



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

FACULTAD DE CIENCIAS BIOLÓGICAS

“Estudio del proteoma de las raíces de *Acourtia cordata* y la influencia de factores edáficos en la producción de perezona”

T E S I S

QUE PARA OBTENER EL GRADO DE

DOCTOR EN CIENCIAS NATURALES

PRESENTA:

M. en B. MA. DEL CARMEN GARCÍA MÉNDEZ

TUTOR: DR. JOSÉ DE JESÚS ARELLANO GARCÍA

CO-TUTOR: DR. SERGIO MANUEL ENCARNACIÓN GUEVARA

CUERNAVACA, MORELOS

AGRADECIMIENTOS.

Definitivamente esta es una de las partes más difíciles de escribir, pensando en que no se me olvide ni un nombre de las personas que contribuyeron a lo largo de estos últimos años para que se llevara a cabo este proyecto de investigación y que vivieron junto conmigo esta experiencia.

Quiero dar GRACIAS A DIOS, el autor y consumidor de mis sueños, por permitirme iniciar y concluir con mucha alegría este proyecto, esta etapa llena de sucesos únicos y de enseñanzas los cuales me han formado como persona e investigadora.

Agardezco a todos los miembros de mi comité tutorial por el tiempo dedicado a este trabajo de investigación, por su interés mostrado y sobretodo por sus acertados comentarios, observaciones y consejos durante cada seminario, los cuales ayudaron a enriquecer este proyecto, gracias a la Dra. Victortia Pando, a la Dra. Laura Alvarez, al Dr. Alexander Taketa, al Dr. Sergio Encarnación y al Dr. Jesús Arellano.

Gracias por sus consejos y sus ideas en la parte proteómica al Dr. Sergio Encarnación, lo cual ayudó para que este proyecto continuara adelante, gracias también por recibirme en su laboratorio de investigación y por apoyarnos con material y equipos de laboratorio para la parte experimental del trabajo.

A la Dra. Laura Alvarez y a la Mra. Silvia Marquina muchas gracias por su apoyo y colaboración en la parte química de este proyecto.

Gracias también al Dr. Efraín Tovar por sus importantes aportaciones a este trabajo, gracias por el apoyo para el área del análisis edáfico, gracias por su tiempo y dedicación al proyecto.

Gracias al Dr. Roque Bru por su importante y crucial colaboración en este proyecto, por recibirme en su laboratorio. Gracias a todos los integrantes del laboratorio de Proteómica y Genómica de Plantas (Alicante, España), gracias por recibirme como una integrante más del laboratorio, gracias por el conocimiento compartido Asun, Elías, Susana y Antonio.

GRACIAS, MUCHAS GRACIAS a mi tutor y director de tesis al Dr. Jesús Arellano, primero por creer en mi y por “arriesgarse” conmigo para desarrollar juntos este proyecto, gracias por animarme y echarme porras durante esta etapa del posgrado. Gracias por el conocimiento compartido, por el tiempo dedicado, por los recursos gestionados y por su importante aportación a mi formación. Gracias por todo.

A mis compañeros del laboratorio de Botánica Estructural del Centro de Investigación en Biotecnología, gracias a todos los que ya se fueron y a también a todos los aun permanecen, gracias por los conocimientos y “tips” compartidos, por la buena convivencia, gracias por las risas compartidas y platicas entre comidas que hicieron siempre más ligero el trabajo. Gracias Victor, Gaby, Norely, Kary, Mariana, Carlos, Manuel y José Luis.

Gracias a mis amigos de “experiencias bonitas”, gracias por sus consejos y también gracias por su conocimiento compartido y el tiempo dedicado a nuestra amistad, gracias Norma y Jorge.

A mis compañeros y amigos del laboratorio de Proteómica Funcional del Centro de Ciencias Genómicas, gracias a Gabriel por apoyarme siempre en la parte experimental y por ser muy paciente conmigo, gracias a la mestra Magda por su apoyo en la parte del análisis de las muestras por MALDI/TOF, gracias a Jasiel, Checa, Ramiro y Angelina por su conocimiento compartido y por explicarme siempre lo que no entendía en la parte teorica y experimental, gracias Victoria y Tania por esas tardes de platicas y risas casi incontrolables debido a sus ocurrencias, gracias por permitirme ser parte de su historia de vida.

Gracias a mi amiga y guía espiritual por su constante apoyo, oraciones e interés sobre mi trabajo y vida, gracias por los consejos sabios, gracias Perla.

Y por último, pero no por eso menos importante, GRACIAS A MI FAMILIA Y A MI ESPOSO, gracias a mis padres Pablo y Beatriz por su apoyo incondicional, gracias por sus consejos, oraciones y aliento día a día durante estos últimos años, gracias a mis hermanos, cuñados y sobrinos por quererme como soy, gracias también por los consejos y animos dados, gracias a mi esposo y compañero de vida, gracias Rene por tu paciencia, por los animos y amor incondicional, gracias a todos, sin ustedes esto no habría sido posible.

Esta investigación fue financiada por la Secretaría de Educación Pública (SEP), número de subvención PROMEP / 103.5 / 13/6626; Consejo Nacional de Ciencia y Tecnología (CONACYT), número de beca: 392123/254165 ”; Laboratorio de la Universidad de Alicante, miembro de Proteored PRB3 y apoyado por la subvención PT17 / 0019, del PE I + D + I 2013-2016, financiado por ISCIII y FEDER.

El trabajo de investigación fue desarrollado en los siguientes laboratorios:

Laboratorio de Botánica Estructural del Centro de Investigación en Biotecnología, Universidad Autónoma del Estado de Morelos, México.

Laboratorio Nacional de Estructura de Macromoléculas del Centro de Investigaciones Químicas, Universidad Autónoma del Estado de Morelos, México.

Laboratorio de Marcadores Moleculares del Centro de Investigación en Biodiversidad y Conservación, Universidad Autónoma del Estado de Morelos, México.

Laboratorio de Proteómica del Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México.

Laboratorio de Proteómica y Genómica funcional de plantas de la Universidad de Alicante, España.



ÍNDICE GENERAL.

I. RESUMEN.	1
II. ANTECEDENTES.	3
II.1. El género <i>Acourtia</i>	3
II.2. Compuestos bioactivos del género <i>Acourtia</i>	8
II.3. Propiedades biológicas del género <i>Acourtia</i>	11
II.4. Producción biotecnológica de perezona.	15
II.5. Biosíntesis de terpenos.	16
II.6. Relación de la producción de terpenos con el estrés biótico o abiótico.....	18
II.6.1. Influencia de los nutrientes en la producción de terpenos.	20
II.7. Sesquiterpeno-sintasas.....	21
II.8. Proteómica vegetal.	23
II.9. Estudio de proteomas de especies "huérfanas".....	25
II.10. Métodos proteómicos.	25
II.11. Categorías de la proteómica.	25
II.12. Enfoques proteómicos.	26
II.12.1. El enfoque top-down.	27
II.12.2. El enfoque bottom-up.	28
II.13. Problemas con el manejo de muestras de proteínas.....	30
II.13.1. Preparación de la muestra.....	30
II.13.2. Separación de la muestra.	31
II.13.2.1 Electroforesis bidimensional (2-DE).	31
II.13.2.2. Cromatografía líquida (Liquid chromatography, LC).	33
II.14. Espectrometría de masas.	36
II.14.1. Identificación de proteínas basada en espectrometría de masas.	36
II.15. Estrategias utilizadas para la identificación de proteínas.	42
II.16. Técnicas de fragmentación.	43
II.17. Motores de búsqueda para la identificación de proteínas.	45
II.18. Cuantificación basada en gel.	47
II.18.1. Método 2-DE-semicuantitativo.	47
II.18.2. Electroforesis en gel de expresión diferencial (DIGE).	48
II.19. Cuantificación basada en espectrometría de masas.	49

II.19.1. Técnicas sin etiqueta.....	49
II.19.2. Etiquetado de isótopos estables.....	51
II.20. Análisis funcional.....	54
III. JUSTIFICACIÓN.....	56
IV. HIPÓTESIS.....	57
V. OBJETIVOS.....	58
CAPÍTULO 1. CUANTIFICACIÓN DE PEREZONA Y ANÁLISIS PROTEÓMICO.....	59
1.1 INTRODUCCIÓN.....	60
1.2. MATERIAL Y MÉTODOS.....	61
1.2.1 Material vegetal.....	61
1.2.2 Cuantificación de perezona.....	62
1.2.2.1 Cromatografía de gases acoplado a espectrometría de masas.....	62
1.2.3 Análisis proteómico.....	63
1.2.3.1 Limpieza del tejido de la raíz.....	63
1.2.4 Extracción de proteína.....	63
1.2.5. Separación de proteínas por 2-DE, adquisición y análisis de imágenes.....	64
1.2.6 Identificación de proteínas por MS y comparación en la base de datos.....	65
1.2.6.1 Digestión de proteínas en gel y análisis por MALDI/TOF.....	65
1.2.7. Digestión de proteínas en líquido y análisis por cromatografía líquida de ultra alto rendimiento acoplada al cuadrupolo de tiempo de vuelo (UHPLC-QTOF).....	65
1.2.8 Análisis cuantitativo sin etiqueta (LC-MS).....	67
1.2.9 Categorización y anotación funcional.....	67
1.3. RESULTADOS.....	68
1.3.1 Cuantificación de perezona en raíces de <i>A. cordata</i>	68
1.3.2 Comparación de perfiles proteicos 2-DE de raíces de <i>A. cordata</i>	69
1.3.3 Identificación y clasificación funcional de las proteínas diferencialmente expresadas.....	71
1.3.6. Comparación proteómica entre plantas altamente productoras y de baja producción de perezona.....	72
1.3.4 Análisis de interacción proteína-proteína.....	75
1.4. DISCUSIÓN.....	80
1.5. CONCLUSIONES.....	82
1.6. PERSPECTIVAS.....	83
CAPÍTULO 2. CORRELACIÓN ENTRE FACTORES ECOLÓGICOS Y LA PRODUCCIÓN DE PEREZONA.....	84

2.1. INTRODUCCIÓN.....	85
2.2. MATERIAL Y MÉTODOS.....	85
2.2.1 Análisis edáfico.....	85
2.2.2. Análisis de datos.....	86
2.3. RESULTADOS.....	86
2.3.1 Correlación entre factores edáficos y la producción de perezona.	86
2.4. DISCUSIÓN.....	88
2.4.1. Influencia de factores edáficos sobre la producción de perezona.	88
2.5. CONCLUSIÓN.....	89
2.6. PERSPECTIVAS.	89
REFERENCIAS.....	90
ANEXOS	98



ÍNDICE DE FIGURAS.

FIGURA 1. Planta silvestre de <i>A. cordata</i> .	4
FIGURA 2. Distribución geográfica del género <i>Acourtia</i> .	4
FIGURA 3. Estructuras química de los compuestos obtenidos de especies del género <i>Acourtia</i> .	10
FIGURA 4. Sistemas de producción de perezona.	16
FIGURA 5. Esquema de las rutas biosintéticas de los terpenos y sus precursores en una célula vegetal.	18
FIGURA 6. Importancia de la producción de compuestos volátiles como los terpenos durante el crecimiento y desarrollo de las plantas.	19
FIGURA 7. Esquema de las estrategias de “bottom-up” y “top-down” en proteómica.	30
FIGURA 8. Representación esquemática de la técnica de electroforesis bidimensional.	33
FIGURA 9. Representación esquemática de las partes principales que componen un espectrómetro de masas.	42
FIGURA 10. Comparación esquemática entre la huella de masa del péptido y la identificación de la proteína MS / MS.	44
FIGURA 11. Fragmentación de la nomenclatura iónica de un tetrapéptido protonado.	45
FIGURA 12. Cuantificación de proteínas con SILAC.	53
FIGURA 13. Ejemplares de <i>A. cordata</i> depositados en HUMO-Herbarium del Centro de Investigación de Biodiversidad y Conservación de la UAEM.	61
FIGURA 14. Flujo de trabajo desarrollado.	62
FIGURA 15. Análisis de las imágenes de 2-DE de los extractos proteicos de las raíces productoras de perezona.	70
FIGURA 16. Análisis de enriquecimiento de la anotación funcional de las proteínas sobreexpresadas en plantas con alta producción de perezona.	73
FIGURA 17. Análisis de enriquecimiento de las vías KEGG de las proteínas diferencialmente expresadas entre los grupos de alta y baja producción de perezona.	75
FIGURA 18. Redes de interacción proteína-proteína de las proteínas diferencialmente sobreexpresadas.	79

ÍNDICE DE TABLAS.

Tabla 1. Especies descritas y aceptadas para el género <i>Acourtia</i> .	5
Tabla 2. Compuestos bioactivos naturales y sintéticos extraídos apartir de plantas del género <i>Acourtia</i> .	9
Tabla 3. Actividades farmacológicas de los compuestos bioactivos extraidos apartir de plantas del género <i>Acourtia</i> .	14
Tabla 4. Proteínas diferencialmente expresadas e identificadas por MALDI-TOF MS en <i>A. cordata</i> .	71
Tabla 5. Media \pm error estándar y resultados de ANOVA para cada parámetro de suelo y para cada carácter macromorfológico medido en tres localidades del corredor biológico el "Chichinautzin", Morelos, México.	87
Tabla 6. Análisis de regresión múltiple que prueban la influencia de diferentes parámetros del suelo y caracteres macromorfológicos en los contenidos de perezona en <i>Acourtia cordata</i> en tres localidades del corredor biológico el "Chichinautzin", Morelos, México.	87



ABREVIATURAS.

ACN	Acetonitrilo (Acetonitrile)
ADN	Ácido desoxirribonucleico
ATP	Trifosfato de adenosina (Adenosine triphosphate)
BCA	Bicarbonato de amonio
BLAST	Basic local alignment search tool
CHAPS	3 - [(3-Cholamidopropil) dimethylamonio] -1-propanosulfonato (3-[(3-cholamidopropyl)dimethylamonio]-1-propanesulfonate)
DIGE	Electroforesis en gel diferencial (Differential gel electrophoresis)
DMAPP	Dimetilalil pirofosfato
DMAPP	Dimetilalil difosfato
DTT	Ditiotreitol (Dithiothreitol)
DXP	Desoxixilulosa 5-fosfato
EDTA	Ácido etilendiaminetetraacético (Ethylenediaminetetraacetic acid)
EM/MS	Espectrometría de masas (Mass spectrometry)
2-DE	Electroforesis en gel bidimensional
ESI	Ionización por electropray (Electrospray ionization)
ESI-MS/MS	Espectrometría de masas en tándem con ionización por electropray (Electrospray ionization tandem mass spectrometry)
EST	Etiquetas de secuencia expresada
FPP	Farnesil difosfato
GAP	Gliceraldehído 3-fosfato
GGPP	Geranilgeranil difosfato
GPP	Geranil difosfato
IAM	Iodoacetamida (Iodoacetamide)
IT	Trampa de iones (Ion trap)
IPP	Isopentenil pirofosfato
IPP	Isopentenil difosfato
LC	Cromatografía Líquida (Liquid chromatography)
M.O.	Materia Orgánica
MALDI-TOF	Matriz asistida por láser desorción / ionización-Tiempo de vuelo (Matrix assisted laser desorption/ionization-Time-of-flight)
MAV/MEV	Ácido mevalónico (Mevalonic acid)
MEP	2-C-metil-D-eritritol-4-fosfato
MPT/PTM	Modificaciones postraduccionales/Post-translational modification
MS	Metabolitos secundarios
MS/MS	Espectrometría de masas en tándem (tandem mass spectrometry)
MudPIT	Tecnología de Identificación de Proteínas Multidimensional (Multidimensional Protein Identification Technology)
NADPH	Nicotinamida adenina dinucleótido fosfato (Nicotinamide adenine dinucleotide phosphate)
NH₄	Amonio

NO₃	Nitrato
N_t	Nitrogeno total
ORF	Marco de lectura abierto (Open reading frame)
PO₄	Fosfato
P_t	Fosforo total
PVP	Polivinilpirrolidona (Polyvinylpyrrolidone)
TCA/acetona	Ácido tricloroacético/acetona
TFA	Ácido trifluoroacético (Trifluoroacetic acid)
TPSs	Terpeno sintasas/ciclasas
UHPLC- QTOF	Cromatografía líquida de ultra alto rendimiento acoplada al cuadrupolo de tiempo de vuelo (Ultra-High-Performance Liquid Chromatography coupled to Quadrupole Time-Of-Flight)



I. RESUMEN.

Acourtia cordata (Cerv.) B. L. Turner (Asteraceae) se conoce comúnmente como "Hierba del zopilote", en la medicina tradicional mexicana se ha empleado para preparar infusiones con propiedades antihelmínticas, laxantes y diuréticas (Flores y Troncoso, 1886). El género *Acourtia* comprende aproximadamente 80 especies, la mayoría de las cuales son endémicas de México (Calderón y Rzedowski 2001). Las raíces de varias especies de este género producen muchos compuestos de origen terpenico, incluida la perezona. La perezona se considera como la primera sustancia aislada y purificada en el nuevo mundo en 1852 y quizá esto sea una de las razones por las que la perezona ha sido ampliamente estudiada desde el punto de vista químico (Joseph-Nathan et al. 1972; Joseph-Nathan et al. 1989).

