteínas *CESA* halladas en *Agave tequilana* (6) y *Agave deserti* (6) en este estudio, y aquellas otras plantas monocotiledóneas como *Zea mays* (13), *Oryza sativa* (11), *Asparagus officinalis* (12), *Triticum aestivum* (9) y *Hordeum vulgare* (8), así como las plantas dicotiledóneas reportadas como *Arabidopsis thaliana* (10), *Populus trichocarpa* (1), *Beta vulgaris* (1), *Eucalyptus grandis* (1), *Glycine max* (1), *Gossypium hirsutum* (1) y *Solanum tuberosum* (1) fueron utilizadas en el análisis filogenético que se muestra en la figura 20. El árbol muestra una agrupación de los ortólogos de monocotiledóneas y dicotiledóneas en diferentes clados. Las longitudes de las ramas, que son indicativas de las distancias evolutivas, se utilizaron para interpretar el árbol filogenético. Los genes parálogos de las especies utilizadas para el análisis filogenético se agruparon en diferentes clados a las de sus ortólogos respectivos, sugiriendo que la divergencia de los genes *CESA* se produjo antes de la separación de monocotiledóneas y eudicotiledóneas (figura 20).

El árbol filogenético para genes *CESA* se integra por tres clados, dos de los cuales agrupan los genes *CESA* que participan en la formación de pared primaria, y un tercer clado que agrupa los genes *CESA* específicos de la formación de pared celular secundaria. Los genes *CESA1*, *CESA3* y *CESA5* de ambas especies de *Agave* fueron hallados como *ZmCESA1*, *ZmCESA2*, *ZmCESA3*, *ZmCESA4*, *ZmCESA5*, *ZmCESA6*, *ZmCESA7*, *ZmCESA8*, *ZmCESA9* de *Z. mays* y *AtCESA1*, *AtCESA2*, *AtCESA3*, *AtCESA5*, *AtCESA6*, *AtCESA9*, *AtCESA10* de *A. thaliana* (Figura 20).

Por otro lado, *CESA4*, *CESA7* y *CESA8* de *A. tequilana* y *A. deserti*, se hallaron dentro del clado para *CESAs* de pared celular secundaria, al igual que *CESA7* de *Arabidopsis* el gen más representativo que codifica para una *CESA* con actividad enzimática durante la formación de pared secundaria (Sakamoto *et al.*, 2018).

Para el caso de *AqCESA4* y *AdCESA4* se agrupa en el mismo clado que

AoCESA4 de *A. officinalis*. Además, *AqCESA8* y *AdCESA8* se agrupa en el mismo clado con *AoCESA2*. Igualmente, *AqCESA7* y *AdCESA7* en los análisis de identidad y similitud (Cuadro 9), presentan mayor homología con *ZmCESA12*, siendo el mismo resultado en el árbol filogenético debido a que se agrupan en el mismo clado. Los resultados del árbol filogenético son congruentes con el análisis de identidad y similitud (Cuadro 9 y Figura 20) y de motivos realizado (Figura 19) y permitiría por primera vez proponer una clasificación para genes *CESA* en agaves sobre su función en la formación de pared primaria o secundaria.

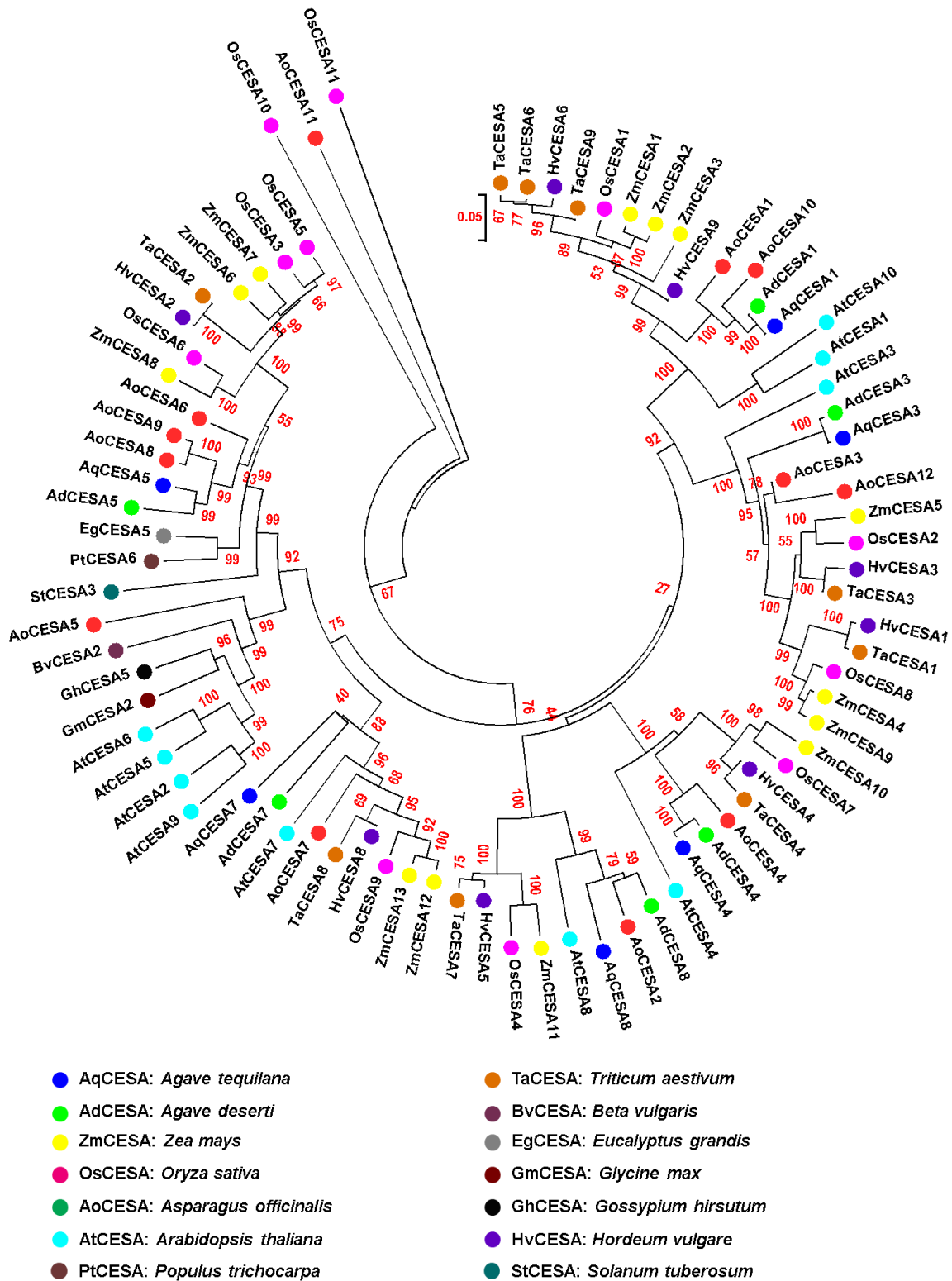


Figura 20. Árbol filogenético no enraizado de los genes *CESA* de *Agave tequilana* (*AqCESA*), *Agave deserti* (*AdCESA*) y de otras especies de plantas en las cuales se han reportado los genes *CESA*.

En el cuadro 10 se muestran los genes ortólogos de *celulosa sintasa* en diferentes especies monocotiledóneas y eudicotiledóneas. Para la pared primaria los genes encontrados en *A. tequilana* y *A. deserti* coinciden con los reportados en maíz, arroz, trigo y *A. thaliana*. Además, los genes identificados *in silico* en la formación de pared secundaria en *A. tequilana* y *A. deserti* coinciden con los reportados en *A. thaliana* y *T. aestivum*.

Cuadro 10. Genes *AqCesA* y *AdCESA* y sus ortólogos de Arabidopsis, trigo, maíz y arroz involucrados en la formación de la pared celular primaria (PCW) y la pared celular secundaria (SCW).

	<i>Zea mays</i>	<i>Oryza sativa</i>	<i>Arabidopsis thaliana</i>	<i>Triticum aestivum</i>	<i>Agave tequilana</i>	<i>Agave deserti</i>
	CESA1, 2 y 3	CESA1	CESA1 y 10	CESA5, 6 y 9	CESA1	CESA1
PCP	CESA4, 5, y 9	CESA2, 8, 10 y 11	CESA3	CESA1 y 3	CESA3	CESA3
	CESA6, 7, y 8	CESA3, 5, y 6	CESA2, 5, 6, y 9	CESA2	CESA5	CESA5
	CESA10	CESA7	CESA4	CESA4	CESA4	CESA4
PCS	CESA12 y 13	CESA9	CESA7	CESA7	CESA7	CESA7
	CESA11	CESA4	CESA8	CESA8	CESA8	CESA8

6.1.5 Genes ortólogos en *A. tequilana* y *A. deserti* involucrados en la biosíntesis de lignina

El análisis de ortología reveló que en los transcriptomas de agave analizados existen 19 genes ortólogos para aquellos genes de *Arabidopsis* que participan en cada uno de los pasos de la ruta de biosíntesis de lignina (Wanner *et al.*, 1995; Cochrane *et al.*, 2004). En ambas especies de agaves fue posible identificar genes ortólogos para las 11 familias génicas involucradas (Cuadro 11). En 5 familias se halló el mismo número de genes que los reportados en *Arabidopsis*. En 4 familias se halló un número menor, siendo el caso extremo *4CL* para la cual *Arabidopsis* posee 4 miembros, mientras que en *A. tequilana* y *A. deserti* solo se hallaron expresados 1 gen ortólogo. Un caso interesante fue *COMT* para la cual en Agave se hallaron dos genes expresados, lo cual contrasta con la familia unigénica en *Arabidopsis*. La familia de genes más importante de esta ruta es *CAD*. En *A. thaliana* se han reportado 9 *CADs* en su genoma, de las cuales a partir de los análisis de ortología se encontraron 7 genes ortólogos expresados en ambas especies de Agave (cuadro 11).

Cuadro 11. Comparación del número de genes por familias involucradas en la biosíntesis de lignina en *A. tequilana*, *A. deserti* y en la planta modelo *A. thaliana*.

Gen	<i>Arabidopsis thaliana</i>	<i>Agave tequilana</i>	<i>Agave deserti</i>
<i>PAL</i>	4	2	2
<i>C4H</i>	1	1	1
<i>4CL</i>	4	1	1
<i>CSE</i>	1	1	1
<i>HCT</i>	1	1	1
<i>C3H</i>	1	1	1
<i>CcoAOMT</i>	1	1	1
<i>CCR</i>	2	1	1
<i>F5H</i>	1	1	1
<i>COMT</i>	1	2	2
<i>CAD</i>	9	7	7

Posteriormente al tamaño de las familias de genes ortólogos expresados en agave, respecto a *Arabidopsis*, se realizó una comparación respecto a tamaño de los genes ortólogos y tamaño de las proteínas codificadas en las 2 especies de Agaves analizadas (Cuadro 12). De las 11 familias génicas, en 7 de ellas existen pequeñas diferencias en cuanto a tamaño de la proteína, excepto para CcoAOMT de *A. deserti* donde el ortólogo identificado codifica una proteína 31 aa más pequeña que la hallada en *A. tequilana*. Respecto a CAD, una de las enzimas clave de la ruta, la longitud de los marcos de lectura abiertos de los CADs en Agave varía de 1059 a 1110 nucleótidos, y codifican polipéptidos de 353 a 370 aminoácidos (Cuadro 12). En *A. thaliana* el gen CAD que tiene mayor longitud es CAD2 ([AT2G21730](#)) con 1131 nucleótidos mientras que *A. tequilana* y *A. deserti* el gen mayor longitud es CAD6 con 1110 nucleótidos. El gen CAD con la menor longitud en *Arabidopsis* es CAD1 ([AT1G72680](#)) con 1068 nucleótidos coincidiendo con los genes de menor longitud en ambas especies de Agave, *AqCAD1* y *AdCAD1* (Cuadro 12).

Cuadro 12. Genes ortólogos identificados en la ruta de biosíntesis de lignina en los transcriptomas *A. tequilana* y *A. deserti*.

Biosíntesis de lignina						
<i>Agave tequilana</i>				<i>Agave deserti</i>		
Gen	Identificador	Longitud nucleótidos	Longitud aminoácidos	Identificador	Longitud nucleótidos	Longitud aminoácidos
PAL1	UN04780	2106	702	UN030045	2121	707
PAL2	UN161333	2121	707	UN062605	2049	683
C4H	UN023970	1506	502	UN137282	1509	503
4CL	UN017824	1656	552	UN029379	1656	552
CSE	UN060774	972	324	UN027048	987	329
HCT1	UN018939	1323	441	UN032672	1326	442
HCT2	UN035132	1299	433	UN007333	1332	444
C3H	UN065853	1536	512	UN025566	1536	512
CcoAOMT	UN068705	768	256	UN080839	675	225
CCR1	UN071278	978	326	UN058652	978	326
CCR2	UN060336	1011	337	UN148129	1011	337
F5H	UN034572	1542	514	UN013652	1545	515
COMT1	UN034645	1089	363	UN117387	1089	363
COMT2	UN057692	1092	364	UN036511	1092	364
CAD1	UN045407	1059	353	UN047496	1059	353
CAD2	UN057761	1089	363	UN055860	1089	363
CAD3	UN146273	1062	354	UN105918	1062	354
CAD4	UN050838	1062	354	UN121435	1062	354
CAD5	UN049346	1068	356	UN160915	1068	356
CAD6	UN050300	1110	370	UN014798	1110	370
CAD7	UN090516	1062	354	UN041364	1062	354

6.1.6 Conservación del gen CAD en Agave respecto a plantas monocotiledóneas y dicotiledóneas

Dada la importancia del gen CAD en la ruta que conduce a la síntesis de los monómeros de lignina, se realizó un análisis de identidad y similitud en las secuencias de los CAD ortólogos identificados en ambas especies de agave y los CAD *bona fide* reportados en otras especies vegetales.

Los alineamientos múltiples de las secuencias de proteínas CAD revelaron que AqCADs y AdCADs tienen un 59-88% de similitud respecto a las secuencias de CAD *bona fide* de otras especies de plantas (Cuadro 13). Dentro de los ortólogos para Agave, AqCAD5 y AdCAD5 son las proteínas que presenta la mayor homología con 79-88% de similitud, sugiriendo que este gen pudiera codificar el CAD *bona fide* para Agave.

Cuadro 13. Secuencias homologas entre las secuencias de aminoácidos ortólogas de AqCAD y AdCAD y los CADs de otras plantas. Los valores muestran el porcentaje de identidades (número izquierdo) y similitudes (número derecho) de aminoácidos.

	PtCAD1	AtCAD4	AtCAD5	NtCAD1	NtCAD2	ZmCAD3	SbCAD2	TaCAD1	OsCAD2	BdCAD6
AqCad1	48/64	43/61	44/62	45/63	45/63	41/59	41/59	42/61	41/60	41/61
AqCad2	47/65	45/63	45/63	47/65	47/66	44/62	44/63	46/63	45/62	44/63
AqCad3	51/69	48/66	51/67	50/68	49/68	47/66	47/66	48/67	48/65	49/67
AqCad4	51/68	46/63	48/65	49/67	49/68	47/64	47/64	51/67	48/65	49/66
AqCad5	68/79	73/82	75/83	76/84	74/83	79/86	79/86	80/88	80/88	79/88
AqCad6	49/66	46/63	47/64	48/65	48/65	45/62	45/62	45/62	45/62	45/62
AqCad7	49/64	45/59	46/62	46/62	46/62	42/59	42/59	43/60	42/59	43/60
AdCad1	48/64	43/61	44/62	45/63	45/63	41/59	40/59	42/61	41/60	42/61
AdCad2	47/65	45/63	45/63	47/65	47/65	44/62	44/63	45/63	45/62	45/63
AdCad3	51/69	48/66	50/67	50/68	49/68	47/66	47/66	48/67	48/65	49/67
AdCad4	50/66	46/62	48/64	49/65	49/66	48/63	48/63	50/65	49/64	49/65
AdCad5	67/79	73/82	74/83	76/84	74/83	79/86	78/86	81/88	79/88	79/88
AdCad6	49/66	46/63	47/64	48/65	48/65	45/62	45/62	45/62	45/62	45/62
AdCad7	49/65	46/59	47/62	47/62	47/62	43/59	42/59	44/60	42/59	43/61

6.1.7 Motivos funcionales y estructurales de los genes *CAD* en *A. tequilana* Weber y *A. deserti*

Los CAD de plantas contienen motivos de unión a Zn altamente conservados (motivo de unión Zn1 PARA Zn²⁺ catalítico y el motivo de unión Zn2 PARA Zn²⁺ estructural) y firmas de unión a NADPH (Figura 21). Estos motivos también están conservados en AqCAD y AdCAD. La secuencia consenso de unión a Zn1 “GHE(X)₂G(X)₅G(X)₂V” esta conservada en todos los AqCAD y AdCAD. El consenso “GD(X)₁₀C(X)₂C(X)₂C(X)₇C” se conserva en todos los AqCAD y AdCAD. Para las firmas de unión a NADPH (Park *et al.*, 2018), “G(X)G(X)₂G”, “SSS(X)₂K” y “G(X)₄P”, están conservados en todas las proteínas CAD de *A. tequilana* y *A. deserti*. La identificación de una conservación muy alta de todos los dominios analizados sugiere la existencia en agave de CADs con funciones biológicas, muy posiblemente también conservadas en la síntesis de los monómeros de lignina.

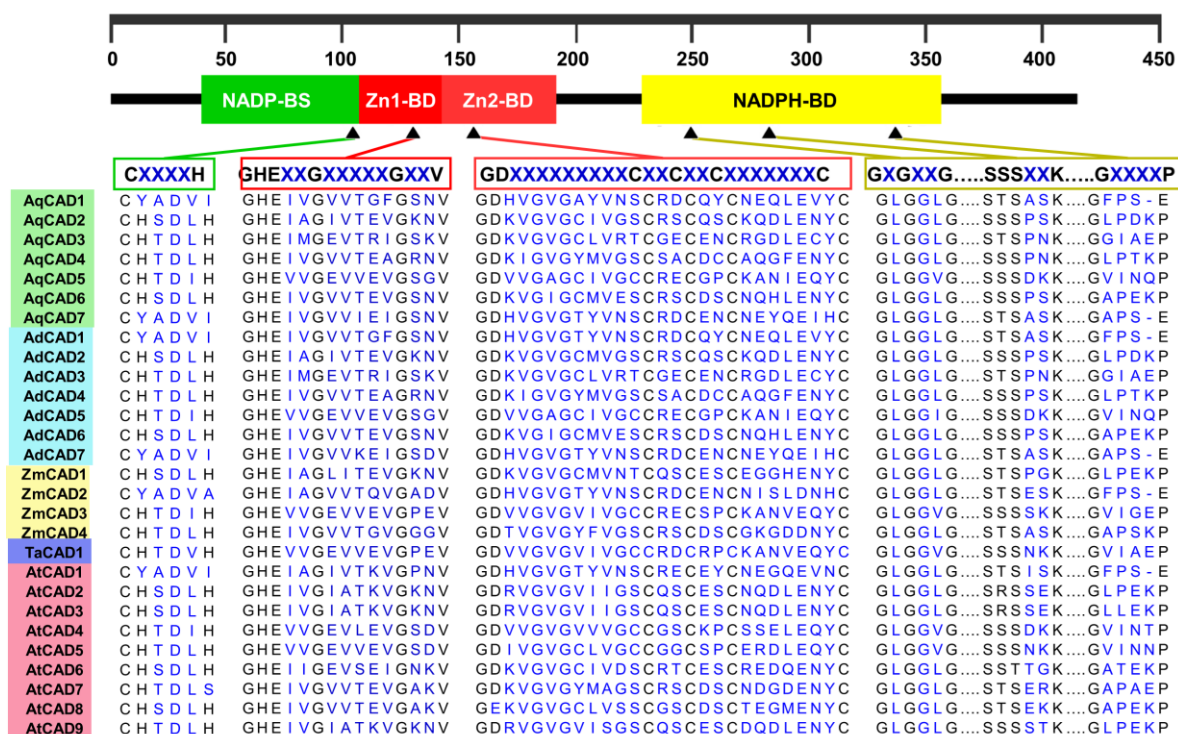


Figura 21. Secuencias de los motivos identificados en los genes ortólogos cinamil alcohol deshidrogenasa (CAD) en (Ad) *A. deserti*, (Aq) *A. tequilana*, (Zm) *Z. mays* y (At) *A. thaliana*.

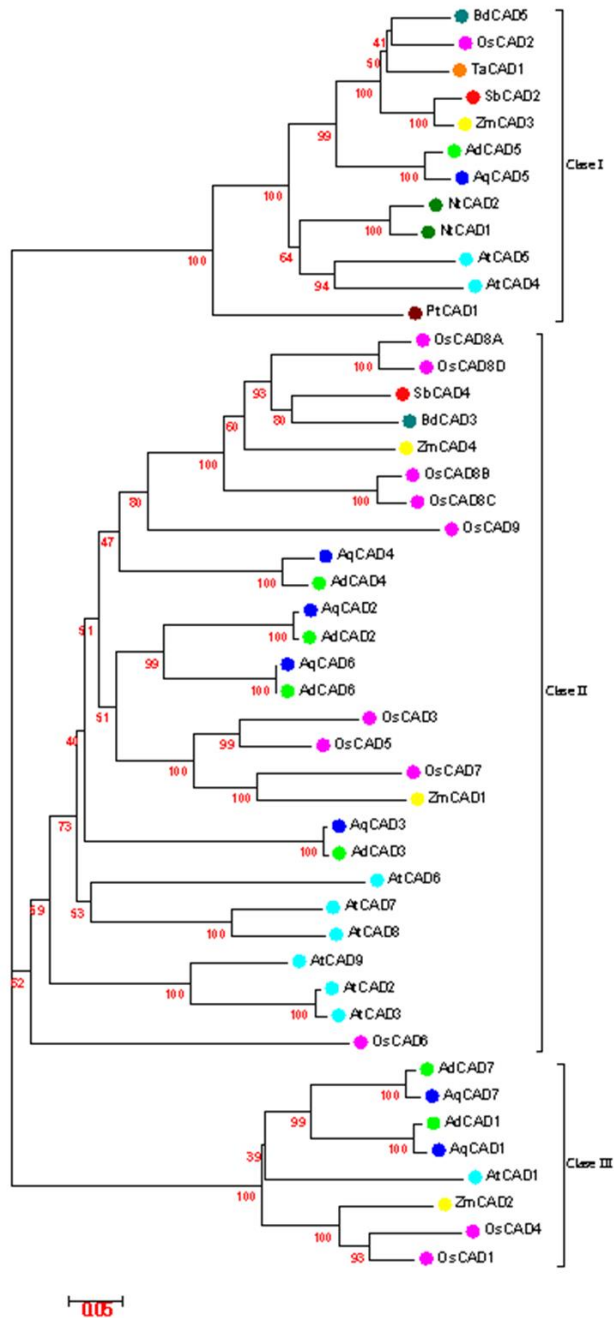
6.1.8 Análisis filogenético de los genes CAD de Agave

Todas las proteínas CAD de las plantas monocotiledóneas como *A. tequilana* (7), *A. deserti* (7), *Z. mays* (4), *O. sativa* (12), *T. aestivum* (1), *Sorghum bicolor* (2), *Brachypodium distachyon* (2) y de las plantas dicotiledóneas como *A. thaliana* (9), *Populus trichocarpa* (1), y *Nicotiana tabacum* (2) fueron utilizadas en un análisis filogenético para poder establecer las relaciones de parentesco para los genes CAD hallados en este estudio (Figura 22).

Se ha demostrado por análisis filogenéticos que los genes CAD se dividen en tres clases: Clase I, II y III (Guo *et al.*, 2010; Tang *et al.*, 2014; Park *et al.*, 2018). La clase I comprende los CAD *bona fide*, para la cual nuestro estudio incluye solo los genes AqCAD5 y AdCAD5 (Figura 22). Esto está en línea con lo previamente

reportado, pues, aunque las familias de genes *CAD* pueden ser multigénicas, solo uno o dos genes se agrupan en la clase I. La evidencia acumulada también muestra que los miembros *CAD* de la clase I participan en la acumulación de lignina en la pared celular secundaria durante el crecimiento y desarrollo de las plantas (Guo *et al.*, 2010; Tang *et al.*, 2014; Park *et al.*, 2018), lo cual para nuestro estudio conduciría a centrar la atención en *AqCAD5* y *AdCAD5*, muy posiblemente los *CAD bona fide* para *Agave* de acuerdo a nuestros análisis de dominios y filogenéticos.