Además, se han descrito varias actividades farmacológicas que la perezona y sus derivados poseen tales como: actividad hipoglucemiante, inhibición de la agregación plaquetaria, inducción de respuesta contráctil en el músculo liso intestinal y protección contra los efectos nocivos de la isquemia y la reperfusión, entre otros. También se han estudiado los efectos antialimentarios de los derivados de perezona contra larvas de plagas de insectos de importancia agrícola y su actividad fitotóxica (Burgueño-Tapia et al. 2008).

Por otro lado, se han realizado estudios sobre la producción biotecnológica de la perezona, los cuales han demostrado que los cultivos de callos, raíces pilosas y las plántulas cultivadas *in vitro* de algunas especies de *Acourtia* producen y acumulan cantidades menores de perezona en comparación con lo que producen las raíces de plantas silvestres (Arellano et al. 1996; Gómez-Serrano et al. 2010; Gómez-Serrano et al. 2012). Esta biosíntesis y acumulación de metabolitos secundarios (MS) puede variar en una especie de planta debido a diferencias ambientales del lugar donde crecen; tales factores ambientales incluyen altitud, latitud, precipitación, temperatura, humedad, radiación solar, entre otros (Bryant et al 1995; Liu et al. 2015).

La variación cuantitativa y cualitativa de MS en las especies de plantas ocurre entre los diferentes tipos de tejidos, las etapas ontogenéticas (Koricheva y Barton 2012), la edad y el tamaño de las plantas (Andrew et al. 2007). Por lo tanto, la estructura y la edad de la población pueden contribuir a la variación de MS (Moore y Foley 2005). Así mismo, los

componentes edáficos también contribuyen a la expresión de diferentes caracteres químicos en las plantas (Hamilton et al. 2001). El fósforo (P) y el nitrógeno (N) son los dos elementos más limitantes en el suelo para las plantas, ya que alteran diversos procesos fisiológicos y bioquímicos (Pradhan et al. 2017; Ramakrishna y Ravishankar 2011). Se ha documentado que N y P pueden aumentar las tasas fotosintéticas y la acumulación de carbohidratos para el crecimiento de las plantas (Pinheiro y Chaves 2011), también pueden modificar la biosíntesis de MS, como ácidos fenólicos, flavonoides, taninos y terpenos, entre otros (Ormeño y Fernández 2012).

Por otro lado, considerada como una herramienta importante, la proteómica permite la descripción de diversos procesos en el reino vegetal, como los cambios morfogenéticos, fisiológicos y bioquímicos durante el desarrollo de tejidos y órganos (Di Michele et al. 2006; Tian et al. 2009; Mohamad et al. 2011; Takáč et al. 2012; Martínez-Márquez et al. 2013; Martínez-Esteso et al. 2013). La respuesta a diversos estímulos ambientales y diferentes estreses bióticos y abióticos (Sargento y Renaut 2010; Fernández-García et al. 2011; Trupiano et al. 2012; Aghaei y Komatsu 2013; Ghosh y Xu 2014; Fang et al. 2015; García de la Garma et al. 2015; Chmielewska et al. 2016; Martínez-Esteso et al. 2016). Además, permite el estudio de proteínas involucradas en rutas biosintéticas que conducen a la biosíntesis de MS de importancia farmacológica (Senthil et al. 2011; Martínez-Esteso et al. 2011; Champagne et al. 2012; Bhattacharyya et al. 2012; Ma et al. 2013; Sud et al. 2014; Bryant et al. 2015; Martínez-Esteso et al. 2015).

Los grandes avances en proteómica vegetal son en gran parte derivados del uso de la cromatografía líquida acoplada a la espectrometría de masas y el progreso en herramientas bioinformáticas. El claro progreso en el flujo de trabajo proteómico ha sido debido al análisis de proteínas mediante la tecnología de identificación multidimensional (MudPIT) utilizando el método “bottom-up”, considerado como uno de los flujos de trabajo más comunes en proteómica de alto rendimiento (Baginsky 2009). Derivado de esta tecnología, se han realizado muchas investigaciones para el análisis de mezclas complejas de proteínas obtenidas de tejidos de especies de plantas modelo y no modelo (Bayer et al. 2006; Armengaud et al. 2014). Aunque hay muchos estudios proteómicos con diferentes enfoques en plantas modelo, estos todavía son pocos en plantas no modelo, principalmente en plantas medicinales (Oldham et al. 2010).

En cuanto a *A. cordata* hasta ahora no existen reportes sobre el proteoma de las raíces ni tampoco se han realizado estudios comparativos con el fin de identificar las proteínas involucradas en la biosíntesis de la perezona y sus derivados. Por tanto, es de gran interés biológico llevar a cabo una exploración y comparación de los perfiles proteómicos de las raíces de *A. cordata* con distintos niveles de producción de perezona. Además de analizar la influencia de factores edáficos en la producción de perezona en plantas silvestres. En el presente trabajo se muestran los resultados de la comparación y exploración de los perfiles de proteínas obtenidos de las raíces de plantas silvestres que producen diferentes niveles de perezona, además de analizar la relación entre la producción de perezona en *A. cordata* con varios parámetros del suelo (contenido de materia orgánica [OM], pH, NO₃, PO₄, NH₄, contenido de N total y P total) y caracteres de las plantas (diámetro basal, altura individual total, número de hojas y cobertura vegetal).

II. ANTECEDENTES.

II.1. El género *Acourtia*.

El género *Acourtia*, consta de aproximadamente 80 especies (tabla 1), algunas de sus especies son conocidas comúnmente como “pipitzáhuac”, “pipichaque”, “matarique”, “hierba del zopilote”, “zazanaca” y “coapatli” (Cabrera R., 2001; Burgueño-Tapia, 2008; Artschwager-Kay, 1977; Alarcon-Aguilar, 1996 y Andrade-Cetto, 2005; Rocha-Gracia., et al, 2011.), dentro de este genero se encuentra *Acourtia cordata* (figura 1), sus plantas son conocidas por su uso medicinal desde épocas precolombinas. La mayoría de las especies son endémicas de México, algunas especies se extienden desde el sur de los Estados Unidos hasta América Central (figura 2). Estas especies se encuentran en áreas secas a mesicas y desde cerca del nivel del mar hasta grandes altitudes en las montañas de México, son hierbas perennes con un caule aéreo bien desarrollado (Cabrera R., 2002). La mayoría de las especies de *Acourtia* tienden a crecer en áreas no perturbadas, aunque algunas están ampliamente distribuidas y toleran hábitats perturbados. En general, las poblaciones de los taxones son pequeñas y con frecuencia se encuentran en áreas aisladas de acceso moderado a difícil (Cabrera R., 2001).



FIGURA 1. Planta silvestre de *A. cordata*



FIGURA 2. Distribución de género *Acourtia*.
*Tomado de *Global Biodiversity Information Facility*

Tabla 1. Especies descritas y aceptadas para el género *Acourtia*.

Especie	Autor	Distribución (Estado de la Republica Mexicana)
<i>Acourtia bravohollisiana</i>	Rzed.	GRO
<i>Acourtia butandae</i>	L. Cabrera	CHIH, SON
<i>Acourtia caltepecana</i>	B.L. Turner	PUE
<i>Acourtia carpholepis</i>	(Sch. Bip. ex A. Gray) Reveal & R.M.King	CHIS, HGO, OAX, PUE, QRO, SLP, VER
<i>Acourtia carranzae</i>	L. Cabrera	QRO, SLP
<i>Acourtia ciprianoi</i>	Panero & Villaseñor	OAX
<i>Acourtia cordata</i>	Cerv.) B.L. Turner	CDMX, GRO, HGO, MEX, MOR, OAX, PUE, QRO, SLP, TLAX, VER
<i>Acourtia coulteri</i>	(A. Gray) Reveal & R.M. King	CHIS, GTO, HGO, NLE, QRO, SLP, TAMS, VER, ZAC
<i>Acourtia cuernavacana</i>	(B.L. Rob. & Greenm.) Reveal & R.M.King	GRO, MEX, MICH, MOR
<i>Acourtia dieringeri</i>	L. Cabrera	CHIH, SON
<i>Acourtia discolor</i>	Rzed.	GRO, OAX
<i>Acourtia dissiticeps</i>	(Bacig.) Reveal & R.M. King	AGS, SLP
<i>Acourtia dugesii</i>	A. Gray) Reveal & R.M. King	AGS, COL, GTO, GRO, HGO, JAL, MEX, MICH, MOR, OAX, PUE, QRO, SLP, VER
<i>Acourtia durangensis</i>	B.L. Turner	DGO
<i>Acourtia elizabethiae</i>	Rzed. & Calderón	HGO, QRO, SLP
<i>Acourtia erioloma</i>	(S.F. Blake) Reveal & R.M. King	OAX
<i>Acourtia fragrans</i>	Rzed.	PUE
<i>Acourtia fruticosa</i>	(Lex.) B.L. Turner	AGS, COL, CDMX, DGO, GTO, HGO, JAL, MEX, MICH, NAY, QRO, SLP, SIN, ZAC
<i>Acourtia gentryi</i>	L. Cabrera	SIN
<i>Acourtia glandulifera</i>	(Nash) B.L. Turner	CHIS
<i>Acourtia glomeriflora</i>	(A. Gray) Reveal & R.M. King	AGS, COL, GRO, JAL, MEX, MICH, MOR, OAX, ZAC
<i>Acourtia grandifolia</i>	(S. Watson) Reveal & R.M. King	JAL, MICH, NAY, ZAC
<i>Acourtia gracilis</i>	L. Cabrera	PUE, VER
<i>Acourtia hidalgoana</i>	B.L. Turner	GTO, HGO, NLE, QRO, SLP, TAMS, VER
<i>Acourtia hooveri</i>	(McVaugh) Reveal & R.M. King	COL, JAL, MICH

<i>Acourtia huajuapana</i>	B.L. Turner	GRO, MEX, MOR, OAX,PUE
<i>Acourtia humboldtii</i>	(Less.) B.L. Turner	COL, CDMX, GTO,GRO, HGO, JAL, MEX, MICH, PUE, QRO, SLP, TLAX, VER
<i>Acourtia intermedia</i>	L. Cabrera	DGO
<i>Acourtia joaquinensis</i>	L. Cabrera	HGO, QRO
<i>Acourtia lepidopoda</i>	(B.L. Rob.) Reveal & R.M. King	JAL,MEX, MICH, MOR
<i>Acourtia lobulata</i>	(Bacig.) Reveal & R.M. King	OAX, PUE
<i>Acourtia longifolia</i>	(S.F. Blake) Reveal & R.M. King	DGO,JAL, NAY, ZAC
<i>Acourtia lozanoi</i>	(Greenm.) Reveal & R.M. King	CDMX,HGO, MEX, PUE, QRO, SLP, TLAX, VER
<i>Acourtia macrocephala</i>	Sch. Bip.	DGO, NAY
<i>Acourtia macvaughii</i>	B.L. Turner	GRO, MICH
<i>Acourtia mexiae</i>	L. Cabrera	JAL
<i>Acourtia mexicana</i>	(Lag. ex D. Don) H. Rob.	AGS, CHIH, DGO,SIN, SON, ZAC
<i>Acourtia microcephala</i>	DC.	BCN
<i>Acourtia michoacana</i>	(B.L. Rob.) Reveal & R.M. King	COL,JAL, MICH
<i>Acourtia moctezumae</i>	Rzed. & Calderón	QRO, SLP
<i>Acourtia moschata</i>	(La Llave) DC.	GTO, GRO, HGO, JAL,MEX, MICH, QRO, TLAX
<i>Acourtia nana</i>	(A. Gray) Reveal & R.M. King	AGS, CHIH,COAH, DGO, GTO, JAL, NLE, QRO, SLP, SON, TAMS, ZAC
<i>Acourtia nelsonii</i>	(B.L. Rob.) Reveal & R.M. King	COL, JAL,MICH, NAY
<i>Acourtia nudicaulis</i>	(A. Gray) B.L. Turner	CHIS
<i>Acourtia nudiuscula</i>	(B.L. Rob.) B.L. Turner	JAL, NAY
<i>Acourtia oaxacana</i>	L. Cabrera	OAX
<i>Acourtia ovatifolia</i>	L. Cabrera	GRO, OAX, PUE
<i>Acourtia oxylepis</i>	(A. Gray) Reveal & R.M. King	GTO, HGO,QRO, SLP
<i>Acourtia palmeri</i>	(S. Watson) Reveal & R.M. King	BCN, BCS
<i>Acourtia parryi</i>	(A. Gray) Reveal & R.M. King	AGS, CHIH,COAH, DGO, GTO, HGO, JAL, NLE, QRO, SLP, TAMS, VER,ZAC
<i>Acourtia patens</i>	(A. Gray) Reveal & R.M. King	CHIH, COL,DGO, JAL, NAY, SIN, SON
<i>Acourtia pilulosa</i>	(Bacig.) B.L. Turner	GRO, MOR, OAX,PUE
<i>Acourtia pinetorum</i>	(Brandege) Reveal & R.M. King	BCS
<i>Acourtia platyphylla</i>	(A. Gray) Reveal & R.M. King	CHIH,DGO
<i>Acourtia platyptera</i>	(B.L. Rob.) Reveal & R.M. King	GRO,MEX, MICH

<i>Acourtia potosina</i>	L. Cabrera	SLP
<i>Acourtia pringlei</i>	(B.L. Rob. & Greenm.) Reveal & R.M. King	GRO, MEX, MICH
<i>Acourtia pulchella</i>	L. Cabrera	AGS, ZAC
<i>Acourtia purpusii</i>	(Brandege) Reveal & R.M. King	COAH,GTO, HGO, JAL, NLE, QRO, SLP, TAMS, VER, ZAC
<i>Acourtia queretarana</i>	B.L. Turner	GTO, NLE, QRO, SLP
<i>Acourtia reticulata</i>	(Lag. ex D. Don) Reveal & R.M. King	CHIS,DGO, GTO, GRO, HGO, JAL, MEX, MICH, MOR, OAX, PUE,QRO, SLP, VER
<i>Acourtia runcinata</i>	(Lag. ex D. Don) B.L. Turner	CHIH, COAH,HGO, NLE, QRO, SLP, TAMS, VER, ZAC
<i>Acourtia rzedowskii</i>	B.L. Turner	PUE, VER
<i>Acourtia scapiformis</i>	(Bacig.) B.L. Turner	GRO, OAX, PUE
<i>Acourtia scaposa</i>	(S.F. Blake) B.L. Turner	MICH
<i>Acourtia simulata</i>	(S.F. Blake) Reveal & R.M. King	AGS,COL, GRO, JAL, MEX, MICH, MOR, ZAC
<i>Acourtia sinaloana</i>	B.L. Turner	SIN
<i>Acourtia tenoriensis</i>	B.L. Turner	OAX, PUE
<i>Acourtia tomentosa</i>	(Brandege) Reveal & R.M. King	GTO,NLE, PUE, QRO, SLP, TAMS
<i>Acourtia turbinata</i>	(Lex.) DC.	AGS, CAM, COL, CDMX,GTO, GRO, JAL, MEX, MICH, MOR, QRO, VER, ZAC
<i>Acourtia umbratilis</i>	(B.L. Rob. & Greenm.) B.L. Turner	OAX,PUE
<i>Acourtia venturae</i>	L. Cabrera	GTO, SLP
<i>Acourtia veracruzana</i>	B.L. Turner	CDMX, HGO, MEX, OAX,PUE, QRO, VER
<i>Acourtia wislizeni</i>	(A. Gray) Reveal & R.M. King	AGS, CHIH,COL, DGO, GTO, JAL, MEX, MICH, NAY, QRO, SLP, SIN,ZAC
<i>Acourtia wrightii</i>	A. Gray) Reveal & R.M. King	AGS, CHIH,COAH, DGO, GTO, JAL, NLE, SLP, SON, TAMS, ZAC
<i>Acourtia zacatecana</i>	B.L. Turner	AGS, DGO, JAL, ZAC

*Información tomada de Villasenor 2016.