El árbol de la figura 22 muestra además que un gran número de miembros de proteínas *CAD* que se agrupan en la clase II, contiene todos los *AqCAD* y *AdCAD* excepto *AqCAD1*, *AqCAD7*, *AdCAD1* y *AdCAD7* (Figura 22). Los *CAD* incluidos en este grupo como *TaCAD1* de trigo, *OsCAD3*, 5, 6, 7, 8A, 8B, 8C, 8D y 9 de arroz, *AtCAD2*, 3, 7 y 8 de *Arabidopsis*, y *SbCAD4* de sorgo, ha demostrado que codifica para proteínas *CAD* catalíticamente activas (Kim *et al.*, 2004; Sattler *et al.*, 2009; Park *et al.*, 2018). La clase III es un pequeño grupo de *CAD*, incluidos *OsCAD1* y 4, *ZmCAD2* y *AtCAD1* (Guo *et al.*, 2010; Park *et al.*, 2018). En el caso de *agave*, el árbol muestra que *AqCAD1*, *AqCAD7*, *AdCAD1* y *AdCAD7* se encuentran en esta clase.



- AqCAD: *Agave tequilana*
- AdCAD: *Agave deserti*
- ZmCAD: *Zea mays*
- OsCAD: *Oryza sativa*
- SbCAD: *Sorghum bicolor*
- NtCAD: *Nicotiana tabacum*
- PtCAD: *Pinus taeda*
- BdCAD: *Brachypodium distachyon*
- TaCAD: *Triticum aestivum*
- AtCAD: *Arabidopsis thaliana*

Figura 22. Árbol filogenético no enraizado de los genes CAD de *Agave tequilana* (AqCAD), *Agave deserti* (AdCAD), *Oryza sativa* (OsCAD), *Zea mays* (ZmCAD) y *Arabidopsis thaliana* (AtCAD).

6.2. Análisis del perfil transcriptómico *in silico* de los genes identificados en los diferentes órganos de *A. tequilana* Weber y *A. deserti*

A partir de los genes ortólogos identificados previamente en las rutas de celulosa y lignina en ambas especies de agave, se analizaron los perfiles de expresión en transcriptomas de diferentes órganos que tienen pared celular primaria y/o están enriquecidos con tejidos con pared secundaria. Para esto, se tuvo acceso a la base de datos del transcriptoma de *A. tequilana* (Ávila de Dios *et al.*, 2019) y del transcriptoma de *A. deserti* (Gross *et al.*, 2013). Los cuadros 21 y 22 muestran los perfiles de expresión en TPM para todos los ortólogos identificados.

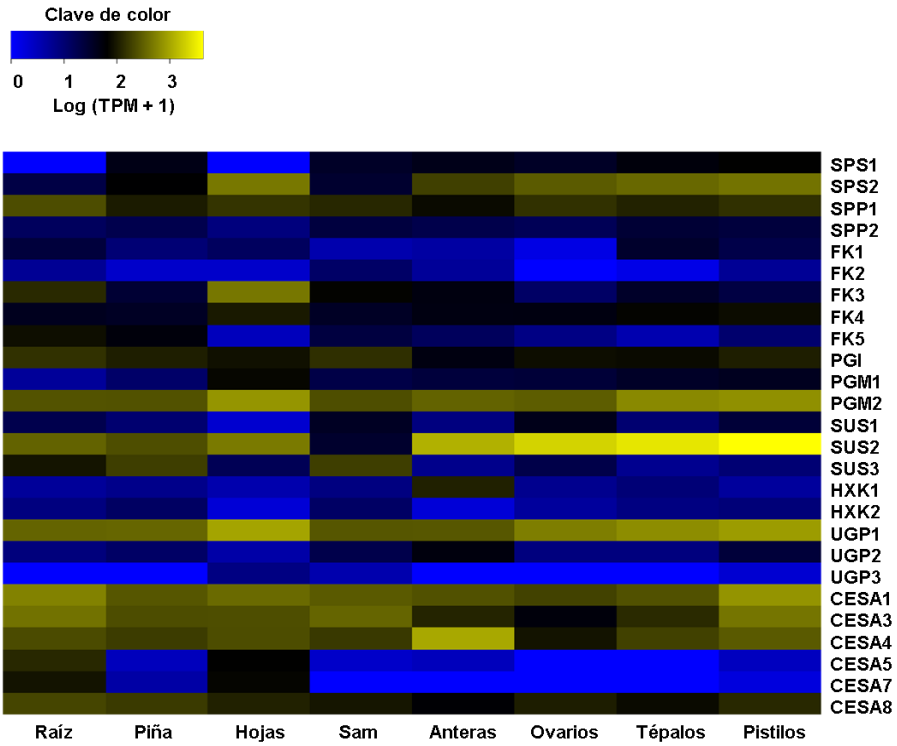
6.2.1 Expresión *in silico* de los genes involucrados en la biosíntesis de celulosa en *A. tequilana* Weber y *A. deserti*

Estudios previos en plantas modelo como *A. thaliana* han demostrado que los genes de la familia celulosa sintasa se expresan de manera específica en pared celular primaria y secundaria (Desprez *et al.*, 2007; Persson *et al.*, 2007; Taylor *et al.*, 2003; Atanassov *et al.*, 2009). En nuestro estudio, aunque los análisis filogenéticos nos indican que *AqCESA1*, *AqCESA3*, *AqCESA5*, *AdCESA1*, *AdCESA3* y *AdCESA1* pudieran ser específicos de la pared celular primaria y que *AqCESA4*, *AqCESA7*, *AqCESA8*, *AdCESA4*, *AdCESA7* y *AdCESA8* son específicos de pared secundaria, los niveles de expresión indican algunas diferencias. La figura 23A muestra que *AqCESA1*, *AqCESA3*, *AqCESA4* y *AqCESA8* presentan un alto nivel de expresión en todos los órganos analizados. Interesantemente, *AqCESA5* y *AqCESA7* presentan un alto nivel de expresión en tejidos que tienen pared celular secundaria como raíz y hojas. En el caso de *AqCESA7*, los análisis de expresión son consistentes con los análisis filogenéticos para ubicar a este gen como un gen *CESA* participando activamente en la síntesis de tejidos con alto contenido de celulosa.

Para *A. deserti* se obtuvieron los perfiles de expresión en órganos como hojas,

meristemo, hijuelos y raíz. Los resultados obtenidos nos sugieren que *AdCESA1* y *AdCESA5* se expresan en todos los transcriptomas analizados. Además, *AdCESA3*, *AdCESA4*, *AdCESA7* y *AdCESA8* presentan un mayor nivel de expresión en el meristemo mientras que en el resto de los tejidos su nivel es bajo (Figura 23B). A diferencia de lo hallado para *A. tequilana*, en este caso, no se observó algún patrón de expresión que indicará algún gen *CESA* con alta expresión en tejidos enriquecidos con pared celular secundaria.

A) Biosíntesis de celulosa en *Agave tequilana*



B) Biosíntesis de celulosa en *Agave deserti*

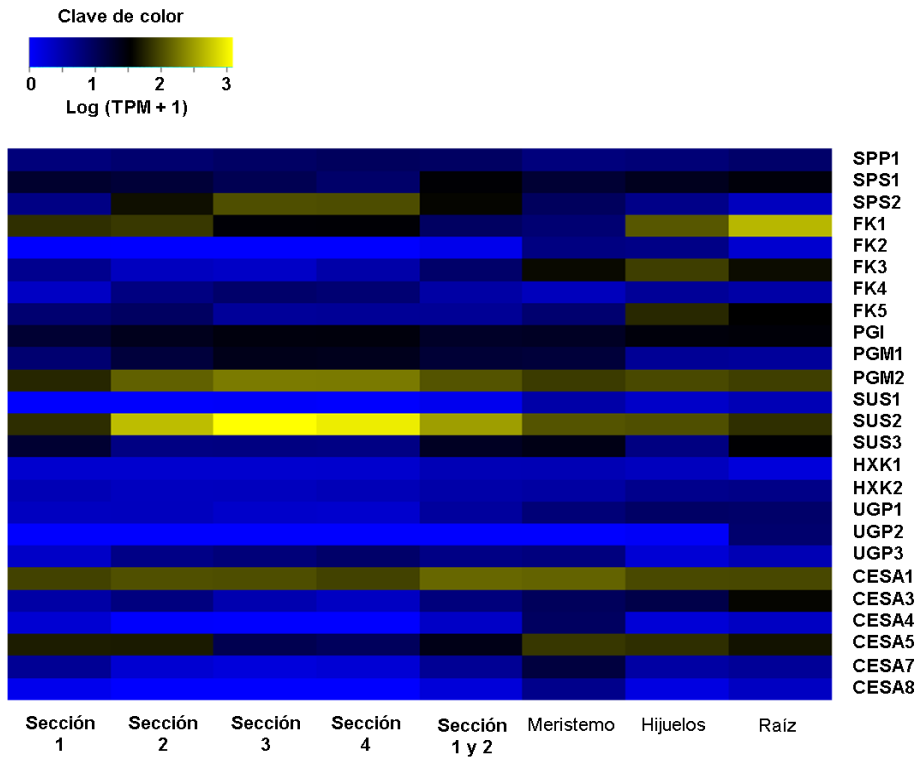


Figura 23. Mapa de calor de los niveles de expresión de los genes identificados en la biosíntesis de celulosa en A) *A. tequilana* y B) *A. deserti*.

6.2.2 Perfil transcriptómico de los genes involucrados en la biosíntesis de lignina en *A. tequilana* Weber y *A. deserti*

Los análisis de expresión para los genes involucrados en la biosíntesis de lignina en los diferentes tejidos de *A. tequilana* señalan un mayor nivel de expresión de casi todos los genes ortólogos en raíz (Figura 24). Estos resultados concuerdan con estudios previamente reportados, sobre la distribución de cDNAs secuenciados en los diferentes tejidos de *A. tequilana* (Simpson *et al.*, 2011).

La biosíntesis de lignina comienza con la enzima fenilalanina amonio liasa (PAL) en el citosol. La vía está modulada por PAL, que es una enzima limitante de la velocidad de biosíntesis de la ruta (Hisano *et al.*, 2009; Jeong *et al.*, 2012; Jia *et al.*, 2015). La expresión de *PAL1* y *PAL2* en *A. tequilana* y *A. deserti* fue discrepante en los diferentes tejidos. Para el caso de *PAL1* de *A. tequilana* los niveles de expresión fueron mayores en raíz, hojas, tépalos y pistilos y *PAL2* raíz, anteras, ovarios, tépalos y pistilos. Además, en *A. deserti* *PAL1* presentó los niveles de expresión más altos en raíz e hijuelos y *PAL2* en la sección 1 y 2 de hojas, en raíz e hijuelos. Este hallazgo sugiere que este gen tiene diferentes funciones en el crecimiento y desarrollo de Agave.

La enzima 4CL cataliza el Ácido p-cumárico a 4-cinamoil CoA y participa en la biosíntesis de otros metabolitos secundarios como las fitoalexinas y los flavonoides (Chowdhury *et al.*, 2013; Jia *et al.*, 2015). El nivel de expresión de 4CL aumentó en raíz, piña, meristemo, anteras, tépalos y pistilos en *A. tequilana*. Y en *A. deserti* aumentó gradualmente en raíz, meristemo e hijuelos.

HCT afecta el contenido de lignina y cambia significativamente la composición de lignina. La regulación negativa de HCT reduce el contenido de lignina y aumenta la proporción de unidades H en comparación con las unidades G y S (Pu *et al.*, 2009). En *A. tequilana* HCT1 y HCT2 presentan una mayor expresión en raíz mientras que en el resto de los demás órganos los niveles de expresión son bajos.

Este resultado concuerda con *A. deserti*, aunque en esta especie también hay un incremento en el nivel de expresión en hijuelos y meristemo.