II.2. Compuestos bioactivos del género *Acourtia*.

En 1852, el doctor Leopoldo Río de la Loza, célebre investigador mexicano, reportó una sustancia extraída a partir de las raíces de estas plantas, que llamó ácido pipitzahóico, hoy conocida como perezona, notable por sus efectos purgantes en el cuerpo humano, sustancia que posteriormente fue reportada por servir como colorante para teñir fibras de lana y algodón, es por tal razón que la perezona tiene un uso potencial en la industria de artes gráficas y textiles (Gómez-Serrano y col., 2010).

La perezona en forma purificada es una sustancia sólida que forma cristales de color amarillo-anaranjado y tiene un peso molecular de 248.31 g/mol. Posee las propiedades de una quinona: se reduce fácilmente, se reoxida espontáneamente y produce derivados aminados (figura 3). Su estructura química fue determinada mediante resonancia magnética protónica por Walls y col., en 1983 y fué confirmada mediante síntesis química por Sánchez y col., en 1983. Esta sustancia ha sido aislada a partir de raíces de *A. adnata*, *A. alamani* y *A. michoacana* entre otras, y es muy abundante en *Acourtia cuernavaca* (Robinson y col., 2006) y *Acourtia cordata*, donde se llegan a obtener rendimientos entre 2 y 8 % en base a peso seco (Joseph-Nathan y Santillán 1989, García-Méndez, 2010 y 2012).

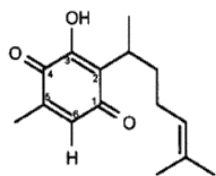
La perezona es un sesquiterpeno benzoquinóico que probablemente deriva de la biosíntesis de terpenos, a través de la ruta citosólica del mevalonato (Sánchez y col., 1983).

Además de la producción de la perezona en las raíces de diversas especies de *Acourtia* existen otros compuestos naturales como el cipreno, acurcumeno, parvifolina, isovalerato extraídos a partir de *Acourtia longifolia*. El α -pipitzol y β -pipitzol a partir de las raíces de *Acourtia carpholepis*. Y también se han producido compuestos sintéticos derivados de la perezona tales como la isoperezona, dihydroperezona, dihydroisoperezona, anilidoperezona, dihidroxiperezona. Así mismo se han identificado otros compuestos como 5-methyl coumarinas, entre otras (tabla 2).

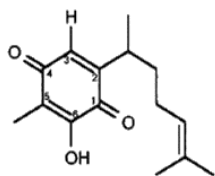
Varios de estos compuestos presentan diversas actividades biológicas, algunas de ellas son descritas en la tabla 3.

Tabla 2. Compuestos bioactivos naturales y sintéticos extraídos apartir de plantas del género *Acourtia*.

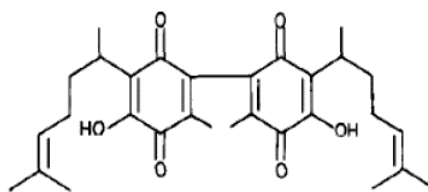
Especie	Compuesto	Referencia
<i>Acourtia alamani</i>	Perezona (2-(1,5-dimethyl-4-hexenyl)-3-hydroxy-5methyl-1,4-benzoquinone), hidroxiperezona, cipereno, parvifolino, isoparvifolinona, 8-hydroxypereflorin y diperezona.	Joseph-Nathan y González 1971; Joseph-Nathan et al., 1982.
<i>Acourtia adnata</i>	Perezona (2-(1,5-dimethyl-4-hexenyl)-3-hydroxy-5methyl-1,4-benzoquinone), pipitzol.	Bugueño-Tapia et al., 2008.
<i>Acourtia carpholepis</i>	Perezona (2-(1,5-dimethyl-4-hexenyl)-3-hydroxy-5methyl-1,4-benzoquinone), cipereno, Parvifolina, diperezona, α -pipitzol y β -pipitzol.	García et al., 1988.
<i>Acourtia cordata</i>	Perezona (2-(1,5-dimethyl-4-hexenyl)-3-hydroxy-5methyl-1,4-benzoquinone)	Joseph-Nathan, 1974, Joseph-Nathan y Santillan, 1989.
<i>Acourtia cuernavacana</i>	Perezona (2-(1,5-dimethyl-4-hexenyl)-3-hydroxy-5methyl-1,4-benzoquinone)	Arellano y Hernandez., 1999; Arellano et al., 1996; Soriano-García et al., 2017; Lozada et al., 2012.
<i>Acourtia longifolia</i>	Cipreno, acurcumeno, parvifolina, isovalerato	García et al., 1988.
<i>Acourtia michoacana</i>	Perezona (2-(1,5-dimethyl-4-hexenyl)-3-hydroxy-5methyl-1,4-benzoquinone)	Caravez y Sandoval., 1988.
<i>Acourtia multiflora</i>	Perezona (2-(1,5-dimethyl-4-hexenyl)-3-hydroxy-5methyl-1,4-benzoquinone), cumarina 3,4,8-trimethoxy-6-hydroxy-5-formyl-2Hl-benzopyran-2-one, isocedrenos	Joseph-Nathan et al., 1978.
<i>Acourtia nana</i>	Isocedrenos, lupeyl acetato,	Takakura y Yamamura, 1999; Zdero et al., 1991.
<i>Acourtia platyphilla</i>	Perezona (2-(1,5-dimethyl-4-hexenyl)-3-hydroxy-5methyl-1,4-benzoquinone)	Escobeso-Gonzale et al., 2019.
<i>Acourtia runcinata</i>	Perezona (2-(1,5-dimethyl-4-hexenyl)-3-hydroxy-5methyl-1,4-benzoquinone), metil senecionato, isocedrenos	Zdero et al., 1988.
<i>Acourtia thurberi</i>	Perezona (2-(1,5-dimethyl-4-hexenyl)-3-hydroxy-5methyl-1,4-benzoquinone), 8- β -D-glucopyranosyloxy-4-methoxy-5-methyl-coumarina, α -pipitzol y β -pipitzol.	Espinoza-Vázquez., et al, 2017; Martínez et al., 2016.
<i>Acourtia turbinata.</i>	2-(Z- 1,5-dimethyl-6-oxo-4-hexenyl)-5-methyl p-benzoquinone, 1,3-acetoxanthorrhizol, parvifolinona, parvifolina, pipitzol, isovalerato de curcuquinol y isoparvifolinona	Mendoza y García 19993.
<i>Acourtia thyrsoides</i>	Perezona (2-(1,5-dimethyl-4-hexenyl)-3-hydroxy-5methyl-1,4-benzoquinone)	Arellano y Hernandez., 1999.
<i>Acourtia humboldtii</i>	Parvifolina	Silva-García et al., 2018.



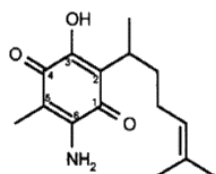
Perezona



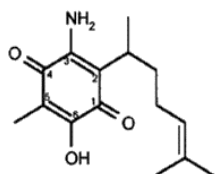
Isoperezona



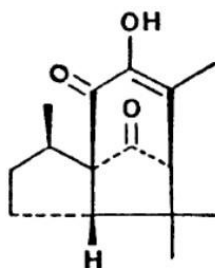
Diperezona



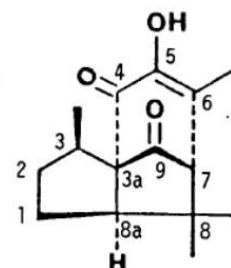
Aminoperezona



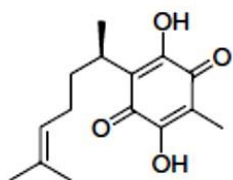
Isoaminoperezona



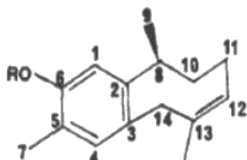
α -pipitzol



β -pipitzol

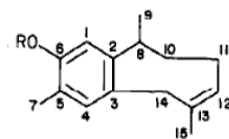


Dihidroiperezona



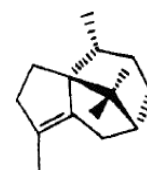
1a R=H
1b R=iVal

Parvifolina (1a)
Parvifolina
isovalerato (1b)

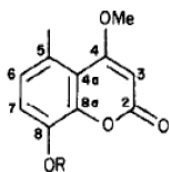


1a R=H
1b R=Ac

Parvifolina (1a)
Acetato de
Parvifolina (1b)

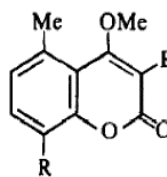


Cipereno

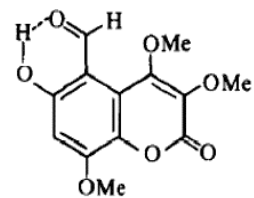


3a R=H
3b R=Me

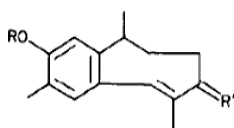
8-hydroxypereflorina (3a)
8-metoxypereflorin (3b)



1: R = R' = H
2: R = H; R' = OMe
3: R = OMe; R' = H
4: R = R' = OMe

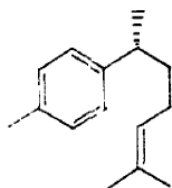


Cumarinas



2a R=H, R'=O
2b R=Ac, R'=H₂

Isoparvifolinona (2a)



α -Curcumeno

Acetato de
Isoparvifolinona (2b)

FIGURA 3. Estructuras químicas de los compuestos obtenidos de especies del género *Acourtia*.

II.3. Propiedades biológicas del género *Acourtia*.

En la medicina tradicional Mexicana las raíces de *Acourtia* se empleaban con efectos, laxantes, diuréticos, regenerativos y analgésicos. Actualmente se le atribuyen otras actividades como la hipoglucemiante, vasodilatador, citotóxico, antialimentaria, fitotóxica, antibacteriana entre otras propiedades biológicas (tabla 3).

a) **Actividad hipoglucemiante.**

Acourtia thurberi ha sido investigada por su efecto reductor de la glucosa en sangre. La actividad hipoglucemiante de esta planta fue evaluada administrando a ratones de manera oral una solución acuosa por decocción a partir de las raíces realizada de acuerdo con el procedimiento hecho por Lorke. Se observó por un lado que los niveles de glucosa en sangre disminuyeron significativamente administrando la solución de la decocción (31.6–316.2 µg / pata) y por el otro lado se reportó que la perezona y 8-β-D-glucopyranosyloxy-4-methoxy-5-methyl-coumarina (3.2–31.6 µg / pata) produjeron una inhibición significativa en el tiempo de lamer durante la prueba de formalina en ratones sanos e hiperglucémicos, demostrando su potencial antinociceptivo y antihiperalgésico respectivamente (Martinez et al., 2016).

b) **Actividad antibacterial.**

El extracto metanólico de las raíces de *Acourtia cordata* presentaron una importante inhibición del crecimiento bacteriano de *Staphylococcus aureus* a una concentración mínima inhibitoria de 250 µg/disc (Rocha-Gracia et al., 2011).

c) **Actividad citotóxica.**

Se han reportado algunos estudios sobre la actividad citotóxica que presenta la perezona y otros sesquiterpenos benzoquinóicos análogos a la perezona.

La perezona aislada de *Acourtia* spp y su isómero la isoperezona mostraron un efecto citotóxico significativo en la línea K-562 (leucemia mielógena crónica humana), aunque los dos compuestos presentaron actividad citotóxica, la perezona comparado con su isómero tuvo un mayor efecto a una concentración de 50 µM, ambos compuestos inducen la muerte celular dependiendo de la actividad de caspasas pero también independientes de este mecanismo (Sanchez-Torres et al., 2010).

Lozada y colaboradores en 2012 reportaron que la perezona y su isómero la isoperezona, compuestos extraídos de *Acourtia cuernavacana*, presentaron actividad citotóxica contra cuatro líneas celulares evaluadas observando un $IC_{50} < 10 \mu M$. Los resultados mostraron que la isoperezona fue dos veces más potente que la perezona contra la línea PC-3 (próstata humana) y la línea HCT-15 (colorrectal humano) con un $IC_{50} = 3.3 \pm 0.2 \mu M$ y un $IC_{50} = 5.9 \pm 0.5 \mu M$ respectivamente, se sugiere que esta diferencia encuento a la actividad citotóxica puede ser debida al cambio en la posición del grupo hidroxilo presente en laperezona y la isoperezona. El ensayo demostró también el compuesto (R)-Methyl 2-amino-3-(1-(4-hydroxy-2-methyl-5-((R)-6-methylhept-5-en-2-yl)-3,6-dioxocyclohexa-1,4-dienyl)-1Hindol-3yl) propanoate es un potente antioxidante con un efecto superior a a-tocoferol ($IC_{50} = 5.564 \pm 0.24 \mu M$).

d) Otras propiedades.

La perezona promueve la liberación de Ca^{2+} intramitocondrial, manteniendo así la producción de ATP durante la reperfusión. La perezona extraída a partir de raíces de *Acourtia adnata* se administró 5 minutos antes de la oclusión coronaria, a una dosis de 3.1 mg / kg, y se pudo observar que este compuesto protege eficazmente contra la alta incidencia de arritmias y la caída de la presión arterial. Además, inhibe la liberación de enzimas deshidrogenasa láctica y creatina-quinasa al plasma. Por lo que se propuso a la perezona como un compuesto con efecto protector de efectos nocivos provocados durante la isquemia y la reperfusión (Téllez et al., 1999).

Además, se comprobó el efecto inhibitorio de la perezona en el proceso de la agregación plaquetaria de la sangre humana. El estudio comparó el efecto inhibitorio de la agregación plaquetaria producido por la perezona, isoperezona, aminoperezona e isoaminoperezona, se observó que solo la perezona tuvo un efecto significativo inhibitorio a $100 \mu M$, mientras que sus derivados no tuvieron el efecto esperado. Los autores proponen que las diferentes respuestas biológicas observadas pueden ser resultado de los distintos grupos químicos presentes cada una de las estructuras de estos compuestos (De la peña y col., 2001).

Por otro lado, estudios previos han demostrado que el sesquiterpeno p-benzoquinona perezona, aislado de *Acourtia adnata*, y sus derivados no naturales isoperezona, dihidroperezona, dihidroisoperezona y anilidoperezona presentan efecto antialimentario

contra los insectos herbívoros *Spodoptera littoralis*, *Leptinotarsa decemlineata* y *Myzus persicae*. Se observó que la perezona y la anilidoperezona exhibieron un fuerte efecto antialimentario provocado en *L. decemlineata* y *M. persicae*, y tuvieron una respuesta baja en *S. littoralis*. Interesantemente, la actividad aumento cuando el hidroxilo del grupo en C-3 en perezona se cambió a C-6 para dar isoperezona. Se encontró el mismo efecto con la hidrogenación del doble enlace de la cadena de alquilo de la perezona para producir dihidroperezona. En contraste, la hidrogenación de este doble enlace en isoperezona para dar dihidroisoperezona provocó una reducción en la actividad antialimentaria. Por otro lado, se determinó la actividad fitotóxica de estos compuestos, se observó que la dihidroperezona tuvo un efecto inhibitor significativo sobre el crecimiento en la longitud de la radícula de *Lactuca sativa* (Burgueño-Tapia y col. 2008).

La perezona también ha sido descrita como un compuesto que ayuda a inhibir la corrosión de metales y ha tomado gran relevancia considerando que la corrosión es el factor responsable de la degradación de oleoductos y tuberías de conductos de gas. En la investigación realizada por Espinoza-Vázquez y colaboradores (2017), la perezona fue extraída a partir de *Acourtia thurberi*, y su potencial inhibitorio de corrosión fue comparado con el de otros compuestos de origen natural. Los resultados mostraron que perezona proporciono buena protección contra la corrosión de acero 1018 sumergido en una solución de NaCl al 3% saturada con CO₂ a una concentración de 10 ppm con eficiencia de = 87% en condiciones estáticas, mientras que en condiciones hidrodinámicas a velocidades de rotación de 100 y 500 rpm y a una concentración de 5 ppm, la eficiencia que alcanzó fue de = 96 y 95%, respectivamente.

Tabla 3. Actividades farmacológicas de los compuestos bioactivos extraídos a partir de plantas del género *Acourtia*.

Actividad	Especie	Parte de la planta	Principio Activo	Organismo testado/ línea celular	Referencia
Antidiabética	<i>Acourtia thurberi</i>	Raíces Parte aerea	-Extracto acuoso de raíces y parte aérea. - Perezona, α -pipitzol, β -pipitzol, 8β -D-glucopiranosiloxi-4D-metoxi-5-metil-coumarina	Ratones y conejos	Aguilar et al., 1994; Alarcon-Aguilar et al., 1997; Andrade et al., 2005; Martinez et al., 2016; Martinez et al., 2017.
Vasodilatador	<i>Acourtia adnata</i>	Raíces	-Perezona y pipitzol	Ratas	Luna-Vazquez et al., 2018.
Antibacterial	<i>Acourtia cordata</i>	Raíces	-Extracto MeOH	<i>Staphylococcus aureus</i>	Rocha-Gracia et al., 2011.
Citotóxico	<i>Acourtia cuernavacana</i>	Raíces	-Perezona e Isoperezona	Línea PC-3 (próstata humana), línea HCT-15 (colorrectal humano), línea, SKLU-1 (cáncer de pulmón) y línea K-562 (leucemia mielógena crónica humana).	Lozada et al., 2012.
	<i>Acourtia spp</i>	Raíces	-Perezona e Isoperezona	K-562 (leucemia mielógena crónica humana).	Sanchez-Torres et al., 2010.

II.4. Producción biotecnológica de perezona.

En cuanto a la producción de perezona empleando métodos biotecnológicos, el primer reporte sobre su producción mediante cultivo de tejidos vegetales, es de Alcántara (1992), quien obtuvo callo a partir de hojas de *A. cuernavacana*. Por otra parte, dado que la perezona se acumula en las raíces de *Acourtia*, se generaron raíces transformadas o “hairy roots”, a partir de segmentos nodales derivados de plántulas germinadas *in vitro* de *A. thyrsoidea* y *A. cuernavacana* vía transformación genética mediada por *Agrobacterium rhizogenes* (Arellano y Hernández, 1999).

Además, se han desarrollado varios sistemas de cultivo *in vitro* de *A. cordata* para la producción biotecnológica de perezona, sin embargo los rendimientos obtenidos son muy bajos (Garrocho-Villegas, 2001; García-Méndez, 2010 y García-Méndez, 2012). Es evidente que el principal problema en la producción de perezona mediante el uso de estas herramientas, es su bajo rendimiento en comparación con la planta silvestre, por lo que desde hace algunos años se han hecho esfuerzos para incrementar dichos rendimientos.

García-Méndez (2012), realizó un análisis histológico para tratar de establecer de manera indirecta el sitio de síntesis y/o acumulación de perezona tanto en raíces de plantas silvestres como en las raíces de plántulas regeneradas *in vitro*. Con base en el hecho de que la perezona se reduce fácilmente y se oxida espontáneamente. El análisis mostró que la perezona se acumula en el floema de las raíces tanto en plantas silvestres como en plántulas cultivadas *in vitro*. Durante el desarrollo de este trabajo se observó que las raíces de plántulas cultivadas *in vitro* requieren de cierta madurez para empezar a producir perezona. Al respecto Gómez-Serrano et al. (2012) establecieron que la producción de perezona en *A. cordata* es dependiente de la edad de la plántula.

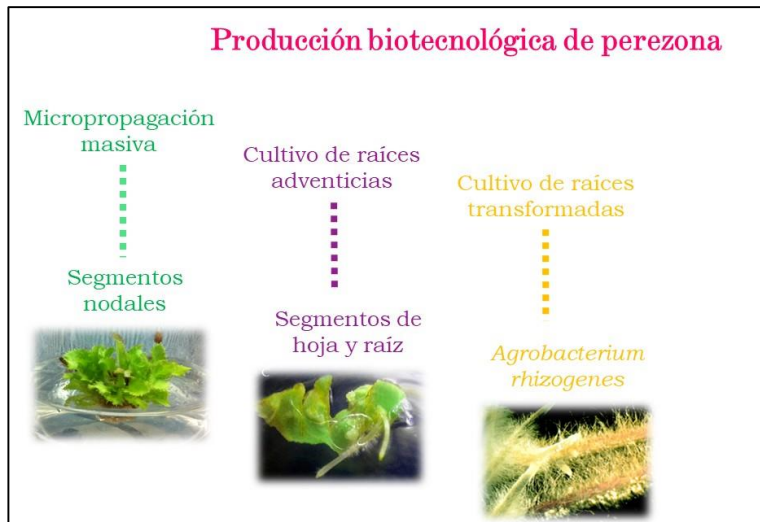


FIGURA 4. Sistemas de producción de perezona.

II.5. Biosíntesis de terpenos.

Los terpenos son compuestos naturales con una amplia variedad de funciones fisiológicas (Croteau y col., 2000). La mayor diversidad de terpenos se encuentra en plantas, los más abundantes en células vegetales son los esteroides y carotenoides, que están involucrados en el mantenimiento de la fluidez de la membrana, así como en la estabilización de las membranas tilacoidales y protección de los componentes fotosintéticos contra el daño foto-oxidativo (Hirschberg, 2001; Fraser y Bramley, 2004). Otros terpenos intervienen en el proceso de modificación post-traducciona de proteínas como la prenilación y la glicosilación y forman parte de la cadena de transporte electrónico. Sin embargo, la mayoría de los terpenos presentes en las plantas son considerados metabolitos secundarios que actúan en la interacción de las plantas con su entorno. Algunos monoterpenos o los sesquiterpenos, desempeñan un importante papel en la defensa de la planta frente a patógenos y herbívoros (Croteau y col., 2000; Chappell, 2002). Además, varios grupos de terpenos son empleados comercialmente como fuente natural de pigmentos, aromas, fibras o ceras. Los terpenos se sintetizan a partir del isopentenil difosfato (IPP) y su isómero dimetilalil difosfato (DMAPP), que a su vez pueden ser sintetizados por dos rutas diferentes: la ruta del ácido mevalónico (Mevalonic acid, MVA) activa en el citosol o a través de la ruta del 2-C-metil-D-eritritol-4-fosfato (2-C-methyl-D-erythritol 4-phosphate, MEP) o ruta plastídica. Cuando se descubrió la vía del ácido mevalónico se asumió que IPP únicamente se sintetizaba a partir de acetil-CoA, sustrato de la ruta, no obstante datos experimentales demostraron que existía la otra vía

independiente de MVA conocida ahora como vía del MEP, en la figura 3 se ilustran estas dos vías (Roberts, 2007; Withers y Keasling, 2007). Se sabe ahora que en las plantas mediante la vía MVA se sintetizan preferentemente sesquiterpenos (15 C) y triterpenos (30 C), mientras que por la vía MEP se sintetizan monoterpenos (10 C), diterpenos (20 C) y tetraterpenos (40 C) (Chen et al., 2004).

La biosíntesis de los monoterpenos y sesquiterpenos es llevada a cabo por una gran familia de enzimas llamadas terpeno sintasas/ciclasas (TPSs). Todas las TPSs tienen propiedades similares con respecto a su masa molecular (ya sean monómeros u homodímeros) y uno de los requisitos para su actividad es la presencia de un ion metálico divalente (normalmente Mg^{2+} o Mn^{2+}).

En los últimos años una serie de monoterpeno sintasas y sesquiterpeno sintasas se han aislado y caracterizado a partir de varias especies de plantas. Generalmente las monoterpeno sintasas son enzimas grandes de entre 600 y 650 aminoácidos, mientras que las sesquiterpeno sintasas están constituidas aproximadamente de 550-580 aminoácidos. Se ha observado que varias terpeno sintasas tienen secuencias homólogas conservadas de aminoácidos ricas en aspartato (Dinesh-Nagegowda., 2010; Degenhardt et al., 2009).

Algunas terpeno sintasas se han purificado y caracterizado apoyándose en herramientas como la cromatografía de permeación en gel y SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis), tal es el caso de la enzima S-linalool sintasa extraída a partir de los estigmas de *Clarkia breweri*, esta enzima cataliza la formación de linalool a partir de geranyl difosfato (GPP) (Pichersky et al., 1995).

También se han caracterizado otras enzimas como la (E)-nerolidol sintasa que participa en la biosíntesis de (E)-nerolidol a partir de farnesil difosfato (FPP) precursor de todos los sesquiterpenos conocidos, este sesquiterpeno es producido como respuesta al ataque de herbívoros en varias especies de plantas tales como *Zea mays* y *Cucumis sativus* entre otras. Se ha propuesto que la conversión de FPP a (E)-nerolidol es una reacción análoga a la conversión de GPP a linalool (Degenhardt et al., 1999).

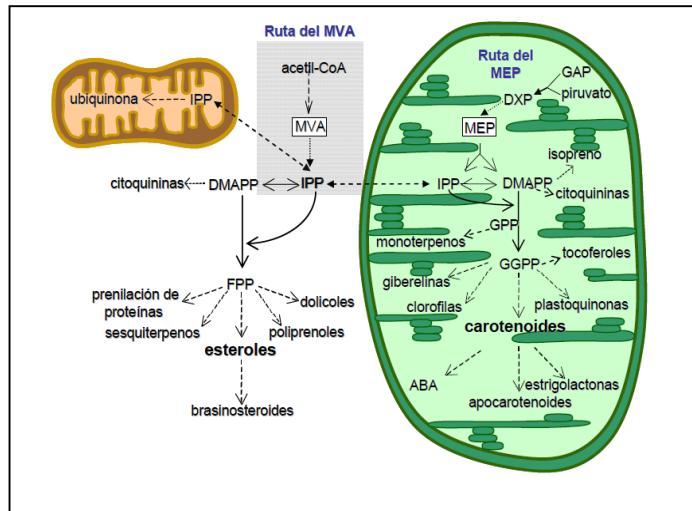


FIGURA 5. Esquema de las rutas biosintéticas de los terpenos y sus precursores en una célula vegetal. MVA, mevalonato; FPP, farnesil difosfato; GAP, gliceraldehído 3-fosfato; DXP, desoxixilulosa 5-fosfato; MEP, 2-C-metil-D-eritritol-4-fosfato; IPP, isopentenil difosfato; DMAPP, dimetilalil difosfato; GPP, geranyl difosfato; GGPP, geranylgeranyl difosfato; ABA, ácido abscísico.

*Editado y tomado de: Ibdah et al., 2019.

II.6. Relación de la producción de terpenos con el estrés biótico o abiótico.

Se sabe que tanto las condiciones bióticas como las abióticas alteran la producción de terpenos en las plantas. Para responder y defenderse del daño ocasionado por una herida y/o el ataque por insectos o microorganismos patógenos, las plantas sintetizan enzimas que degradan la pared celular del microorganismo o tienen la capacidad de inactivar substancias tóxicas de origen microbiano. La composición y la estructura de la pared celular vegetal también cambian, formando una barrera más rígida y menos digerible para insectos. Estas respuestas de defensa a su vez, se combinan con el desarrollo de estructuras contra sus depredadores, tales como las espinas, las espigas, los tricomas y los pelos glandulares. Así mismo y como parte de la protección química, otra estrategia utilizada por las plantas es la producción de MS con actividad antimicrobiana, en contra de herbívoros, o con actividad antioxidante (Croteau et al., 2000). Los MS son compuestos de bajo peso molecular que tienen una gran importancia ecológica porque participan en los procesos de adaptación de las plantas a su ambiente, pues una síntesis activa de MS se induce cuando las plantas son expuestas a condiciones adversas tales como: a) el consumo por herbívoros (artrópodos y vertebrados), b) el ataque por microorganismos: virus, bacterias y hongos, c) la competencia

por el espacio de suelo, la luz y los nutrientes entre las diferentes especies de plantas y d) la exposición a la luz solar u otros tipos de estrés abiótico (Abbas et al., 2017).

Los metabolitos secundarios como los terpenoides están involucrados en diversos procesos fisiológicos y funciones ecológicas basadas en la expresión diferencial de genes y enzimas como las terpeno sintasas como respuesta a factores ambientales bióticos y abióticos durante todo el desarrollo de la planta. Los terpenoides poseen diferentes funciones benéficas para las plantas. Estos compuestos suelen proteger a muchas especies de plantas, contra patógenos, competidores y depredadores transmitiendo mensajes a mutualistas y congéneres sobre enemigos (Gershenson y Dudareva 2007). Los terpenoides también están involucrados en la interacción planta-planta, ayudan a repeler plagas y atraer enemigos de plagas (Unsicker et al. 2009). Los terpenoides volátiles están involucrados especialmente en atraer polinizadores, en la dispersión de semillas, en la defensa contra los herbívoros de abajo y arriba del suelo (Ali et al. 2012; Delory et al., 2016) (figura 6).