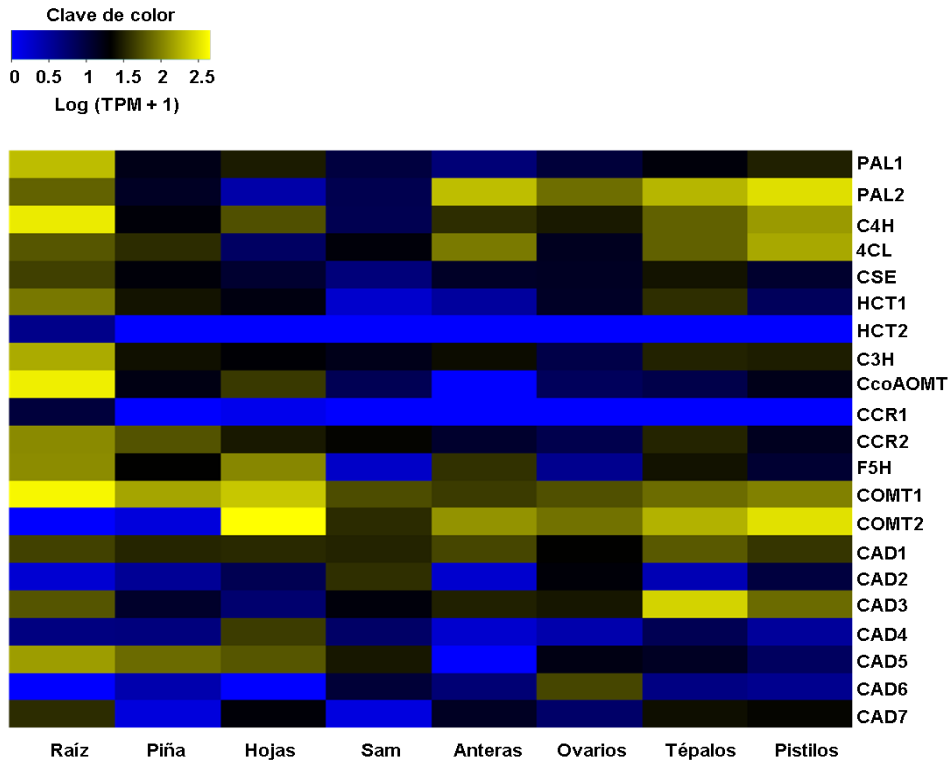
F5H y COMT son enzimas clave implicadas en la formación de unidades de siringilo (Golfier *et al.*, 2019). F5H cataliza la 5-hidroxilación de coniferaldehído y alcohol coniferílico, que luego son metilados por COMT (Vanholme *et al.*, 2010). *F5H* presenta el mayor nivel de expresión en raíz en ambas especies de agave. Además, *COMT1* de *A. tequilana* y *COMT2* de *A. deserti* se expresa en todos los tejidos. CCR y CAD catalizan los pasos finales de la biosíntesis de lignina, estas enzimas pueden determinar la composición y el contenido total de lignina de una planta (Yan *et al.*, 2019). En la figura 24A muestra que *AqCAD1*, *AqCAD3*, *AqCAD5* y *AqCAD7* presentan un alto nivel de expresión en raíz. Mientras que *AqCAD1*, *AqCAD4*, *AqCAD5* y *AqCAD7* su nivel de expresión es alto en hojas. Interesantemente genes como *AqCAD1* y *AqCAD3* están siendo expresados en tejidos que como meristemo, anteras, ovarios, tépalos y pistilos que carecen de pared secundaria, sugiriendo otras funciones biológicas para las subunidades formadas potencialmente por este par de genes.

Para *A. deserti* se obtuvieron los perfiles de expresión en tejidos como hojas, meristemo, hijuelos y raíz. Los resultados obtenidos nos muestran que la mayoría de los genes involucrados en la biosíntesis de lignina se están expresando altamente en hijuelos y raíz (Figura 24B). Para el caso de los genes *CAD*, la mayoría se está expresando en raíz a excepción de *AdCAD2*. Similarmente a lo hallado en *A. tequilana*, los mapas revelaron genes *AdCAD1*, *AdCAD2* y *AdCAD4*, con expresión en meristemo, un tejido con nulo crecimiento secundario donde difícilmente las subunidades sintetizadas conformarían el polímero lignina.

Estudios previos han demostrado que la familia de genes *CAD* se divide en tres clases. La clase I es donde se encuentran los genes *CAD bona fide*, estos genes están implicados en síntesis de subunidades de lignina que forma parte de la pared celular secundaria. Así mismo en la clase II y III es donde se concentra la

mayoría de los genes *CAD* según estudios filogenéticos realizados previamente en diferentes especies de plantas monocotiledóneas y eudicotiledóneas (Guo *et al.*, 2010; Tang *et al.*, 2014; Park *et al.*, 2018). En este trabajo, los análisis filogenéticos y de expresión sugieren fuertemente que *AqCAD5* y *AdCAD5* están involucrados en la formación de pared celular secundaria y son de hecho, los *CAD bona fide* en agave (Figura 24).

A) Biosíntesis de lignina en *Agave tequilana*



B) Biosíntesis de lignina en *Agave deserti*

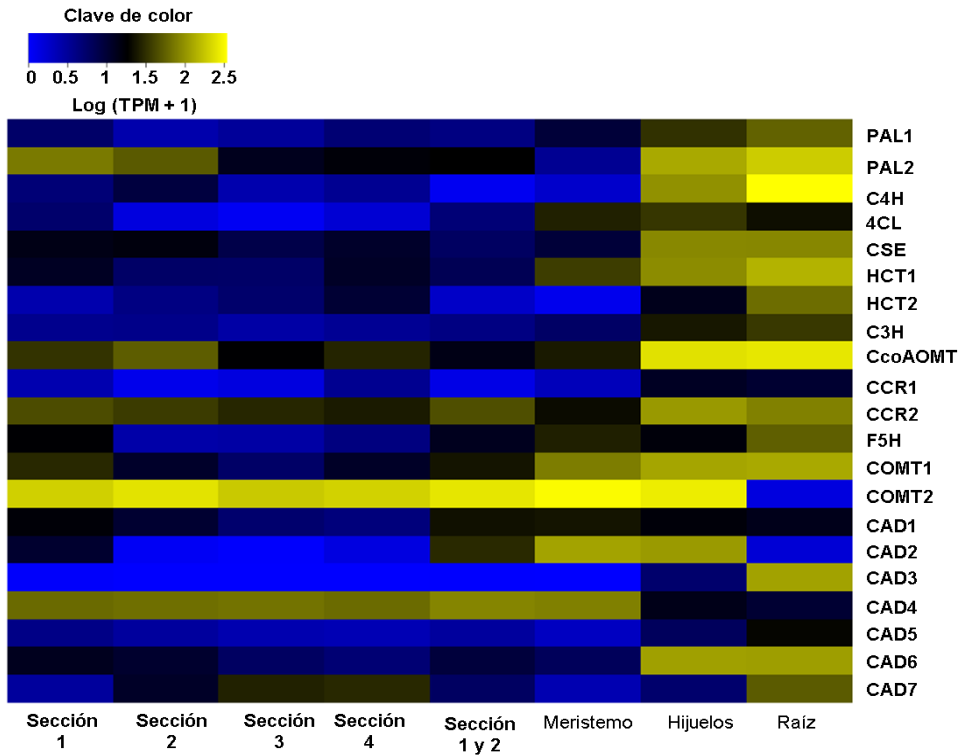


Figura 24. Mapa de calor de los niveles de expresión de los genes identificados en la biosíntesis de lignina en A) *A. tequilana* y B) *A. deserti*.

6.3. Modificaciones menores al protocolo estándar de la extracción de RNA total

Para analizar experimentalmente la expresión de algunos genes involucrados en la biosíntesis de celulosa y lignina en *A. tequilana*, se realizaron 2 experimentos en *A. tequilana*. El primer experimento consistió en analizar la expresión de *SUS2*, *CESA1*, *CESA7*, *CCR2* y *CAD5* en raíz, piña meristemo y hoja. El segundo experimento consistió en analizar la expresión de *SUS2*, *CESA1*, *CESA7*, *CCR2* y *CAD5* en hojas y fibras.

Para ello el primer paso fue obtener RNA de buena calidad de los diferentes tejidos para los ensayos RT-PCR punto final. Las extracciones de RNA total en los diferentes tejidos permitieron obtener un buen rendimiento de RNA total y de buena calidad, excepto en el caso de raíz. El Cuadro 14 muestra rendimientos de cantidad y calidad de RNA de raíces con el compuesto rojo (raíces rojas) y sin el compuesto rojo (raíces blancas). Aún no existen reportes sobre la naturaleza química de este compuesto; una hipótesis sugiere que puede ser betalaina, un pigmento de color rojo presente también en raíces como la remolacha (Hattestad *et al.*, 2015). Nuestro protocolo modificado para extracción de RNA total de raíz sin el pigmento rojo (raíces blancas) aumentó significativamente los cocientes absorción 260/280, mientras que las raíces rojas contienen índices de pureza 260/280 más bajos, lo cual concuerda con la presencia del metabolito (Cuadro 14). Incluso, el protocolo de purificación fue incapaz de incrementar la pureza en las muestras contaminadas (Cuadro 14), las cuales muestran un patrón electroforético anormal (Figura 25). Probablemente el contaminante en las extracciones de raíz roja está asociado a los RNAs e impiden su movimiento a lo largo del gel.

Cuadro 14. Rendimiento y pureza del RNA total extraído sin purificar y purificado de raíz de *A. tequilana*.

	Tejido	Muestra	ng/ μ L	260/230	260/280
RNA total extraído sin purificar	Raíz roja	1	306	0.89	1.62
		2	775	0.75	1.58
		3	406	0.96	1.63
		4	592	0.55	1.60
	Raíz blanca	1	627	1.02	1.93
		2	580	0.79	1.95
		3	694	0.85	1.93
		4	495	0.75	1.86
RNA total extraído purificado	Raíz roja	1	211	1.09	1.78
	Raíz blanca	1	899	2.18	2.09

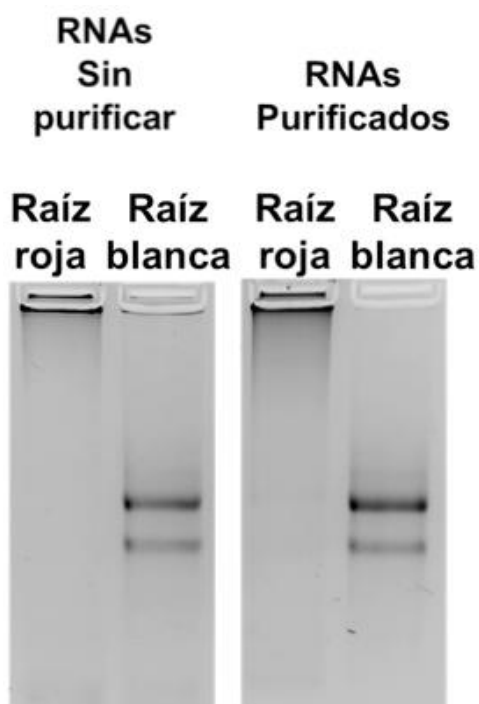


Figura 25. Gel Electroforesis de RNA total de raíz roja y raíz blanca de *A. tequilana* en gel de agarosa al 1%. Las dos bandas intensas corresponden a la subunidad mayor y menor del RNA ribosomal.

Para los demás tejidos se obtuvo RNA total de calidad excelentes después de la purificación: tejidos de raíz, piña, meristemo y hoja (Análisis 1), y hoja y fibra (Análisis 2). En ambos experimentos, las muestras tienen valores 260/280 cercanos a 2 (Cuadro 15), el cual es un parámetro de calidad, y las electroforesis muestran la ausencia de DNA genómico y la presencia de bandas bien definidas correspondientes a las subunidades grande y pequeña del RNA ribosomal (28S rRNA y 18S rRNA, respectivamente) (Figura 26).

Cuadro 15. Rendimiento y pureza del RNA total extraído de raíz, piña, meristemo y hoja (Análisis 1) y hojas y fibras (Análisis 2) de *A. tequilana*.