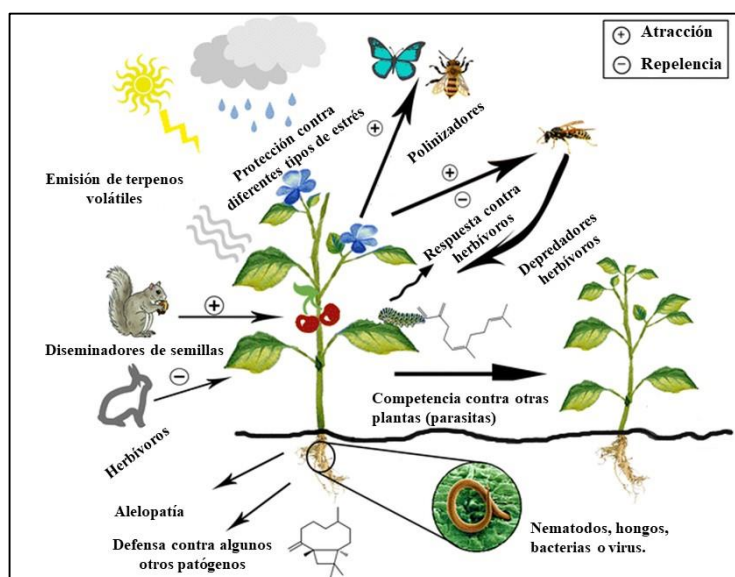


FIGURA 6. Importancia de la producción de compuestos volátiles como los terpenos durante el crecimiento y desarrollo de las plantas. A) Interacción planta-animal: interacciones que comprenden la atracción de diseminadores de semillas y polinizadores. La repelencia /atracción de herbívoros, así como la atracción de enemigos naturales de herbívoros agresivos. B) Interacción con la rizosfera y la atmósfera: las interacciones sobre el suelo que incluye el cebado o la obtención de las respuestas de defensa contra plantas cercanas, por otra parte, en el subsuelo la interacción consiste en la actividad alelopática en la germinación, crecimiento y desarrollo de plantas cercanas competitivas. La emisión de de terpenoides por las raíces y órganos reproductivos poseen actividad antimicrobiana protegiendo de esta manera a la planta contra el ataque de patógenos. Además la emisión por las hojas de estos compuestos le confieren a la planta termotolerancia y fotoprotección.

*Editado y tomado de Abbas et al., 2017.

II.6.1. Influencia de los nutrientes en la producción de terpenos.

La variación cuantitativa y cualitativa de los metabolitos secundarios en las distintas especies de plantas ocurre entre los diferentes tipos de tejidos, las etapas ontogenéticas (Koricheva y Barton 2012), la edad y el tamaño de las plantas (Andrew et al., 2007). Por lo tanto, la estructura y la edad de la población pueden contribuir a la variación de MS (Moore y Foley 2005). Así mismo, los parámetros edáficos también contribuyen a la expresión de diferentes caracteres químicos en las plantas (Hamilton et al., 2001). El fósforo (P) y el nitrógeno (N) son los dos elementos más limitados en las plantas, que alteran diversos procesos fisiológicos y bioquímicos (Pradhan et al., 2017; Ramakrishna y Ravishankar 2011). Se ha documentado que N y P pueden aumentar la acumulación de carbohidratos para el crecimiento de las plantas y las tasas fotosintéticas (Pinheiro y Chaves 2011), también pueden modificar la biosíntesis de MS a base de carbono, como ácidos fenólicos, flavonoides, taninos, terpenos, entre otros (Ormeño y Fernández 2012).

En este sentido uno de los nutrientes que podría promover la síntesis de terpenos en las plantas es el nitrógeno debido a que este nutriente aumenta la tasa de transporte de electrones y la de la fotosíntesis que a su vez provee los requerimientos de adenosín trifosfato (ATP) y la disponibilidad de sustratos como el de carbono. Es probable que el nitrógeno favorezca la biosíntesis de terpenos, especialmente las emisiones de isopreno, y las emisiones de mono y sesquiterpenoides en especies de plantas en que no se almacenan, que a menudo presentan respuestas a corto plazo más fuertes y más rápidas a los factores ambientales debido a la ausencia de reservorios de terpenoides que amortigüen la influencia directa del medio ambiente (Blanch et al., 2007).

Por otro lado, es de esperarse que el fósforo influya en la producción de terpenoides, ya que los precursores de los terpenoides: isopentenil difosfato (IPP) y dimetilalil difosfato (DMAPP) contienen alta energía. Los enlaces fosfato y el fósforo son un componente clave de ATP y nicotinamida adenina dinucleótido fosfato (NADPH) que se requieren para la síntesis de terpenoides. El fósforo podría por lo tanto ser un nutriente limitante involucrado en la emisión de terpenoides y su almacenamiento (Niinemets et al., 2002).

Se cree que un equilibrio existente entre los nutrientes del suelo en donde crece la planta, influye en la producción de compuestos secundarios a nivel de regulación metabólica en las plantas (Herms y Mattson, 1992). La hipótesis del balance de carbono / nitrógeno, según lo propuesto por Bryant y colaboradores en 1983, postula que la fertilización con nutrientes limitantes del crecimiento conducirá a una disminución de las concentraciones de metabolitos secundarios a base de carbono. La hipótesis explica que las concentraciones de metabolitos secundarios en los tejidos vegetales se encuentran en función de la abundancia relativa de los recursos vegetales, particularmente el nitrógeno. La hipótesis se basa en los supuestos de que el crecimiento (metabolismo primario) es una prioridad para las plantas sobre el metabolismo secundario, y que el carbono y el nitrógeno se asignan a la producción de metabolitos secundarios solo después de que se cumplan los requisitos para el crecimiento. También supone que la tasa de producción de metabolitos secundarios está determinada por las concentraciones de moléculas precursoras (Reichardt et al., 1991). Cuando el crecimiento está limitado por la deficiencia de nitrógeno en las plantas, la hipótesis del balance de carbono / nitrógeno predice que los carbohidratos se acumularán en los tejidos de las plantas, lo que conducirá a una mayor síntesis de compuestos secundarios a base de carbono. La interpretación de estos resultados es que la acumulación de metabolitos basados en carbono cuando la relación carbono-nitrógeno es alta se debe a un exceso relativo de carbono fijo disponible para la planta. Por lo tanto, la planta solo puede usar los fotosintatos para los metabolitos que contienen carbono, hidrogeno y oxígeno (C, H y O), como los fenólicos y los terpenos (Waterman y Mole, 1994). De esta manera, se predice que la limitación de nutrientes conduce a una disminución en la producción de compuestos que contienen N, como alcaloides y glucósidos cianogénicos. Aunque muchos estudios han arrojado resultados que son consistentes con estas predicciones (Dixon y Paiva, 1995, Kouki y Manetas, 2002, Simon et al., 2010), en muchos otros estudios ha fallado repetidamente (Koricheva et al., 1998, Riipi et al., 2002).

II.7. Sesquiterpeno-sintasas

Las sesquiterpeno-sintasas también conocidas como terpeno ciclasas son enzimas que catalizan a partir del farnesil difosfato (FPP) la formación de una clase de productos naturales de gran diversidad en estructura, estereoquímica, función biológica y aplicación. Estas enzimas participan en la conversión del FPP a aproximadamente 300 sesquiterpenos

diferentes todos representados con la fórmula química $C_{15}H_{24}$. Las sesquiterpeno-sintasas se caracterizan por su homología estructural, mostrando todas ellas la procedencia de un antecesor común. Entre ellas se encuentran enzimas altamente específicas y otras que producen numerosos sesquiterpenos a partir de un mismo sustrato (Miller y Allemann, 2012). Están formadas por aproximadamente 550-580 amino ácidos equivalentes a 60-65 kDa organizados en su estructura terciaria como α -hélices unidas por cortos “loops” siendo el dominio C-terminal el que contiene el sitio activo de la enzima (Degenhardt y col., 2009). En este sitio activo se reconocen 3 iones divalentes de Mg^{2+} como cofactores tal es el caso de las enzimas α -Farneseno sintasa, Nerolidol sintasa, Amorfa-4,11 dieno sintasa, Epi-cedrol sintasa, etc (Wymore y col., 2011 y Rupasinghe y col., 2000).

Se ha reportado en algunos estudios que el punto isoeléctrico teórico de estas enzimas es de aproximadamente entre 5-6 (Guo et al., 2015; Ee et al., 2014; Su-Fang Ee et al., 2014).

En las plantas, las FPPS son enzimas homodiméricas codificadas por pequeñas familias génicas. En *Arabidopsis* se distinguen tres isoformas FPPS codificadas por dos genes con gran homología de secuencia, FPPS1 y FPPS2. FPPS1 se expresa ampliamente en todos los tejidos de la planta y a lo largo de todas las etapas del desarrollo. Una transcripción diferencial da lugar a dos mRNAs de distinta longitud: una versión larga (FPPS1L) que se expresa principalmente en tejido floral y otra más corta (FPPS1S) que presenta una expresión ubicua (Cunillera et al., 1996, 1997, 2000). Las proteínas codificadas por estos dos transcritos solamente se diferencian en la secuencia N terminal. FPPS1L presenta una extensión de 41 aminoácidos en esta región que la lleva a la mitocondria (Cunillera et al., 1997; Manzano et al., 2004) mientras que FPPS1S se localiza en el citosol (Keim et al., 2012). La isoforma FPPS2 es citosólica (Keim et al., 2012) y presenta un patrón de expresión restringido en órganos y etapas del desarrollo concretos. Se detectan transcritos principalmente en órganos florales, específicamente en los granos de polen, en la raíz, en las zonas de formación de raíces laterales y en los puntos de unión entre la raíz principal y las secundarias y en semillas maduras (Cunillera et al., 2000; Keim et al., 2012). Este patrón de expresión diferencial entre las isoformas, sugiere que FPPS1, está involucrada en la síntesis de isoprenoides que realizan funciones básicas en la célula vegetal mientras que FPPS2 podría estar relacionada con la síntesis de isoprenoides particulares con funciones especializadas. La pérdida de función total de FPPS en mutantes dobles es letal, provocando un arresto del desarrollo del embrión en el

estadio pre-globular, lo que demuestra que la síntesis de FPP es esencial en el desarrollo vegetal. Además, la falta de actividad FPPS dificulta la transmisión del gametofito masculino (Closa et al., 2010).

Se propone que la perezona por ser un sesquiterpeno, sea un compuesto biosintetizado por enzimas como las sesquiterpeno-sintasas a través de la ruta del MVA (figura 5). Y se esperaría que tuvieran las características antes descritas. Hasta el momento, como ya se mencionó, no existen reportes acerca de la elucidación estructural de las enzimas que participan en la producción de la perezona o alguno de sus derivados, para lo cual se ha recurrido a la utilización de las técnicas englobadas en la proteómica, esta disciplina permitirá realizar un estudio que nos proporcione información a través de perfiles proteicos, de toda la gama de proteínas presentes en las raíces productoras y no productoras de *A. cordata*, muchas de las cuales participan en el metabolismo primario y algunas otras que probablemente participan en rutas del metabolismo secundario.

II.8. Proteómica vegetal.

La proteómica vegetal se considera un campo de investigación emergente, que tiene como objetivo principal identificar todas las proteínas expresadas en una célula, órgano o tejido. A través de la proteómica vegetal también es posible crear mapas de expresión de dichas proteínas indicando la localización de las mismas. El creciente interés en los estudios proteomicos en plantas genera un gran número de investigaciones particularmente sobre la división celular, elongación, diferenciación celular, formación de órganos y la inhibición de alguno de estos procesos debido a algún tipo de estrés (Takác, Pechan y Samaj, 2001).

Se sabe que la mayoría de las funciones biológicas que ocurren en los organismos vivos son mediadas por las proteínas. El proteoma de los organismos es modificado en función de factores bióticos y abióticos. Así, el análisis del proteoma de células, órganos o tejidos de plantas se ha llevado a cabo para analizar, por ejemplo, la respuesta a hormonas o moléculas señal, la respuesta a estímulos del ambiente y cambios asociados al desarrollo entre otros.

La proteómica ha ayudado a reconocer proteínas involucradas en los procesos metabólicos básicos (glicolisis, gluconeogénesis, la síntesis del ATP, generación de aminoácidos y nucleótidos) de la célula, o alguno de sus precursores que son esenciales para

la actividad de la célula (Takác, Pechan y Samaj, 2001). La mayoría de los estudios de proteómica vegetal se han llevado a cabo con plantas modelo como *Arabidopsis thaliana*, *Oryza sativa* y *Medicago truncatula* entre otros.

En tales estudios se ha conseguido, mediante la utilización de la proteómica, llevar a cabo la caracterización de los genotipos de dichas especies (Thiellement y col., 2002).

Algunas de las características y procesos son únicos para cada una de las especies de plantas o familias y no pueden ser abordados a través del estudio de una planta modelo. Debido a esto se han llevado a cabo investigaciones en plantas “no modelo” que son esenciales como alimentos o fuentes de energía tales como el maíz, tabaco, trigo, tomate, etc. (Carpentier y col., 2008).

En la actualidad se han realizado estudios proteómicos con plantas como *Panax ginseng*, *Helianthus annuus*, *Spartium junceum* de las cuales su genoma se encuentra parcialmente secuenciado o no ha sido secuenciado.

Desde el año 2006 se han puesto en marcha investigaciones relacionadas con la secuenciación de genomas de algunas especies de la familia Asteraceae, como *Helianthus annuus* (Girasol) y *Lactuca sativa* (Lechuga). Estos avances permitirán en un futuro tener una referencia para el conocimiento de los genes que codifican para proteínas involucradas en la biosíntesis de sesquiterpenos. Esta información es muy útil para el análisis proteómico que se plantea en el presente proyecto de investigación.

Más recientemente, en diversas especies se ha utilizado este tipo de estudio para analizar la interacción de proteínas o sus modificaciones postraduccionales. Además, se han analizado mediante esta metodología la equivalencia y seguridad de alimentos derivados de plantas transgénicas, por lo que los campos de aplicación de la proteómica en plantas han ido en aumento en los últimos años (Chen y Harmon, 2006).

La carencia de la secuencia genómica total de la especie con la que se ha venido trabajando es la primer limitante en este estudio proteómico, pero también es el principal motivo al que debe su mayor atraso o rezago la proteómica en general. Desde otro punto de vista el desarrollo de la proteómica vegetal ha sido posible gracias al avance en técnicas como las de separación de proteínas, la espectrometría de masas y de herramientas bioinformáticas (Chen y Harmon 2006).

II.9. Estudio de proteomas de especies "huérfanas".

A pesar de los muchos logros obtenidos, lo cierto es que la proteómica no ha satisfecho totalmente las expectativas generadas durante sus inicios. Una de las razones fundamentales de ello es que los estudios de proteómica global solo han podido realizarse en organismos modelo, entendiéndose organismo modelo en el campo de la proteómica como aquel bien descrito a nivel de proteína. Este hecho ha excluido o minimizado el uso de las técnicas proteómicas en campos de gran relevancia científica y económica donde no se trabaja con estos organismos. Y el problema se agrava para aquellos organismos mal representados en las bases de datos utilizadas habitualmente en los análisis proteómicos. A estos organismos se les conoce como organismos huérfanos (Carpentier y America 2014).