	Tejido	[ng/μL]	260/280
Análisis 1	Raíz	207.1	2.01
	Piña	786.6	2.05
	Meristemo	1163	2.04
	Hoja	254.9	2.06
Análisis 2	Hoja	213.15	2.09
	Fibras	82.5	2.09

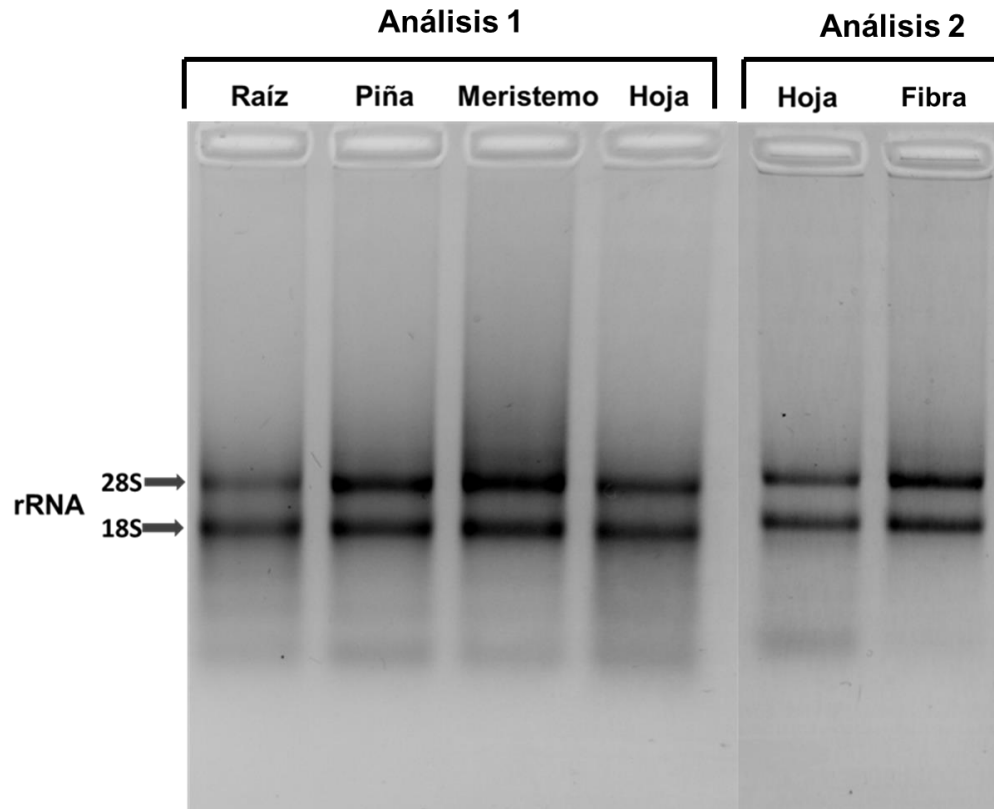


Figura 26. Gel de agarosa teñido con bromuro de etidio en el que se observan los RNAs totales de raíz, piña, meristemo y hoja (Análisis 1) y hojas y fibras (Análisis 2) de *A. tequilana*.

6.4 Análisis de expresión de los genes *SUS2*, *CESA1*, *CESA7*, *CCR2* y *CAD5* en raíz, piña, meristemo y hoja de *A. tequilana*

Este análisis mostró los niveles de expresión de algunos genes ortólogos hallados en este estudio para celulosa y lignina en órganos y/o tejidos que presentan pared primaria y algunos que tiene ya pared secundaria.

6.4.1 Identificación de genes endógenos como controles de carga

Para los ensayos tipo RT-PCR punto final, los genes control para carga utilizados fueron *G3P* y *TEF1- α* . Para cada gen se realizaron PCRs con 25 ciclos, y para *TEF1- α* se probó también 30 ciclos. Los resultados indican que ninguno de los dos genes seleccionados tiene niveles similares de expresión en todos los tejidos analizados (experimento 1) (Figura 27), excepto *G3P* para el experimento 2 de hojas y fibras (Figura 28).

6.4.2 Expresión de *SUS2*, *CESA1* y *CESA7*

No obstante, para el experimento 1 se careció de un gen control, las cantidades de RNA total para la síntesis de cDNA fueron muy iguales para cada órgano/tejido analizado (Figura 27). Para el caso de celulosa, los análisis muestran que el gen *SUS2* presenta el mayor nivel de expresión en hoja y piña en Agave (Figura 27). Este resultado concuerda con lo reportado en Arabidopsis donde *SUS1*, *SUS2*, *SUS3* y *SUS4* tiene los mayores niveles de expresión en hojas (Fernández *et al.*, 2011), respecto a otros órganos de la planta. En plantas monocotiledóneas como *O. sativa* se ha reportado que *SUS1*, *SUS2* y *SUS4* están expresados en cariósides mientras que los niveles de expresión de *SUS2* son altos en la mayoría de los tejidos de arroz (Wang *et al.*, 1999; Hirose *et al.*, 2008). En el caso de agave, la mayor expresión de *SUS2* coincide con los órganos donde existe una mayor acumulación de azúcares. *SUS2* es una sucrosa sintasa que cataliza la conversión reversible de sacarosa y un nucleósido difosfato en nucleósido difosfato glucosa y fructosa (Fernández *et al.*, 2011). Los resultados del RT-PCR punto final para *SUS2* también coinciden con los niveles de expresión *in silico* para todos los tejidos analizados, excepto para raíz (Figura 23A y Figura 27)

El gen *CESA1* se ha reportado que está presente en la pared primaria en *A. thaliana*, además este gen es el ortólogo al gen *CESA1* de *A. tequilana*. Al ser un gen que está presente en pared primaria sus niveles de expresión deberían de ser idénticos en todos los tejidos como ocurre con *P. trichocarpa* (Takata *et al.*, 2015). En nuestros análisis de expresión *in silico* el gen *CESA1* tiene niveles de expresión muy similares en todos los tejidos. Estos resultados concuerdan con los resultados de RT-PCR punto final para piña, meristemo y hoja (Figura 27). Para el caso de raíz el nivel de expresión de *CESA1* es muy bajo. Interesantemente, también fue posible la amplificación de una banda inespecífica menor a 345 pb, que pudiera corresponder a alguna una isoforma del gen *CESA1*. Cabe aclarar que *CESA1* de *A. thaliana* es muy similar a *CESA10* de esa misma especie y que en el caso particular de *A. tequilana* no fuimos capaces de encontrar el gen ortólogo a *CESA10* por lo que probablemente esa isoforma corresponda a *CESA10*.

El gen *CESA7* al igual que *CESA4* y *CESA8* participan en la formación de la pared secundaria en *A. thaliana* (Taylor *et al.*, 2003; Handakumbura *et al.*, 2013; Takata *et al.*, 2015). En *A. tequilana*, los análisis *in silico* muestran un mayor nivel de expresión en hojas y raíz mientras que en el resto de los tejidos su nivel es bajo. Los análisis de expresión por RT-PCR punto final muestran discrepancias respecto a este análisis previo. *AqCESA7* mostró mayor expresión en piña y meristemo, y bajo nivel de expresión en hojas y raíces (Figura 27). Una posible explicación sería una amplificación inespecífica para raíz, piña y meristemo (de hecho, se observa una banda extra menor a la esperada), y sub-representación de los mRNA para *AqCESA7* en hoja. En hoja, *CESA7*, acorde a su función biológica debería ser expresada en tejido con pared secundaria o proximal a las fibras.

6.4.3 Expresión de *CCR2* y *CAD5*

Los genes seleccionados que están involucrados en la biosíntesis de lignina fueron *CCR2* y *CAD5*. La familia de genes *CCR* y *CAD* catalizan los pasos finales de la biosíntesis de lignina, y modulan por tanto el contenido de lignina (Yan *et al.*, 2019). Los análisis del RT-PCR punto final mostraron mayor expresión en piña,

pero casi nulo para raíz (Figura 27), lo cual es diferente a lo hallado *in silico* previamente (Figura 24). Lo interesante es que raíz es un tejido lignificado por lo cual es posible que este gen esté presente; en este caso, sería interesante analizar del otro ortólogo *CCR1*.

El gen *CAD5* fue seleccionado debido a que en base a los análisis filogenéticos es un gen que es específico de pared secundaria y además porque es el único gen de *A. tequilana* que pertenece a la clase I de los genes *CAD* que es donde se ubican los genes *CAD bona fide*. Este gen es ortólogo a *CAD4* y *CAD6* de *A. thaliana* y a *CAD3* de *Z. mays*. Los análisis *in silico* de los niveles de expresión de este gen indican que tiene los mayores niveles de expresión en raíz, piña y hojas. Este resultado coincide con los análisis de expresión mediante RT-PCR, aunque en raíz tiene un menor nivel de expresión.

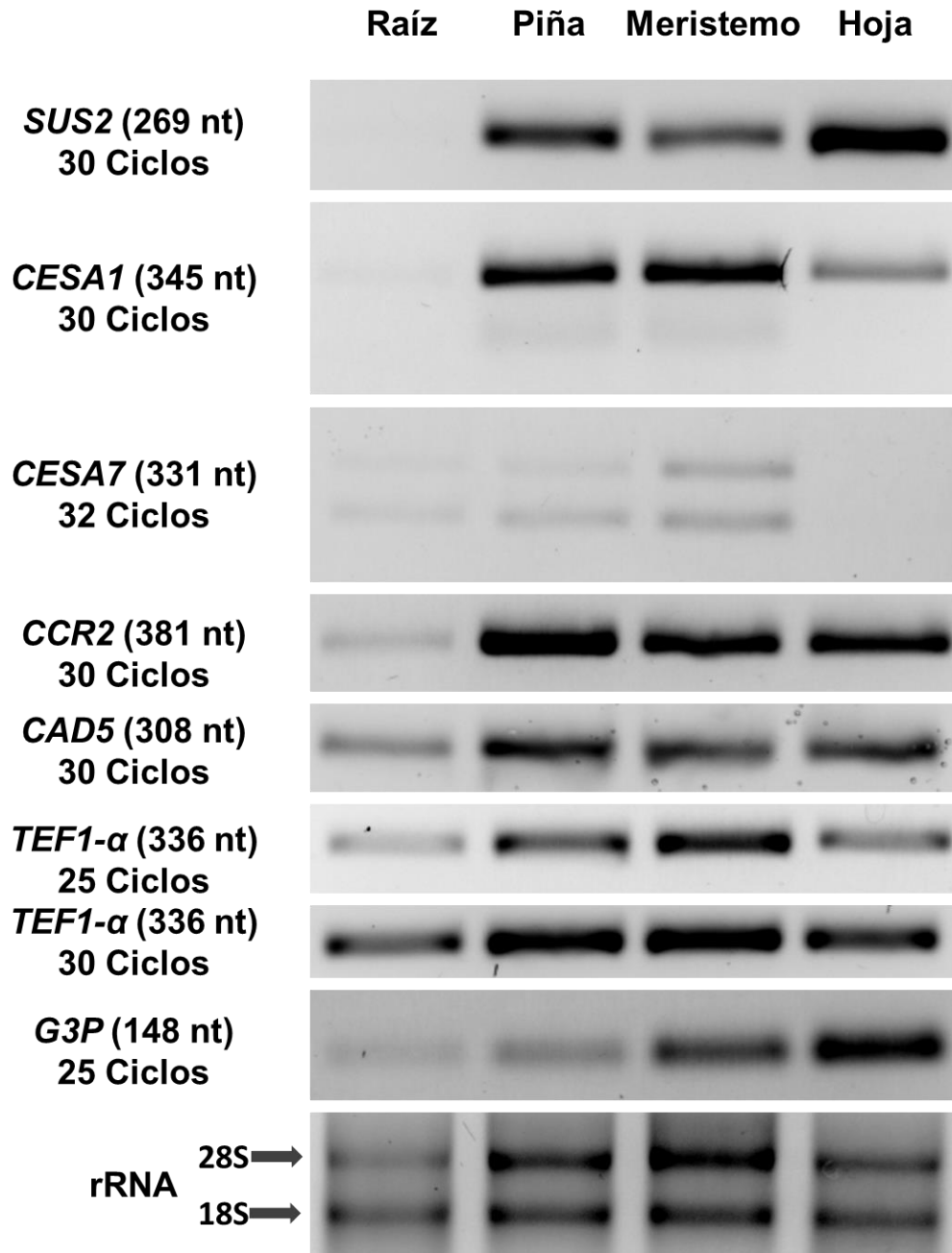


Figura 27. RT-PCR de *SUS2*, *CESA1*, *CESA7*, *CCR2* y *CAD5* en raíz, piña meristemo y hoja en *A. tequilana*.

6.5 Análisis de expresión de los genes *SUS2*, *CESA1*, *CESA7*, *CCR2* y *CAD5* en tejido adyacente a las fibras

Los resultados del análisis de expresión mediante RT-PCR punto final de algunos genes para las rutas de celulosa y lignina en órganos/tejidos mostraron inconsistencias respecto a los obtenidos en los perfiles transcriptómicos. Una razón plausible en muchos casos fue la amplificación inespecífica a partir de isoformas, pero también la sub-representación de mRNA de genes con mayor expresión en pared celular secundaria. La segunda hipótesis incluso se desprende de los perfiles transcriptómicos donde genes *CESA*, específicamente aquellos de pared secundaria, no muestran diferencias respecto a los otros tejidos (Figura 23). Para validar esta hipótesis, en un segundo análisis por RT-PCR punto final se colectó solamente fibras con tejido adyacente (Figura 15), y se contrastó con una muestra compuesta de hojas, colectada tradicionalmente.

Los resultados muestran que para el caso de hoja el gen *G3P* es un buen control de carga; se observó un nivel de expresión similar en ambas muestras aún a 25 ciclos de PCR. Los genes *SUS2* y *CESA1* muestran niveles de expresión similares en ambos tejidos, lo cual es consistente con las funciones de ambas enzimas codificadas. *SUS2* aporta glucosa y fructosa a ambos tejidos, mientras que *CESA1* participaría en la formación de la pared celular primaria, ubicada en todas las células. Interesantemente, *CESA7* mostró un mayor nivel de expresión en tejido adyacente a las fibras. Para el caso de lignina, *CCR2* tuvo una expresión ligeramente mayor en tejido proximal a fibras, comparado a hoja completa. El gen *CAD5*, por su parte, está mayormente expresada en tejido cercano a las fibras (Figura 28). La expresión de los 2 genes clave para pared celular secundaria, *CESA7* y *CAD5*, en muestras enriquecidas con pared secundaria valida la hipótesis de la sub-representación de RNAm en los transcriptomas y RT-PCRs punto final. Esto es además consistente con lo reportado en *Zinnia*, donde las células vivas del parénquima adyacentes a elementos de xilema aportan los monómeros para la biosíntesis de lignina (Serk *et al.*, 2015).

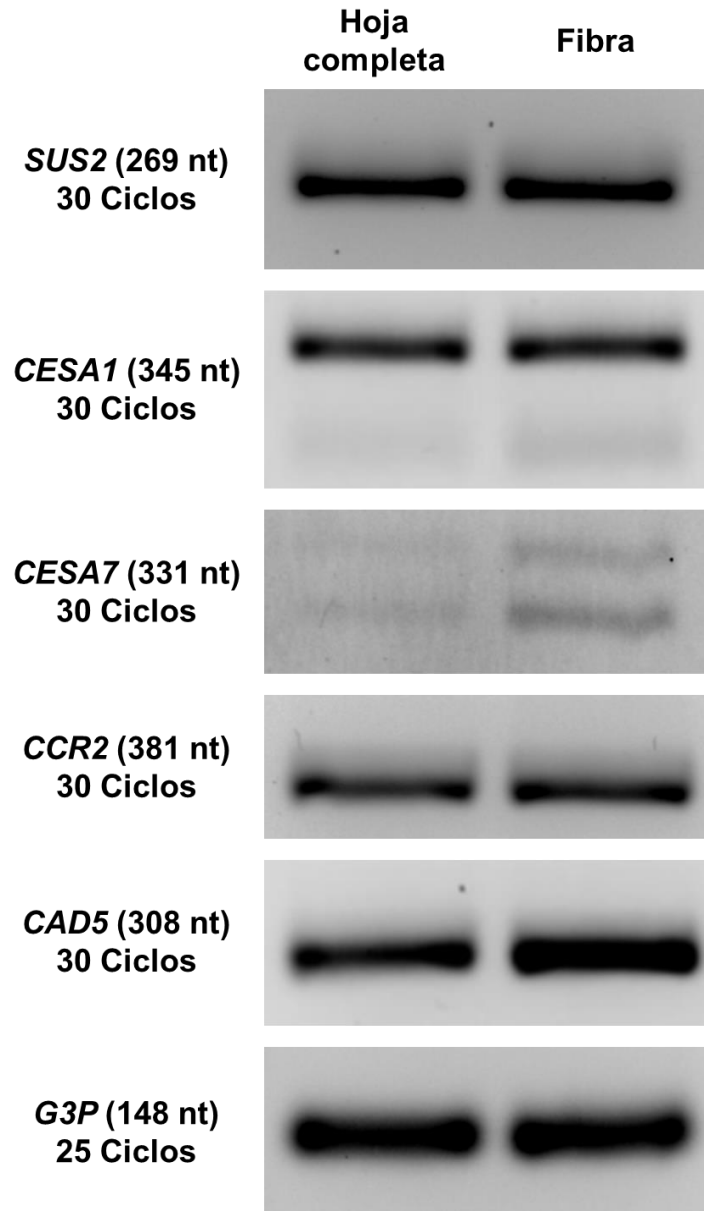


Figura 28. RT-PCR de *SUS2*, *CESA1*, *CESA7*, *CCR2* y *CAD5* en hoja y fibras de *A. tequilana*.

VII. CONCLUSIONES Y PERSPECTIVAS

7.1 Conclusiones

- El enfoque de ortología usado en este estudio fue útil para identificar genes potencialmente codificantes para cada una de las enzimas involucradas en la síntesis de celulosa y lignina en agaves. Los resultados muestran que dichos genes ortólogos están altamente conservados evolutivamente en *A. tequilana* y *A. deserti* respecto a plantas dicotiledóneas como *A. thaliana* y monocotiledóneas como *O. sativa* y *Z. mays*.
- Los análisis de expresión mostraron que algunos genes relacionados con celulosa y lignina están siendo diferencialmente expresados, en la mayoría de los casos acorde con su función esperada para pared celular. Los resultados mostraron también que los genes seleccionados como control de carga para ensayos tipo RT-PCR tienen diferentes niveles de expresión en los tejidos analizados (raíz, piña, meristemo y hoja). Sin embargo, *AqG3P* puede utilizarse como control endógeno en experimentos con tejido foliar. Interesantemente, genes como *AqCESA7* y *AqCAD5* tienen una expresión enriquecida en tejidos proximales a fibras. Este hallazgo sugiere un mecanismo cooperativo para la depositación de componentes estructurales de tejidos de soporte, similar al reportado para elementos de xilema.
- La modificación al protocolo de extracción de RNA en raíces de *A. tequilana* fue efectiva al obtener RNA de buena calidad e integridad.
- La mayoría de los oligonucleótidos diseñados en este estudio produjeron un solo amplicon, excepto aquellos diseñados para genes *CESA*, lo cual pudiera ser resultado de amplificación inespecífica por las isoformas que presenta esta familia génica.

7.2 Perspectivas

- Es indispensable secuenciar el genoma de *A. tequilana* para tener una mayor cobertura de los genes involucrados en la biosíntesis de celulosa y lignina.
- Será necesario corroborar nuestros análisis de ortología con futuros datos genómicos de *Agave sp.* Los estudios de los genes involucrados en las biosíntesis de celulosa y lignina aun son pocos en plantas monocotiledóneas, y más en agave, por lo que es necesario aplicar estas metodologías en diferentes especies de Agave.
- Un enfoque similar al usado en este trabajo podría ser empleado para identificar moduladores genéticos de las rutas de celulosa y lignina en agave, los cuales ya son conocidos en *A. thaliana*.
- A fin de sugerir funciones biológicas y optimizar kits de extracción y purificación para RNA es necesario realizar la identificación del metabolito (s) de color rojo presente en las raíces de *A. tequilana*.

VIII. LITERATURA CITADA

Acker, R. V., Déjardin, A., Desmet, S., Hoengenaert, L., Vanholme, R., Morreel, K., Boerjan, W. (2017). Different Routes for Conifer- and Sinapaldehyde and Higher Saccharification upon Deficiency in the Dehydrogenase CAD1. *Plant Physiology*, 175(3), 1018-1039.

Adrain, C., & Freeman, M. (2012). New lives for old: evolution of pseudoenzyme function illustrated by iRhoms. *Nature reviews Molecular cell biology*, 13(8), 489.

Akpan, E. I. (2019). Chemistry and Structure of Lignin. In *Sustainable Lignin for Carbon Fibers: Principles, Techniques, and Applications* (pp. 1-50). Springer, Cham.

Ambavaram, M. M., Krishnan, A., Trijatmiko, K. R., & Pereira, A. (2011). Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. *Plant physiology*, 155(2), 916-931.

Atanassov, I. I., Pittman, J. K., & Turner, S. R. (2009). Elucidating the mechanisms of assembly and subunit interaction of the cellulose synthase complex of Arabidopsis secondary cell walls. *Journal of Biological Chemistry*, 284(6), 3833-3841.

Babb, V. M., & Haigler, C. H. (2001). Sucrose Phosphate Synthase Activity Rises in Correlation with High-Rate Cellulose Synthesis in Three Heterotrophic Systems. *Plant Physiology*, 127(3), 1234-1242.

Ballesteros, M., & Manzanares, P. (2019). Liquid Biofuels. In *The Role of Bioenergy in the Bioeconomy* (pp. 113-144). Academic Press.

Blokhina, O., Laitinen, T., Hatakeyama, Y., Delhomme, N., Paasela, T., Zhao, L., ... & Fagerstedt, K. (2019). Ray parenchymal cells contribute to lignification of tracheids in developing xylem of Norway spruce. *Plant Physiology*, 181(4), 1552-1572.

Brostow, W., Datashvili, T., & Miller, H. (2010). Wood and wood derived materials. *Journal of Materials Education*, 32(3), 125.

Bušić, A., Mardetko, N., Kundas, S., Morzak, G., Belskaya, H., Šantek, M. I., ... & Šantek, B. (2018). Bioethanol production from renewable raw materials and its separation and purification: a review. *Food technology and biotechnology*, 56(3), 289.

Campanella, J. J., Bitincka, L., & Smalley, J. (2003). MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. *BMC bioinformatics*, 4(1), 29.

- Carrillo-Nieves, D., Alanís, M. J. R., de la Cruz Quiroz, R., Ruiz, H. A., Iqbal, H. M., & Parra-Saldivar, R. (2019). Current status and future trends of bioethanol production from agro-industrial wastes in Mexico. *Renewable and Sustainable Energy Reviews*, 102, 63-74.
- Castillo, Q.D., Sáenz, R.J.T., Narcia, V.M. y Vázquez R.J.A. 2013. Propiedades físico-mecánicas de la fibra de *Agave lechuguilla* Torr. de cinco procedencias bajo plantaciones. *Revista Mexicana de Ciencias Forestales* 4: 78-91.
- Cesarino, I., Araújo, P., Domingues Júnior, A. P., & Mazzafera, P. (2012). An overview of lignin metabolism and its effect on biomass recalcitrance. *Brazilian Journal of Botany*, 35(4), 303-311.
- Chowdhury, M. E. K., Choi, B., Cho, B. K., Kim, J. B., Park, S. U., Natarajan, S., ... & Bae, H. (2013). Regulation of 4CL, encoding 4-coumarate: coenzyme A ligase, expression in kenaf under diverse stress conditions. *Plant Omics*, 6(4), 254.
- Cochrane, F. C., Davin, L. B., & Lewis, N. G. (2004). The Arabidopsis phenylalanine ammonia lyase gene family: kinetic characterization of the four PAL isoforms. *Phytochemistry*, 65(11), 1557-1564.
- Demirbaş, A. (2005). Bioethanol from Cellulosic Materials: A Renewable Motor Fuel from Biomass. *Energy Sources*, 27(4), 327-337.
- Demirbas, A. (2007). Progress and recent trends in biofuels. *Progress in energy and combustion science*, 33(1), 1-18.
- Emms, D. M., & Kelly, S. (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16(1).
- Endler, A., & Persson, S. (2011). Cellulose Synthases and Synthesis in *Arabidopsis*. *Molecular Plant*, 4(2), 199-211.
- Eudes, A., Dutta, T., Deng, K., Jacquet, N., Sinha, A., Benites, V. T., . Loqué, D. (2017). SbCOMT (Bmr12) is involved in the biosynthesis of triclin-lignin in sorghum. *Plos One*, 12(6).
- F.O. Licht. (2016). World ethanol & biofuels report, Vol. 2008–2016
 Franke, R., McMichael, C. M., Meyer, K., Shirley, A. M., Cusumano, J. C., &
- Chapple, C. (2000). Modified lignin in tobacco and poplar plants over-expressing the *Arabidopsis* gene encoding ferulate 5-hydroxylase. *The Plant Journal*, 22(3), 223-234.
- García-Mendoza, A. 2002. "Distribution of the genus *Agave* (Agavaceae) and its endemic species in Mexico", *Cactus and Succulent Journal* (us), núm. 74, 177-

187.