II.10. Métodos proteómicos.

Las técnicas empleadas en la proteómica como la electroforesis en dos dimensiones combinada con la espectrometría de masas son usadas a gran escala en la investigación para identificar muchas proteínas de varios organismos.

Uno de los descubrimientos más significativos para la investigación en proteomas, ha sido la identificación de proteínas por medio de espectrometría de masas, separadas mediante electroforesis en geles de doble dimensión, lo cual ha ampliado el análisis de estas más allá de su simple visualización. La espectrometría de masas esencialmente ha substituido la técnica clásica de secuenciación de la porción N-terminal de la proteína, por medio de la degradación de Edman.

II.11. Categorías de la proteómica.

La mayoría de los estudios proteómicos de plantas hasta la fecha se pueden dividir en dos categorías básicas (Rose et al., 2004): proteómica analítica y proteómica de expresión (o proteómica de visualización diferencial) (Simpson, 2003a). El primero involucra el perfil de proteínas del material biológico con el objetivo de separar, secuenciar y catalogar tantas proteínas como sea posible. El objetivo es establecer el mapa de proteínas de un sistema biológico comparable a una colección de etiquetas de secuencia expresada (Expressed sequence tag, EST) que proporciona una instantánea del complemento de la transcripción. Sin embargo, los proteomas completos de tipos de células individuales aún no se pueden

mapear completamente en este momento, y se necesita el uso de estrategias complementarias para compilar mapas de proteomas parciales (Ahrens et al., 2010). Por lo tanto, se deben seguir diferentes estrategias para lograr un mapa completo de proteínas dentro de un tejido como el fraccionamiento subcelular. Esta estrategia complementaria podría reducir drásticamente la complejidad de la proteína de un extracto particular y revelar información importante sobre la localización subcelular (Dreger, 2003).

La segunda categoría básica de análisis proteómico se puede denominar como proteómica de expresión diferencial, donde el objetivo no es identificar el conjunto completo de proteínas en una muestra particular, sino caracterizar las diferencias entre las diferentes poblaciones de proteínas. Este enfoque es, por lo tanto, un tanto análogo al perfil comparativo de microarrays de ácido desoxirribonucleico (ADN) y es poderoso para analizar proteínas de plantas silvestres versus mutantes (Kim et al., 2006; Konishi et al., 2005), o tejidos en diferentes etapas de desarrollo (Hochholdinger et al., 2005) o siguiendo las respuestas a estímulos externos (Ndimba, 2005; Chivasa, 2005) como ejemplos ilustrativos.

II.12. Enfoques proteómicos.

Dadas las numerosas opciones técnicas disponibles para el análisis del proteoma, han surgido varias estrategias generales de identificación de proteínas. Un análisis proteómico comienza con la preparación de la muestra en la que las proteínas se digieren enzimáticamente en péptidos (enfoque “bottom-up”) (MacDonald Yates, 2002, 2003; Chait, 2006) o se analizan intactos (enfoque “top-down”) (Breuker et al., 2008; Reid y McLuckey, 2002). La separación se realiza a nivel de proteína para el enfoque de arriba hacia abajo “top-down” o a nivel de péptido para el enfoque “bottom-up”. La separación basada en gel o cromatográfica se utiliza para reducir la complejidad de la muestra antes del análisis de masas. Finalmente, se lleva a cabo la identificación de proteínas a través de la espectrometría de masas. De manera general, la identificación de proteínas requiere la adquisición de datos espectrales de masas en tándem, en la que los iones de péptidos aislados se someten a fragmentación, generalmente mediante disociación inducida por colisión (Collision induced dissociation, CID), para producir un patrón de fragmentación a partir del cual la secuencia de aminoácidos puede reconstruirse.

II.12.1. El enfoque top-down.

La estrategia o proteómica top-down (de arriba hacia abajo), se centra exclusivamente en la caracterización e identificación de proteínas intactas a partir del análisis por MS de las mismas, sin necesidad de recurrir al análisis de sus péptidos (Kelleher, 2004; Bogdanov y Smith, 2005). Esta aproximación identifica las proteínas determinando su masa molecular relativa (M_r) y/o junto con la fragmentación de los iones proteicos obtenidos, generando así información sobre su secuencia. Para ello, es necesario utilizar espectrómetros de masas de alta resolución, tales como el espectrómetro de masas de resonancia ciclotrónica de iones (Fourier Transform Ion Cyclotron Resonance, FTICR) (Bagdanov y Smith, 2005), o las trampas iónicas tipo Orbitrap (Hu y col., 2005), capaces de aislar iones proteicos, determinar su envoltura isotópica e inducir su fragmentación. Esta estrategia confía también en la búsqueda mediante bases de datos, facilitando la interpretación de la secuencia y la identificación de la proteína de interés. Se ha demostrado la posibilidad de realizar estudios de proteómica top-down y obtener fragmentación altamente informativa a partir de proteínas mayores de 200 kDa (Han et al., 2006).

La fuerza de este enfoque radica en la detección directa de la masa molecular nativa o experimental de las especies de proteínas biológicas. Se ha propuesto que los beneficios del enfoque “top-down” incluyen una mayor cobertura de secuencia de proteínas (Kelleher et al., 1999), una mejor caracterización de modificaciones postraduccionales (Post-translational modification, PTM) (Siuti y Kelleher, 2007; Borchers et al., 2006), permite la identificación de isoformas específicas de proteínas (Uttenweiler-Joseph et al., 2008) y de proteínas nativas (Chen et al., 2010) y una mayor confiabilidad de la cuantificación de proteínas (Du et al., 2006). Esta estrategia utiliza métodos de fraccionamiento alternativos al CID, como la disociación de captura de electrones (ECD) y la transferencia de electrones (ETD) (Zubarev et al., 1998)

Se requiere hacer uso de la disociación (ETD) (Coon et al., 2005) para lograr una secuenciación más completa del esqueleto de proteínas. Sin embargo, existen varias limitaciones tecnológicas que mantienen a los métodos de secuenciación descendente lejos de su uso generalizado. En este sentido, el alcance de este enfoque “top-down” se ha limitado al análisis de proteínas simples y mezclas de proteínas simples y se requieren varios esfuerzos para el análisis sistemático de mezclas de proteínas complejas (Yates et al., 2009).

II.12.2. El enfoque bottom-up.

La aproximación clásica de la proteómica Bottom-Up comienza típicamente con la separación electroforética de la mezcla compleja de proteínas en geles de poliacrilamida de una dimensión (One Dimensional gel Electrophoresis, 1-DE), o dos dimensiones (Two Dimensional gel Electrophoresis, 2-DE). Las proteínas resueltas en el gel son entonces visualizadas utilizando distintas tinciones convencionales y seguidamente, los geles son escaneados y sometidos a análisis de imagen mediante diferentes programas informáticos desarrollados para tal fin. La caracterización e identificación de las especies proteicas de interés, se lleva a cabo mediante la digestión en gel con enzimas, tales como la tripsina, y los péptidos así formados son analizados mediante MS. En la mayoría de los casos, pero no siempre, se recurre a la separación de la mezcla de péptidos, por medio de sistemas de cromatografía líquida (Liquid Chromatography, LC) acoplada a los equipos de MS. La identificación de la proteína original se lleva a cabo a partir de la comparación e interpretación adecuada del mapa de péptidos o huella peptídica (Peptide Mass Fingerprinting, PMF), o bien a partir del espectro de fragmentación de uno o más de sus péptidos, frente a los valores teóricos calculados a partir de los datos contenidos en las bases de datos (digestión *in silico*).

Esta estrategia “clásica” es la más utilizada hasta el momento, principalmente en estudios de proteómica de expresión y proteómica de expresión diferencial. Se trata de una aproximación basada normalmente en la separación mediante geles 2-DE, que ofrece una buena resolución y compatibilidad aunque adolece de reproducibilidad, es difícil de estandarizar y automatizar, y no permite el análisis de todos los componentes de un determinado proteoma, ya que presenta un rango dinámico limitado.

Una segunda aproximación de la estrategia bottom-up, está basada en la tecnología multidimensional de identificación de proteínas (Multi Dimensional Protein Identification Technology, MudPIT), propuesta por el grupo de Yates en 1999 (Link y col., 1999; Washburn y col., 2001). El término shotgun (del inglés, escopeta), se refiere al estudio de las proteínas contenidas en una muestra a partir del análisis directo y rápido de la mezcla compleja de péptidos resultantes de su digestión enzimática (Wu y MacCoss, 2002). En esta aproximación, se emplean digestiones peptídicas completas, obtenidas a partir de mezclas de

proteínas, sin necesidad de una separación previa mediante 2-DE u otro procedimiento cromatográfico. Posteriormente, la mezcla de digeridos peptídicos se somete a separaciones cromatográficas multidimensionales, acopladas en una última etapa, a un espectrómetro de masas como detector. Éste finalmente induce la fragmentación de los péptidos, conforme van eluyendo de la columna. Típicamente la separación se realiza en dos dimensiones, atendiendo a dos propiedades diferentes de los péptidos, primero según su carga, utilizando para ello una LC de intercambio catiónico fuerte (Strong Cation eXchange, SCX) y posteriormente según su hidrofobidad, empleando una LC en fase reversa (RP, Reverse Phase). Sin embargo, si la suma de péptidos a analizar no es muy compleja, el análisis shotgun puede ser realizado sirviéndose simplemente de una única separación cromatográfica, normalmente utilizando una LC en RP. Finalmente, se realiza la comparación bioinformática de las fragmentaciones obtenidas frente a las fragmentaciones in silico de las proteínas contenidas en las bases de datos, lo que da lugar a la identificación de los péptidos y consecuentemente de las proteínas presentes en la muestra de partida.

Las ventajas del enfoque “bottom-up” incluyen una mejor resolución en la separación a nivel de péptido que a nivel de proteína y una sensibilidad más alta que el método “top-down”. Los principales inconvenientes del enfoque de abajo hacia arriba incluyen una cobertura de secuencia de proteínas limitada por los péptidos identificados, una pérdida de PTM lábiles, la pérdida de conectividad entre péptidos y proteínas y la ambigüedad en la determinación del origen de las secuencias de péptidos degenerados (figura 7), es decir, péptidos compartidos por múltiples proteínas. Además, solo se identifica una fracción de la población peptídica total de una proteína dada. Por lo tanto, se obtiene información en solo una porción de la secuencia proteica.

Sin embargo, para aquellos proteomas cuyas secuencias genéticas o proteicas no estén previamente registradas en las bases de datos, como puede ser el caso de muchas especies del género Asteraceae, ambas estrategias aun siendo complementarias presentan una dificultad añadida, puesto que se trabaja con un recurso todavía no identificado ni descrito. Lógicamente el problema se complica aún más si se opta por el análisis de proteomas no caracterizados, mediante métodos multidimensionales, supeditados a la fragmentación masiva de péptidos. En todos estos casos, la estrategia a seguir necesita de la secuenciación “de novo” de los péptidos (bottom-up) o de proteínas intactas (top-down), con el objeto de

determinar su secuencia de aminoácidos completa, o al menos aquellos tags de secuencias susceptibles de dar información sobre el posible origen del péptido mediante análisis de homología.

Actualmente, existen distintos programas automáticos diseñados para secuenciar péptidos de novo, a partir de la interpretación directa de sus espectros de fragmentación, sin embargo todavía se está lejos de la interpretación de novo de forma automática a partir de los datos “top-down”.

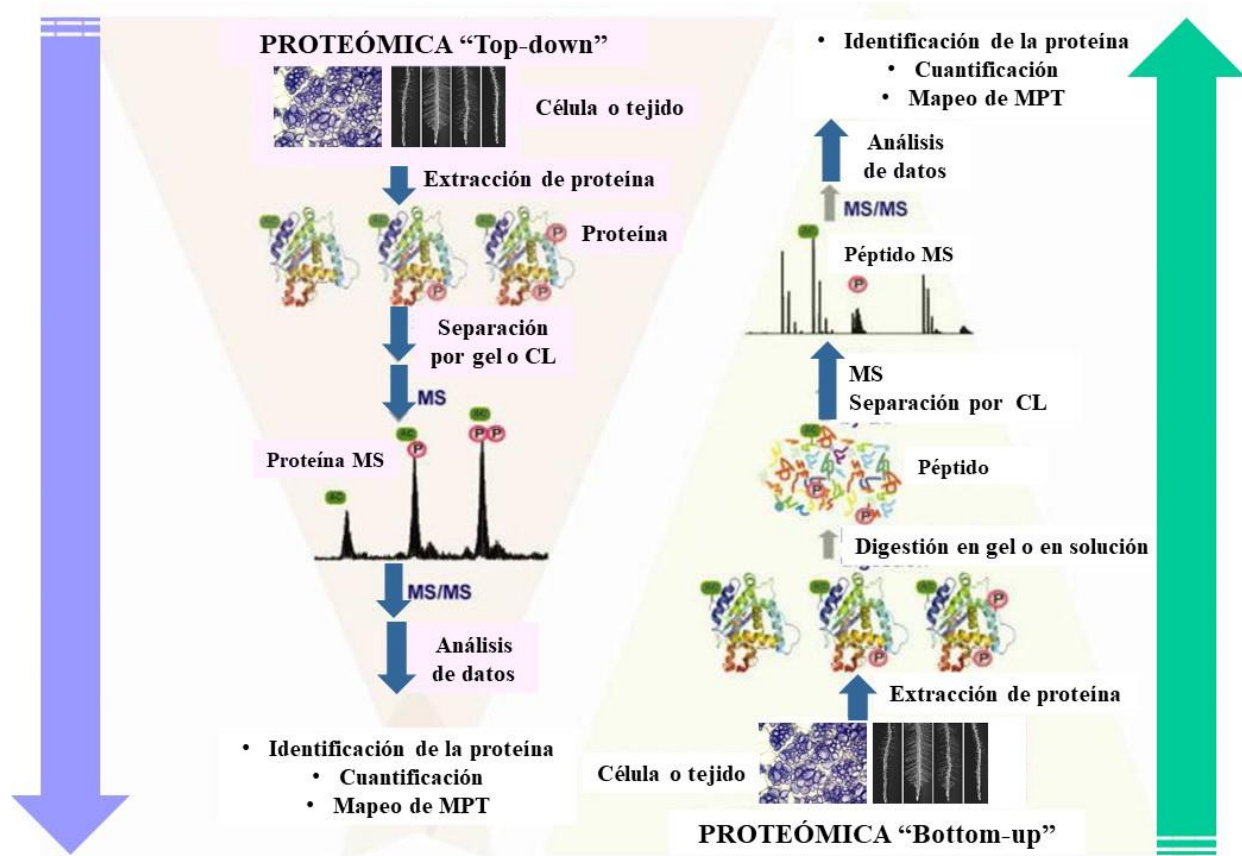


FIGURA 7. Esquema de las estrategias de “bottom-up” y “top-down”
 *Editado y tomado de: Gregorich et al., 2014.

II.13. Problemas con el manejo de muestras de proteínas.

II.13.1. Preparación de la muestra.

La preparación de la muestra es el paso más crítico en cualquier estudio proteómico y las plantas no son la excepción, pues comparado con otros organismos las plantas son un poco

más complejas en cuanto a estructura celular, ya que las células vegetales son usualmente muy ricas en proteasas y otros compuestos de interferencia como los lípidos, polisacáridos, polifenoles de la pared celular y varios metabolitos secundarios, que podrían no favorecer la extracción proteica deseada. Sin embargo, debido a la diversidad de la abundancia de las proteínas, del peso molecular, de la carga, de su hidrofobicidad, de sus modificaciones postraduccionales y de su interacción con otras moléculas, no existe un protocolo “universal” que pueda adaptarse a todas las muestras proteicas (Chen y Harmon 2006).