Golfier, P., Unda, F., Murphy, E. K., Xie, J., He, F., Zhang, W., ... & Wolf, S. (2019). Distinct and overlapping functions of *Miscanthus sinensis* MYB transcription factors SCM1 and MYB103 in lignin biosynthesis. *bioRxiv*, 629709.

Grenville-Briggs, L. J., Anderson, V. L., Fugelstad, J., Avrova, A. O., Bouzenzana, J., Williams, A., ... & van West, P. (2008). Cellulose synthesis in *Phytophthora infestans* is required for normal appressorium formation and successful infection of potato. *The Plant Cell*, 20(3), 720-738.

Gross, S. M., Martin, J. A., Simpson, J., Abraham-Juarez, M., Wang, Z., & Visel, A. (2013). De novo transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genomics*, 14(1), 563.

Guarnizo Franco A, Martínez Yepes P. N. Valencia Sánchez HA. (2009). Pretratamientos de la celulosa y biomasa para la sacarificación. *Scientia et Technica*, XV(42):284-289.

Guo, D. M., Ran, J. H., & Wang, X. Q. (2010). Evolution of the cinnamyl/sinapyl alcohol dehydrogenase (CAD/SAD) gene family: the emergence of real lignin is associated with the origin of bona fide CAD. *Journal of Molecular Evolution*, 71(3), 202-218.

Hamann, T., Osborne, E., Youngs, H. L., Misson, J., Nussaume, L., & Somerville, C. (2004). Global expression analysis of CESA and CSL genes in *Arabidopsis*. *Cellulose*, 11(3-4), 279-286.

Handakumbura, P. P., & Hazen, S. P. (2012). Transcriptional Regulation of Grass Secondary Cell Wall Biosynthesis: Playing Catch-Up with *Arabidopsis thaliana*. *Frontiers in Plant Science*, 3.

Hatlestad, G. J., Akhavan, N. A., Sunnadeniya, R. M., Elam, L., Cargile, S., Hembd, A., ... & Lloyd, A. M. (2015). The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. *Nature genetics*, 47(1), 92.

Heidari, P., Ahmadizadeh, M., Izanlo, F., & Nussbaumer, T. (2019). *In silico* study of the CESA and CSL gene family in *Arabidopsis thaliana* and *Oryza sativa*: Focus on post-translation modifications. *Plant Gene*, 100189.

Hirose, T., Scofield, G. N., & Terao, T. (2008). An expression analysis profile for the entire sucrose synthase gene family in rice. *Plant Science*, 174(5), 534-543.

Hoffmann, L., Maury, S., Martz, F., Geoffroy, P., & Legrand, M. (2002). Purification, Cloning, and Properties of an Acyltransferase Controlling Shikimate and Quinate Ester Intermediates in Phenylpropanoid Metabolism. *Journal of*

Biological Chemistry, 278(1), 95-103.

Hu, H., Zhang, R., Tao, Z., Li, X., Li, Y., Huang, J., ... & Peng, L. (2018). Cellulose synthase mutants distinctively affect cell growth and cell wall integrity for plant biomass production in *Arabidopsis*. *Plant and Cell Physiology*, 59(6), 1144-1157.

Iniguez-Covarrubias G, Díaz-Teres R, Sanjuan-Duenas R, Anzaldo-Hernández J, Rowell RM (2001). Utilización de subproductos de la industria del tequila. Parte 2: Valor potencial de *Agave tequilana* Weber hojas azul. *Bioresource Technology*, 77, 101 - 108.

Jeong, M. J., Choi, B. S., Bae, D. W., Shin, S. C., Park, S. U., Lim, H. S., ... & Bae, H. (2012). Differential expression of kenaf phenylalanine ammonia-lyase (PAL) ortholog during developmental stages and in response to abiotic stresses. *Plant Omics*, 5(4), 392.

Jia, X. L., Wang, G. L., Xiong, F., Yu, X. R., Xu, Z. S., Wang, F., & Xiong, A. S. (2015). De novo assembly, transcriptome characterization, lignin accumulation, and anatomic characteristics: novel insights into lignin biosynthesis during celery leaf development. *Scientific reports*, 5, 8259.

Kasem, S., Rice, N., & Henry, R. (2008). Plant Genotyping II: SNP Technology.

Keera, S. T., El Sabagh, S. M., & Taman, A. R. (2011). Transesterification of vegetable oil to biodiesel fuel using alkaline catalyst. *Fuel*, 90(1), 42-47.

Kim, S. J., Kim, M. R., Bedgar, D. L., Moinuddin, S. G., Cardenas, C. L., Davin, L. B., ... & Lewis, N. G. (2004). Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 101(6), 1455-1460.

Labate, C. A., Riaño-Pachón, D. M., Pinto, G. C., Mazzafera, P., & Valledor, L. (Eds.). (2019). *Integration of OMICS Data to Understand Plant Metabolism*. Frontiers Media SA.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359.

Li, B., & Dewey, C. (2014). Rsem. *Bioinformatics*, 41-74.

Li, X., Bonawitz, N. D., Weng, J., & Chapple, C. (2010). The Growth Reduction Associated with Repressed Lignin Biosynthesis in *Arabidopsis thaliana* Is Independent of Flavonoids. *The Plant Cell Online*, 22(5), 1620-1632.

Mahmood, H., Moniruzzaman, M., Iqbal, T., & Khan, M. J. (2019). Recent advances in the pretreatment of lignocellulosic biomass for biofuels and value-added products. *Current Opinion in Green and Sustainable Chemistry*.

Martínez-Ramírez, S.; A. Trinidad-Santos; G. Bautista-Sánchez y E. C. Pedro-Santos. 2013. Crecimiento de plántulas de dos especies de mezcal en función del tipo de suelo y nivel de fertilización. *Revista Fitotecnia Mexicana* 36(4): 387-393.

Maury, S., Delaunay, A., Mesnard, F., Crônier, D., Chabbert, B., Geoffroy, P., & Legrand, M. (2010). O-methyltransferase(s)-suppressed plants produce lower amounts of phenolic vir inducers and are less susceptible to *Agrobacterium tumefaciens* infection. *Planta*, 232(4), 975-986.

Mielenz, J. R., Rodriguez, M., Thompson, O. A., Yang, X., & Yin, H. (2015). Development of Agave as a dedicated biomass source: production of biofuels from whole plants. *Biotechnology for Biofuels*, 8(1).

Mussatto, S. I., Dragone, G., Guimarães, P. M., Silva, J. P., Carneiro, L. M., Roberto, I. C., Teixeira, J. A. (2010). Technological trends, global market, and challenges of bio-ethanol production. *Biotechnology Advances*, 28(6),

Nair, R., & Rost, B. (2002). Sequence conserved for subcellular localization. *Protein Science*, 11(12), 2836-2847.

Nakano, Y., Yamaguchi, M., Endo, H., Rejab, N. A., & Ohtani, M. (2015). NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. *Frontiers in Plant Science*, 6.

Nawaz, M. A., Lin, X., Chan, T. F., Imtiaz, M., Rehman, H. M., Ali, M. A., ... & Chung, G. (2019). Characterization of Cellulose Synthase A (CESA) Gene Family in Eudicots. *Biochemical genetics*, 57(2), 248-272.

Nuruzzaman, M., Manimekalai, R., Sharoni, A. M., Satoh, K., Kondoh, H., Ooka, H., & Kikuchi, S. (2010). Genome-wide analysis of NAC transcription factor family in rice. *Gene*, 465(1-2), 30-44.

Ohtani, M., & Demura, T. (2019). The quest for transcriptional hubs of lignin biosynthesis: beyond the NAC-MYB-gene regulatory network model. *Current opinion in biotechnology*, 56, 82-87.

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., ... & Hayashizaki, Y. (2003). Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA research*, 10(6), 239-247.

Park, H., Kim, T., Bhoo, S., Lee, T., Lee, S. W., & Cho, M. H. (2018). Biochemical Characterization of the Rice Cinnamyl Alcohol Dehydrogenase Gene Family. *Molecules*, 23(10), 2659.

Pear, J. R., Kawagoe, Y., Schreckengost, W. E., Delmer, D. P., & Stalker, D. M. (1996). Higher plants contain homologs of the bacterial celA genes encoding the

catalytic subunit of cellulose synthase. *Proceedings of the National Academy of Sciences*, 93(22), 12637-12642.

Peng, L., Zhang, L., Cheng, X., Fan, L., & Hao, H. (2012). Disruption of cellulose synthesis by 2,6-dichlorobenzonitrile affects the structure of the cytoskeleton and cell wall construction in *Arabidopsis*. *Plant Biology*, 15(2), 405-414.

Perdices, M. B. (2018). El papel de la bioenergía en la nueva bioeconomía. *Mediterráneo económico*, (31), 353-369.

Persson S., Paredez A., Carroll A., Palsdottir H., Doblin M., Poindexter P., Khitrov N., Auer M., Somerville C.R. (2007). Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 104(39), 15566-15571.

Preston, J., Wheeler, J., Heazlewood, J., Li, S. F., & Parish, R. W. (2004). AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. *The Plant Journal*, 40(6), 979-995.

Pu, Y., Chen, F., Ziebell, A., Davison, B. H., & Ragauskas, A. J. (2009). NMR characterization of C3H and HCT down-regulated alfalfa lignin. *BioEnergy Research*, 2(4), 198.

Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., ... & Mielenz, J. R. (2006). The path forward for biofuels and biomaterials. *science*, 311(5760), 484-489.

Ralph, J., Lapierre, C., & Boerjan, W. (2019). Lignin structure and its engineering. *Current opinion in biotechnology*, 56, 240-249.

Rawat, J., Saxena, J., & Sanwal, P. (2019). Biochar: a sustainable approach for improving plant growth and soil properties. In *Biochar-An Imperative Amendment for Soil and the Environment*. IntechOpen.

Riazi, M. R., & Chiaramonti, D. (2017). *Biofuels Production and Processing Technology*. CRC Press.

Richmond, T. A., & Somerville, C. R. (2000). The cellulose synthase superfamily. *Plant physiology*, 124(2), 495-498.

Robinson, M. D., Mccarthy, D. J., & Smyth, G. K. (2009). EdgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140.

Ruprecht, C., Mutwil, M., Saxe, F., Eder, M., Nikoloski, Z., & Persson, S. (2011). Large-Scale Co-Expression Approach to Dissect Secondary Cell Wall Formation

Across Plant Species. *Frontiers in Plant Science*, 2.

S.O. Han, H.Y. Choi., 2010. Morphology and surface properties of natural fibre treated with electron beam. *Microsc. Sci. Technol. Appl. Educ.*, 3, pp. 1880–1887.

Sakamoto, S., Somssich, M., Nakata, M. T., Unda, F., Atsuzawa, K., Kaneko, Y., ... & Brady, S. M. (2018). Complete substitution of a secondary cell wall with a primary cell wall in *Arabidopsis*. *Nature plants*, 4(10), 777.

Sambrook, J., & Russell, D. W. (2001). *Molecular cloning: A laboratory manual*, the third edition.

Sattler, S. E., & Funnell-Harris, D. L. (2013). Modifying lignin to improve bioenergy feedstocks: strengthening the barrier against pathogens?†. *Frontiers in Plant Science*, 4.

Sattler, S. E., Saathoff, A. J., Haas, E. J., Palmer, N. A., Funnell-Harris, D. L., Sarath, G., & Pedersen, J. F. (2009). A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the sorghum brown midrib6 phenotype. *Plant Physiology*, 150(2), 584-595.

Serk, H., Gorzsás, A., Tuominen, H., & Pesquet, E. (2015). Cooperative lignification of xylem tracheary elements. *Plant signaling & behavior*, 10(4), e1003753.