La extracción de proteínas vegetales generalmente involucra la disrupción física del tejido o célula en combinación con métodos mecánicos, lisis química o enzimática. Dos métodos químicos más usados en diferentes protocolos para la extracción de proteínas: el de precipitación con ácido tricloroacético/acetona (TCA/acetona) y la extracción basada en fenoles con precipitación con metanol o acetato de amonio (Malcevski y Marmiroli, 2011).

II.13.2. Separación de la muestra.

Para lograr la separación de la muestra biológica se emplean dos técnicas en los estudios proteómicos, una de ellas es la cromatografía líquida bidimensional (LC) y otra es la electroforesis bidimensional, la cual consiste en separar las moléculas dependiendo fundamentalmente de su carga y de su peso molecular bajo la influencia de un campo eléctrico. Se pueden conocer a través de la electroforesis las características ácido-básicas, peso molecular, punto isoelectrico etc (García, 2000).

II.13.2.1 Electroforesis bidimensional (2-DE).

La electroforesis bidimensional (2-DE) fue la primera técnica empleada en Proteómica para separar mezclas complejas de proteínas, que eran identificadas posteriormente en espectrómetros de masas tipo MALDI-TOF, en la actualidad sigue siendo una técnica de elección para estudios proteómicos. Está basada en la electroforesis convencional en geles de poliacrilamida o PAGE-SDS, en la que se separan las proteínas por su masa molecular en presencia del detergente SDS. A este tipo de electroforesis, que constituye la segunda dimensión, se le añade una electroforesis adicional previa en base al punto isoelectrico (Isoelectric point, pI) denominada isoelectroenfoco (Isoelectric focusing, IEF), que constituye la primera dimensión. La separación de las proteínas en la segunda dimensión se realiza de manera perpendicular a la primera, de tal manera que el resultado obtenido es un

mapa de manchas proteicas en dos dimensiones: punto isoeléctrico y masa molecular (Figura 8).

Desde que O'Farrel (1975) informó por primera vez, 2-DE ha mejorado en reproducibilidad y resolución con el desarrollo de gradientes de pH inmovilizados (IPG) (Rabilloud et al., 1997a; Rabilloud et al., 1997b), y por consiguiente en el número de manchas de proteínas distintas que se pueden detectar (Görg et al., 2000). Sin embargo, además de estas mejoras técnicas, aún se requiere un grado de competencia técnica y, para obtener resultados óptimos, la alta calidad de las muestras de proteínas es crítica, ya que los contaminantes pueden interferir dramáticamente con la separación y distorsionarla.

Las limitaciones técnicas de la 2-DE incluyen la capacidad limitada para separar clases específicas de proteínas, en particular hidrófobas y para visualizar proteínas de baja abundancia (Harry et al., 2000; Rabilloud, 2002). Por ejemplo, está bien establecido que las proteínas de membrana hidrófobas no se fraccionan bien utilizando 2-DE (Santoni et al., 1999; 2000). Estas raramente están presentes en la lista de identificación de proteínas de las muestras de proteínas separadas por 2-DE, que están dominadas por las proteínas solubles contenidas en los tejidos bajo análisis. Las técnicas alternativas de fraccionamiento sin gel basadas en la separación de las proteínas por cromatografía, se utilizan de manera efectiva para detectar muchas proteínas también en mezclas de proteínas complejas, siendo altamente reproducibles y permitiendo estudios comparativos. Además de los enfoques existentes, 2-DE sigue siendo el enfoque más efectivo para el análisis proteómico comparativo (Rabilloud, 2002). Sin embargo, como ya se ha mencionado la técnica presenta algunas limitaciones, éstas principalmente son la variación entre geles y muchas otras variables que complican sustancialmente la superposición de múltiples imágenes de gel y, por lo tanto, la comparación de manchas entre las muestras analizadas, incluido el problema de los puntos de migración conjunta en los que más de dos proteínas por lugar puede aparecer (Vincent et al., 2007; Gygi et al., 2000). La variación experimental en los patrones de puntos entre geles es una combinación de variación de posición y variación de abundancia, por ejemplo, la misma cantidad de la misma proteína puede correr en diferentes posiciones con diferentes grados de resolución espacial y, por lo tanto, tienen diferentes intensidades de puntos en diferentes geles. Se logró un avance importante en la mejora y la superación de este paso limitante con el desarrollo de la electroforesis fluorescente en gel en 2-D (Difference gel electrophoresis,

DIGE) que permite distinguir la variabilidad biológica de la técnica mediante la introducción de estándares internos (Ünlü et al., 1997).

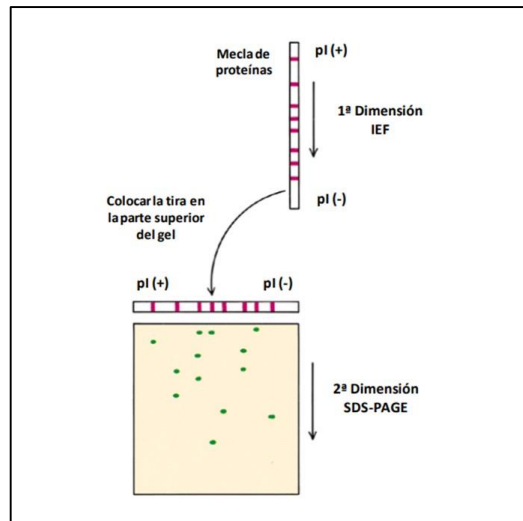


FIGURA 8. Representación esquemática de la técnica de electroforesis bidimensional. La técnica de 2-DE está basada en la separación de una mezcla de proteínas en dos dimensiones. En la primera dimensión las proteínas se separan según su punto isoeléctrico en una tira de poliacrilamida con un gradiente inmóvilizado de pH. A continuación, la tira se coloca en la parte superior de un gel SDS-PAGE, y se separan las proteínas en base a su masa molecular (segunda dimensión).

*Editado y tomado de Lodish et al., 2000.

II.13.2.2. Cromatografía líquida (Liquid chromatography, LC).

Se denomina cromatografía a toda técnica de separación que distribuye a los componentes de una mezcla entre dos fases: una estacionaria y otra móvil. Las moléculas que se van a separar se distribuyen entre ambas fases, bien por solubilidad, bien por su interacción con alguna de las fases (habitualmente la estacionaria), llevándose a cabo, de esta manera, la separación en una columna. Las moléculas con menor solubilidad/interacción con la fase estacionaria se moverán más rápidamente porque tenderán a permanecer en la fase móvil, mientras que las que tengan mayor solubilidad/interacción se moverán más despacio, tratando de permanecer asociadas a la fase estacionaria el mayor tiempo posible (Link *et al.*, 1999; Washburn *et al.*, 2001)

La técnica cromatográfica más comúnmente utilizada en proteómica es la cromatografía líquida (Liquid chromatography, LC), ya que puede ser directamente acoplada a espectrometría de masas (Mass spectrometry, MS) y puede realizarse tanto con proteínas

como con péptidos. En los métodos cromatográficos que se emplean en proteómica se utiliza como fase estacionaria una matriz porosa, generalmente en forma de columna. Existen diferentes tipos de LC según la propiedad física o química en base a la cual se lleve a cabo la separación de las proteínas.

a) Cromatografía de afinidad. Las proteínas se separan en base a su afinidad de unión específica a ligandos unidos a la matriz de la columna. Una columna de afinidad puede retener específicamente a una o varias proteínas o péptidos, lo que implica el uso de anticuerpos unidos a la matriz, o puede unir grupos específicos de proteínas y péptidos por su unión a determinadas moléculas como sucede, por ejemplo, en el caso de la cromatografía de afinidad mediante metales inmovilizados (Immobilized metal affinity chromatography, IMAC) en el que fosfoproteínas o fosfopéptidos se unen a los iones cargados positivamente de la fase estacionaria. Esta técnica conlleva un método de elución en dos pasos. En la primera fracción eluyen las proteínas o péptidos que no interaccionan con los ligandos unidos a la matriz y en la segunda aquellos que han sido retenidos por éstos.

b) Cromatografía de intercambio iónico. En este tipo de LC las proteínas o péptidos se separan por su carga. La matriz empleada será un soporte cromatográfico de tipo catiónico o aniónico, que atraerá a las moléculas con carga opuesta, en mayor medida cuanto mayor sea su carga. Para conseguir la elución progresiva de las proteínas según su carga, se suele aumentar la fuerza iónica o el pH de la fase móvil.

c) Cromatografía de fase reversa. El empleo de una fase enlazada hidrofóbica en la matriz de las columnas de fase reversa permite su empleo para llevar a cabo la separación de una mezcla de proteínas o péptidos en base a su hidrofobicidad. La elución gradual de las moléculas se lleva a cabo mediante un aumento progresivo de un componente orgánico en la fase móvil que produce la disociación de las interacciones hidrofóbicas. La cromatografía de fase reversa es la modalidad cromatográfica de mayor resolución y la más empleada en Proteómica, fundamentalmente debido a la compatibilidad de las fases móviles con la ionización por electrospray (Electrospray Ionization, ESI) en el espectrómetro de masas, lo que permite su acoplamiento en línea.

d) Cromatografía de penetrabilidad. También se conoce como cromatografía de exclusión por tamaño. El soporte cromatográfico que se emplea en este tipo de LC está

constituido por unas microesferas porosas formadas por un entramado tridimensional de fibras de polímeros naturales como la agarosa o los dextranos. Los espacios presentes entre estas fibras formarán los poros. Las proteínas o péptidos de menor tamaño pueden entrar por los poros de las microesferas, lo que les hace tardar más tiempo en llegar al final de la columna, al contrario de lo que les sucede a los de mayor tamaño, que al no acceder a través de los poros de las esferas se desplazan más rápidamente a través de los huecos existentes entre éstas. Este principio de separación, denominado exclusión molecular, provoca que las proteínas o péptidos eluyan en volúmenes inversamente proporcionales a su masa molecular.

e) Cromatografía líquida multidimensional. Para estudiar un proteoma por LC-MS suele ser conveniente utilizar varios métodos cromatográficos concatenados, lo que permite obtener la resolución necesaria para estudiar complejas mezclas de proteínas y péptidos. Es bastante frecuente el empleo de una técnica bidimensional de LC con una primera separación mediante intercambio iónico y una segunda en fase reversa. Esta metodología permite realizar una separación bidimensional, al igual que en el caso de una 2-DE, con la ventaja de que se puede acoplar directamente a MS, lo que permite realizar el estudio proteómico de una manera automatizada (Link *et al.*, 1999; Washburn *et al.*, 2001). La adición de pasos extra de LC en este método de cromatografía bidimensional permite realizar las dimensiones que sean requeridas para un adecuado fraccionamiento del proteoma de estudio, siempre teniendo en cuenta la necesidad de utilizar tampones de elución en las LC compatibles con el ciclo cromatográfico posterior. En una aproximación bastante reciente, se han desarrollado columnas con dos soportes cromatográficos en tándem en busca de una completa automatización del método de LC bidimensional referido anteriormente. Estas columnas contienen un soporte de intercambio catiónico en la zona proximal seguida de una de fase reversa en la distal. Esto permite realizar la LC bidimensional por cromatografía líquida de alta resolución (High performance liquid chromatography, HPLC) y acoplarla de manera automatizada a un espectrómetro de masas con fuente ESI y ha sido denominada como tecnología multidimensional de identificación de proteínas (Multidimensional protein identification technology, MudPIT).

II.14. Espectrometría de masas.

II.14.1. Identificación de proteínas basada en espectrometría de masas.

La espectrometría de masas es una técnica analítica que se emplea para determinar la masa molecular de un compuesto mediante el uso de instrumentos que pueden medir la relación masa/carga (m/z) de iones, trabajando en condiciones de vacío. Estos instrumentos se denominan espectrómetros de masas. En sus inicios, la MS no era aplicable a las proteínas, debido a la ausencia de técnicas suaves de ionización que permitieran realizar el paso de grandes biomoléculas a iones en fase gaseosa sin una fragmentación excesiva de las mismas, es decir, que no se volatilizaban. En su lugar, desde mediados de los 80 se empleaba una técnica para identificar proteínas procedentes de geles 2-DE transferidos a membranas de polivinildifluoruro (PVDF), basada en el empleo de la degradación de Edman. A finales de los 90 se desarrollaron dos técnicas de ionización que permitieron realizar análisis de proteínas y péptidos, entre otras biomoléculas, por MS. La primera de ellas es la denominada ionización mediante desorción por láser asistida por matriz (Matrix-assisted laser desorption ionization: MALDI) desarrollada por Karas y Hillenkamp en 1988. La segunda de ellas es la ionización por electrospray (Electrospray ionization, ESI), desarrollada por Fenn y colaboradores. La identificación mediante degradación de Edman es una técnica laboriosa y requiere largos periodos de tiempo, por lo que a mediados de los 90, con el desarrollo de la MS de proteínas, fue desplazada por ésta como método de identificación.

La espectrometría de masas (MS) es un enfoque instrumental que permite la medición masiva de iones derivados de moléculas. Durante la última década, los espectrómetros de masas se han convertido cada vez más en una herramienta esencial en la investigación proteómica, debido en parte a los nuevos métodos de ionización y las mejoras importantes en la precisión, resolución y sensibilidad de la masa. El diseño y la función de estos instrumentos juegan un papel importante en la determinación del tipo de experimento de proteómica que se puede realizar. A continuación, se proporciona una descripción de los diferentes tipos de analizadores de masa con más frecuencia utilizados en las investigaciones de estudios proteómicos (Domon et al., 2006).

Los espectrómetros de masas suelen constar de tres partes principales: la fuente de ionización, el analizador de masas y el detector (Figura 8). La fuente de iones se utiliza para convertir los péptidos en iones analitos que luego se transfieren al analizador de masas, donde

se separan de acuerdo con su relación masa-carga (m/z). Los métodos más utilizados en proteómica son la ionización por electrospray (ESI), desarrollada por Fenn y colaboradores (1989) y la desorción e ionización láser asistida por matriz (MALDI), desarrollada por Karas y Hillenkamp (1988). Los analizadores de masa son una parte integral de cada instrumento porque pueden acumular iones y separarlos en función de las relaciones de masa a carga (m/z) dentro de un campo eléctrico o magnético.

1. Fuente de iones. Las macromoléculas biológicas, como proteínas y péptidos, no son fácilmente volatilizadas. El desarrollo de las técnicas de ionización suave permitió que péptidos y proteínas ionizados y relativamente intactos pudieran ser introducidos, en fase gaseosa, en un sistema de vacío en los espectrómetros de masas para ser analizados. La ionización ESI y MALDI son las más comunes en proteómica, aunque existen también otros métodos un poco menos utilizados.

MALDI. El analito se mezcla con de un compuesto aromático o matriz en presencia de un disolvente orgánico y se deposita en una superficie metálica. Al evaporarse el compuesto orgánico la matriz co-cristaliza junto con el analito. La placa metálica se introduce en la cámara de ionización del espectrómetro de masas, que se encuentra en alto vacío, y a continuación se incide con un pulso de un láser ultravioleta (UV) sobre ella. Los cristales absorben la energía del láser, algo que no haría el analito por sí solo, y se produce un importante aumento de energía en una superficie pequeña, lo que conlleva una rápida sublimación de la mezcla, dejando al analito ionizado y en fase gaseosa. La acción de un campo eléctrico (20-25kV) produce el movimiento de los iones del analito hacia el analizador de masas (Fenn et al., 1989; Karas et al., 1988).