Simpson, J., Hernández, A. M., Juárez, M. J., Sandoval, S. D., Villarreal, A. S., & Romero, C. C. (2010). Genomic resources and transcriptome mining in *Agave tequilana*. *GCB Bioenergy*, 3(1), 25-36.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences.

Takata, N., & Taniguchi, T. (2015). Expression divergence of cellulose synthase (CesA) genes after a recent whole genome duplication event in *Populus*. *Planta*, 241(1), 29-42.

Tang, R., Zhang, X. Q., Li, Y. H., & Xie, X. M. (2014). Cloning and *in silico* analysis of a cinnamyl alcohol dehydrogenase gene in *Pennisetum purpureum*. *Journal of genetics*, 93(1), 145-158.

Taylor, N. G. (2000). Multiple Cellulose Synthase Catalytic Subunits Are Required for Cellulose Synthesis in *Arabidopsis*. *The Plant Cell Online*, 12(12), 2529-2540.

Taylor, N. G., Howells, R. M., Huttly, A. K., Vickers, K., & Turner, S. R. (2003). Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proceedings of the National Academy of Sciences*, 100(3), 1450-1455.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22), 4673-4680.

Tu, W. C., & Hallett, J. P. (2019). Recent advances in the pretreatment of lignocellulosic biomass. *Current Opinion in Green and Sustainable Chemistry*.

Turner, S. R., & Somerville, C. R. (1997). Collapsed Xylem Phenotype of *Arabidopsis* Identifies Mutants Deficient in Cellulose Deposition in the Secondary Cell Wall. *The Plant Cell*, 9(5), 689.

Valdivia, E. R., Herrera, M. T., Gianzo, C., Fidalgo, J., Revilla, G., Zarra, I., & Sampedro, J. (2013). Regulation of secondary wall synthesis and cell death by NAC transcription factors in the monocot *Brachypodium distachyon*. *Journal of experimental botany*, 64(5), 1333-1343.

Vanholme, R., De Meester, B., Ralph, J., & Boerjan, W. (2019). Lignin biosynthesis and its integration into metabolism. *Current opinion in biotechnology*, 56, 230-239.

Vanholme, R., Ralph, J., Akiyama, T., Lu, F., Pazo, J. R., Kim, H., ... & Rohde, A. (2010). Engineering traditional monolignols out of lignin by concomitant up-regulation of F5H1 and down-regulation of COMT in *Arabidopsis*. *The Plant Journal*, 64(6), 885-897.

Wang, A. Y., Kao, M. H., Yang, W. H., Sayion, Y., Liu, L. F., Lee, P. D., & Su, J. C. (1999). Differentially and developmentally regulated expression of three rice sucrose synthase genes. *Plant and cell physiology*, 40(8), 800-807.

Wang, H., Avci, U., Nakashima, J., Hahn, M. G., Chen, F., & Dixon, R. A. (2010). Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. *Proceedings of the National Academy of Sciences*, 107(51), 22338-22343.

Wang, Y. X., Teng, R. M., Wang, W. L., Wang, Y., Shen, W., & Zhuang, J. (2019). Identification of genes revealed differential expression profiles and lignin accumulation during leaf and stem development in tea plant (*Camellia sinensis* (L.) O. Kuntze). *Protoplasma*, 256(2), 359-370.

Wanner, L. A., Li, G., Ware, D., Somssich, I. E., & Davis, K. R. (1995). The phenylalanine ammonia-lyase gene family in *Arabidopsis thaliana*. *Plant molecular biology*, 27(2), 327-338.

Wasiak, A. (2019). Biomass Resources for Biofuel Production. In *Modeling Energetic Efficiency of Biofuels Production* (pp. 3-27). Springer, Cham.

Xiao, W., Yang, Y., & Yu, J. (2018). ZmNST3 and ZmNST4 are master switches for secondary wall deposition in maize (*Zea mays* L.). *Plant Science*, 266, 83-94.

Yamaguchi, M., Ohtani, M., Mitsuda, N., Kubo, M., Ohme-Takagi, M., Fukuda, H., & Demura, T. (2010). VND-INTERACTING2, a NAC domain transcription factor, negatively regulates xylem vessel formation in *Arabidopsis*. *The Plant Cell*, 22(4), 1249-1263.

Yan, X., Liu, J., Kim, H., Liu, B., Huang, X., Yang, Z., ... & Muddiman, D. C. (2019). CAD 1 and CCR 2 protein complex formation in monolignol biosynthesis in *Populus trichocarpa*. *New Phytologist*, 222(1), 244-260.

Yang, H., Yan, R., Chen, H., Lee, D. H., & Zheng, C. (2007). Characteristics of hemicellulose, cellulose and lignin pyrolysis. *Fuel*, 86(12-13).

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, 13(1), 134.

Yoon, J., Choi, H., & An, G. (2015). Roles of lignin biosynthesis and regulatory genes in plant development. *Journal of integrative plant biology*, 57(11), 902-912.

Yoshida, K., Sakamoto, S., Kawai, T., Kobayashi, Y., Sato, K., Ichinose, Y., ... & Ohme-Takagi, M. (2013). Engineering the *Oryza sativa* cell wall with rice NAC transcription factors regulating secondary wall formation. *Frontiers in plant science*, 4, 383.

Zhong, R., Cui, D., & Ye, Z. H. (2019). Secondary cell wall biosynthesis. *New Phytologist*, 221(4), 1703-1723.

Zhong, R., Demura, T., & Ye, Z. (2006). SND1, a NAC Domain Transcription Factor, Is a Key Regulator of Secondary Wall Synthesis in Fibers of *Arabidopsis*. *The Plant Cell Online*, 18(11).

Zhong, R., Lee, C., Zhou, J., Mccarthy, R. L., & Ye, Z. (2008). A Battery of Transcription Factors Involved in the Regulation of Secondary Cell Wall Biosynthesis in *Arabidopsis*. *The Plant Cell Online*, 20(10), 2763-2782.

IX. ANEXOS

Anexo A: Proteínas utilizadas en el análisis filogenético de CESA y CAD en *A. tequilana* y *A. deserti*

Anexo A1. Número de acceso de las proteínas CESA usadas en el análisis filogenético.

Nombre científico	Número de acceso de la proteína
<i>Asparagus officinalis</i>	AoCESA1 (LOC109828772), AoCESA2 (LOC109847015), AoCESA3 (LOC109834482), AoCESA4 (LOC109836391), AoCESA5 (LOC109836559), AoCESA6 (LOC109843790), AoCESA7 (LOC109844015), AoCESA8 (LOC109841011), AoCESA9 (LOC109841019), AoCESA10 (LOC109845371), AoCESA11 (LOC109847523) y AoCESA12 (LOC109820019)
<i>Arabidopsis thaliana.</i>	AtCESA1 (AT4G32410), AtCESA2 (AT4G39350), AtCESA3 (AT5G05170), AtCESA4 (AT5G44030), AtCESA5 (AT5G09870), AtCESA6 (AT5G64740), AtCESA7 (AT5G17420), AtCESA8 (AT4G18780), AtCESA9 (AT2G21770) y AtCESA10 (AT2G25540)
<i>Agave deserti</i>	AdCESA1 (UN005336), AdCESA3 (UN024289), AdCESA4 (UN031321), AdCESA5 (UN085584), AdCESA7 (UN100087) y AdCESA8 (UN050033)
<i>Agave tequilana</i>	AqCESA1 (UN017499), AqCESA3 (UN003733), AqCESA4 (UN045020), AqCESA5 (UN052428), AqCESA7 (UN068270) y AqCESA8 (UN481687)
<i>Beta vulgaris</i>	BvCESA2 (XP_010678670)
<i>Eucalyptus grandis</i>	EgCESA5 (XP_010063196)
<i>Gossypium hirsutum</i>	GhCESA5 (AFB18634)
<i>Glycine max</i>	GmCESA2 (XP_003531396)
<i>Hordeum vulgare</i>	HvCESA1 (AY483150), HvCESA2 (AY483152), HvCESA3 (AY483151), HvCESA4 (AY483154), HvCESA5 (AY483153), HvCESA6 (AY483155), HvCESA8 (AY483156) y HvCESA9 (AK367031)
<i>Oryza sativa</i>	OsCESA1 (LOC_Os05g08370), OsCESA2 (LOC_Os03g59340), OsCESA3 (LOC_Os07g24190), OsCESA4 (LOC_Os01g54620), OsCESA5 (LOC_Os03g62090), OsCESA6 (LOC_Os07g14850), OsCESA7 (LOC_Os10g32980), OsCESA8 (LOC_Os07g10770), OsCESA9 (LOC_Os09g25490), OsCESA10 (LOC_Os12g29300) y OsCESA11 (LOC_Os06g39970)

<i>Populus trichocarpa</i>	PtCESA6 (XP_002319002)	
<i>Solanum tuberosum</i>	StCESA3 (XP_006354075)	
<i>Triticum aestivum</i>	TaCESA1 (Traes_2AS_665AF9500.1), (Traes_4AL_941C0E3EF.2), (Traes_5AL_E176291CC.1), (Traes_1AL_F420A1BBE.1), (Traes_1AS_10C467127.1), (Traes_6BS_8DA635027.1), (TRAES3BF028900030CFD), (Traes_5BL_51C858A97.1) (Traes_2BS_9B34A7A43.2)	TaCESA2 TaCESA3 TaCESA4 TaCESA5 TaCESA6 TaCESA7 TaCESA8 TaCESA9 y
<i>Zea mays</i>	ZmCESA1 (AF200525), ZmCESA3 (NP_001292792.1), (AF200528), (AF200530), (AF200532), (AY372244), (AY372246)	ZmCESA2 (AF200526), ZmCESA4 ZmCESA5 (AF200529), ZmCESA6 ZmCESA7 (AF200531), ZmCESA8 ZmCESA9 (AF200533), ZmCESA10 ZmCESA11 (AY372245), ZmCESA12 y ZmCESA13 (KJ874174)

Anexo A2. Número de acceso de las proteínas CAD usadas en el análisis filogenético.

Nombre científico	Número de acceso de la proteína
<i>Agave deserti</i>	AdCAD1 (UN047496), AdCAD2 (UN055860), AdCAD3 (UN105918), AdCAD4 (UN121435), AdCAD5 (UN160915), AdCAD6 (UN014798) y AdCAD7 (UN041364)
<i>Agave tequilana</i>	AqCAD1 (UN045407), AqCAD2 (UN057761), AqCAD3 (UN146273), AqCAD4 (UN050838), AqCAD5 (UN049346), AqCAD6 (UN050300) y AqCAD7 (UN090516)
<i>Arabidopsis thaliana</i>	AtCAD1 (AT1G72680), AtCAD2 (AT2G21730), AtCAD3 (AT2G21890), AtCAD4 (AT3G19450), AtCAD5 (AT4G34230), AtCAD6 (AT4G37970), AtCAD7 (AT4G37980), AtCAD8 (AT4G37990) y AtCAD9 (AT4G39330)
<i>Brachypodium distachyon</i>	BdCAD3 (AFK80371) y BdCAD5 (AFK80372)
<i>Nicotiana tabacum</i>	NtCAD1 (CAA44216) y NtCAD2 (CAA44217)
<i>Oryza sativa</i>	OsCAD1 (NP_001064283.1), OsCAD2 (NP_001046132.1), OsCAD3 (LOC_Os10g29470), OsCAD4 (LOC_Os11g40690), OsCAD5 (LOC_Os08g16910), OsCAD6 (LOC_Os04g15920), OsCAD7 (LOC_Os04g52280), OsCAD8A (LOC_Os09g23530), OsCAD8B (LOC_Os09g23540), OsCAD8C (LOC_Os09g23550), OsCAD8D (LOC_Os09g23560) y OsCAD9 (LOC_Os03g12270)
<i>Pinus taeda</i>	PtaCAD1 (CAA86073)
<i>Sorghum bicolor</i>	SbCAD2 (AEM63608) y SbCAD4 (XP_002462348)
<i>Triticum aestivum</i>	TaCAD12 (ANW09910)
<i>Zea mays</i>	ZmCAD1 (NM_001137309), ZmCAD2 (XM_008659825), ZmCAD3 (NM_001112184) y ZmCAD4 (NM_001147705)