ESI: En este caso el analito disuelto en un tampón volátil, a pH ácido, es llevado desde la salida de la columna HPLC a un fino capilar metálico que está sometido a un alto voltaje. Esto produce la salida de la disolución en forma de spray de gotas cargadas que contienen varios analitos. Las microgotas son sometidas a la acción de una corriente de gas a alta temperatura, lo que elimina el disolvente restante y deja a los analitos ionizados y en fase gaseosa, lo que provoca que sean acelerados por el campo eléctrico hacia el analizador de masas.

La ionización por ESI permite acoplar métodos de electroforesis capilar y cromatografía líquida directamente al espectrómetro de masas, lo que es hoy en día muy utilizado para realizar la identificación a gran escala de complejas mezclas de proteínas. La técnica MALDI se emplea para la identificación de mezclas de muy bajo número de proteínas, lo que limita su uso a la identificación de manchas proteicas de geles 2-DE o fracciones de baja complejidad procedentes de métodos multidimensionales de LC. Aún así sigue siendo una técnica complementaria al ESI por su gran sensibilidad y tolerancia a los contaminantes de la muestra (Domon et al., 2006).

2. Analizador de masas. El analizador de masas es la parte del instrumento en la que los iones se separan en base su relación entre la masa y carga (m/z). Es el elemento que se usa generalmente para definir el tipo de instrumento. Existen varios tipos y además pueden combinarse en los llamados espectrómetros híbridos. Así, un analizador de tipo cuadrupolo puede encontrarse acoplado con un analizador de tiempo de vuelo o una trampa iónica para formar un QTOF (Quadrupole-Time Of Flight, Cuadrupolo-Tiempo de Vuelo) o QTrap (Quadrupole-Ion Trap, Cuadrupolo-Trampa Iónica) respectivamente. Algunos de los analizadores de masas más empleados en proteómica son los siguientes:

- Analizadores de tipo TOF (Time Of Flight, Tiempo de Vuelo). Estos analizadores usan un campo eléctrico para acelerar los iones de analito. La separación se produce por la diferencia en el tiempo que éstos invierten en recorrer una distancia en el vacío en el interior del analizador, el llamado Tiempo de Vuelo. La aceleración y por tanto el Tiempo de Vuelo es una función de la relación m/z de los iones que impactan en el detector a diferentes tiempos. Los analizadores tipo TOF separan todo el paquete de analitos ionizados en una escala dependiente del tiempo y por tanto se usan generalmente en combinación con ionización MALDI, que introduce iones en el analizador en pulsos de láser en un régimen discontinuo controlado.
- Cuadrupolos. Los analizadores de tipo cuadrupolo reciben su nombre porque constan de cuatro varillas metálicas de sección hiperbólica en su cara interna enfrentadas en pares llamados polos con cargas opuestas. Sobre estos pares, además del potencial eléctrico de corriente continua, se aplica también una corriente alterna de radiofrecuencia. Esta conformación permite crear un campo eléctrico oscilante

controlado que desvía selectivamente los iones que pasan através y de esta forma solo los iones en un estrecho margen de valores m/z podrán llegar a impactar en el detector mientras que el resto son desviados y filtrados.

- Las trampas iónicas funcionan bajo el mismo principio físico que los cuadrupolos, pero la conformación en forma de cámara de las trampas iónicas permite confinar y acumular los iones que luego son liberados selectivamente. La trampa iónica QIT (Quadrupole Ion Trap, Trampa Iónica Cuadrupolar), descrita por el Nobel Wolfgang Paul en 1960, consta de dos electrodos metálicos de sección hiperbólica enfrentados y un electrodo toroidal que conforman una cámara donde se acumulan los iones de analito. En el interior los iones orbitan en el vacío. El ajuste de la radiofrecuencia permite filtrar selectivamente los iones, estabilizando aquellos con determinados valores m/z y desestabilizando el resto, que colisionan con el electrodo y no llegan al detector (Louris et al., 1987). Las trampas iónicas LTQ (Linear Trap Quadrupole, Cuadrupolo de confinamiento lineal) o LIT (Linear Ion Trap, Trampa Iónica Lineal) consisten en un sistema de cuadrupolo, que sitúa los iones en un eje radial, y dos electrodos terminales, uno en cada extremo, que confina los iones longitudinalmente. Con respecto a las trampas iónicas tipo QIT, las trampas iónicas LTQ tienen una mayor capacidad de acumulación de iones y de barrido. Además, pueden funcionar como un cuadrupolo convencional.
- El Orbitrap es un tipo de trampa iónica, relativamente reciente, desarrollado a finales de los años 90 del siglo XX (Makarov, 2000). Consta de un electrodo en un eje interno rodeado por un electrodo externo cilíndrico. Los iones son introducidos tangencialmente desde la fuente de ionización y al ajustar la diferencia de potencial, son atrapados en órbitas elípticas longitudinales, en las que la atracción hacia el eje interno es compensada por la fuerza centrífuga. La relación m/z se determina a partir de la frecuencia angular de la oscilación de los iones en torno al electrodo longitudinal interno.

La identificación de las proteínas mediante MALDI-TOF se realiza mediante una radiación láser y con el uso de una matriz adecuada, se denomina MALDI por sus siglas en inglés Matrix-Assisted Laser Desorption/Ionization (desorción/ionización de matriz asistida

por láser) y TOF por el detector de iones que se acopla al MALDI y cuyo nombre procede también de sus siglas en inglés Time-Of-Flight (tiempo de vuelo).

Se opera mezclando una disolución de la muestra con una sustancia matriz, la mezcla se deposita en un dispositivo portamuestras especialmente diseñado para este sistema de ionización. Tras la evaporación del disolvente, los cristales de la mezcla muestra-matriz, se irradian con un haz láser. Los péptidos son identificados por la medición del "tiempo de vuelo" de los iones que se basa en la relación entre la masa y la velocidad de dichos iones. Este procedimiento permite la identificación rápida e inequívoca de la proteína, mediante "la huella digital" de la proteína obtenida mediante MALDI-TOF, sin embargo, al utilizar esta técnica es importante que se conozca la secuencia del genoma del organismo en estudio.

Otra de las técnicas relevantes en espectrometría de masas para la identificación de proteínas es la trampa de iones (IP) con fuente de ionización de electrospray (ESI). Conocida por sus siglas en inglés como ElectroSpray Ionization-Mass Spectrometry, ESI-MS es una técnica también llamada de "ionización suave", ya que hay muy poca fragmentación de la molécula en estudio. La técnica consiste en introducir el analito (que será la muestra de proteínas a ionizar) disuelto en un solvente más volátil por un capilar de metal muy pequeño y cargado. Debido a la repulsión de las cargas eléctricas, el líquido sale del capilar y forma un aerosol, una nube de pequeñas gotas altamente cargadas. Conforme el solvente se evapora, las moléculas de analito se aproximan, se repelen y finalmente, cuando la repulsión de las cargas positivas vence la tensión superficial, estallan las gotas (Explosión de Coulomb). El proceso se repite hasta que el analito está libre de solvente y finalmente los iones se mueven hacia el analizador de masas. El analizador de masas en este caso es el analizador tipo trampa de iones (IP); este analizador está formado por tres electrodos: el electrodo anular, el electrodo de entrada y el electrodo de salida, que forman una cavidad en la cual es posible almacenar y analizar iones (Domon et al., 2006).

Los iones formados en la fuente entran en el analizador donde se aplican diferentes voltajes generando un campo eléctrico tridimensional en la cavidad de la trampa. Este campo atrapa y concentra los iones dada su trayectoria de oscilación estable. La naturaleza de la trayectoria depende del potencial y de la relación masa-carga (m/z) de los iones. Durante la detección, los potenciales de los electrodos se alteran para provocar inestabilidad en las

trayectorias de los iones y expulsarlos en dirección axial en función de su relación m/z dando lugar a un espectro de masas (MS). Una de las características del cuadrupolo de la trampa iónica es su capacidad para aislar un ión y fragmentarlo obteniéndose así lo que se denomina “espectro de fragmentación” ó espectro MS/MS. Los espectros de fragmentación contienen la información acerca de la secuencia aminoácidica del péptido aislado. La ventaja de esta técnica es que no requiere del conocimiento previo de la secuencia del organismo en estudio, sin embargo, se requiere de mayor tiempo para completar el análisis de cada proteína y es más costosa.

Cada analizador de masas tiene propiedades únicas, como rango de masa, velocidad de análisis, resolución, sensibilidad, transmisión de iones y rango dinámico. Los analizadores de masas ION trap (IT), Orbitrap y Fourier Reson Cyclotron Resonance (FTICR) separan los iones en función de su frecuencia de resonancia m/z , los cuadrupolos (Q) y el tiempo de vuelo (TOF) utilizan la estabilidad m/z . Se han construido espectrómetros de masas híbridos que combinan más de un analizador de masas para responder a necesidades específicas durante el análisis (Gross, 2004; De Hoffman y Stroobant, 2007). Los iones de un valor m/z particular pasan desde el analizador de masas para golpear el detector. La magnitud de la corriente producida en el detector en función del tiempo se utiliza para determinar la intensidad o la abundancia de ese ion. La relación de la abundancia de iones generados en una fuente de ionización suave se mide en el detector y los resultados se muestran como valores m/z en abscisas y abundancias de iones en ordenadas.

Hay dos categorías amplias de analizadores de masas: los espectrómetros de masas en tándem en el espacio, que tienen analizadores de masas discretos para cada etapa de la espectrometría de masas y los espectrómetros de masas en tándem en el tiempo, que tienen solo un analizador de masas y cada etapa de la espectrometría de masas tiene lugar en la misma región, pero está separada en el tiempo por una secuencia de eventos (Bush et al., 1988). Las siguientes configuraciones son las soluciones más utilizadas en el campo de la proteómica: triple cuadrupolo (TQ o QqQ), TOF, TOF/TOF y los híbridos LTQ-Orbitrap, Q-TOF e IT-TOF para la categoría anterior; trampa de iones cuadrupolo (QIT), trampa de iones (3D), trampa de iones lineal (LIT) para la última categoría. Los analizadores de masa de barrido, como TOF, suelen estar conectados con MALDI para realizar análisis pulsados, mientras que los instrumentos de captura están acoplados a una fuente de ESI continua.

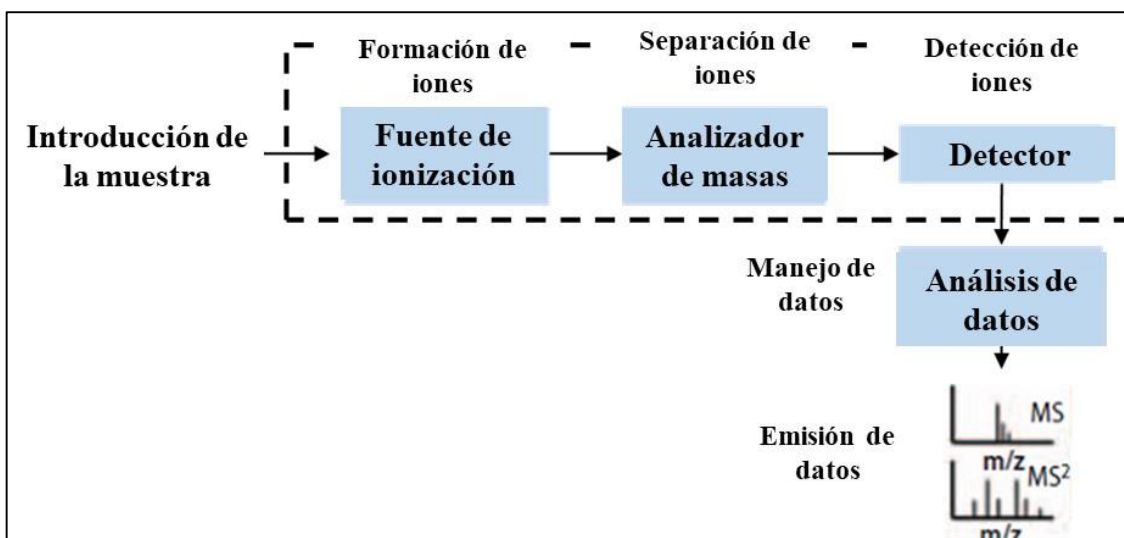


FIGURA 9. Esquemática de las partes principales que componen un espectrómetro de masas.
*Editado y tomado de Simpson, 2003b.

II.15. Estrategias utilizadas para la identificación de proteínas.

a) Huella peptídica (Peptide mass fingerprint, PMF). Consiste en la identificación de una proteína mediante el empleo de las masas de los péptidos obtenidos tras digestión enzimática, que generalmente se realiza con tripsina. Cada proteína puede ser inequívocamente identificada por las masas de los péptidos que la forman, y eso constituye lo que denominamos huella peptídica. Este tipo de identificación se realiza para proteínas únicas o mezclas simples de proteínas, que pueden provenir de manchas proteicas de un gel 2-DE y bandas de geles unidimensionales o fracciones de LC de baja complejidad. Una vez obtenido el espectro en un espectrómetro de masas tipo MALDI-TOF, se utiliza un programa informático denominado motor de búsqueda que compara el espectro obtenido con los espectros teóricos de una base de datos que contiene las digestiones teóricas de todas las proteínas.

b) Identificación de proteínas por espectrometría de masas en tándem (MS/MS). Actualmente se emplean en la mayoría de los casos, espectrómetros de masas con dos analizadores de masas en tándem para llevar a cabo identificación de proteínas. Estos equipos permiten realizar la medición de m/z de los péptidos intactos en el primer analizador de masas y tras una fragmentación de algunos de los péptidos en una cámara de colisión por CID, la

medición de las m/z de los fragmentos de los mismos en el segundo analizador de masas. Esto permite realizar la identificación de la proteína con información adicional a la realizada por huella peptídica, ya que el espectro de fragmentación de un péptido es característico de cada péptido. Este espectro de fragmentación puede incluso permitir la secuenciación total o parcial del péptido. La obtención de la secuencia de aminoácidos de un péptido, utilizando exclusivamente los datos presentes en el espectro de fragmentación se conoce como secuenciación de novo. Para realizar el análisis MS/MS de mezclas poco complejas se emplean los espectrómetros de masas tipo MALDI-TOF-TOF, que permiten realizar la identificación de proteínas con mayor confiabilidad que en un MALDI-TOF. El resto de espectrómetros de masas en tándem tienen fuentes de ionización ESI y diferentes tipos de analizadores de masas. Las configuraciones más habituales hasta la fecha son: Q3, Q-TOF, IT, Q-QLIT, FT y Orbitrap. Recientemente se han desarrollado nuevas técnicas de fragmentación como alternativa a la CID en busca de una mejor secuenciación peptídica. Existen dos nuevos métodos basados en la transferencia de electrones de los iones presentes en la cámara de colisión denominados disociación por captura de electrones (Electron capture dissociation, ECD) y disociación por transferencia de electrones (Electron transfer dissociation, ETD). En la figura 10 se muestran comparativamente ambas estrategias de MS para la identificación de proteínas.

. II.16. Técnicas de fragmentación.

La técnica más utilizada para realizar la fragmentación de las proteínas es la disociación inducida por colisión (CID), o alternativamente, la llamada (CAD). El proceso CID implica la activación colisional del ión por colisiones energéticas con un gas blanco neutro para convertir la energía de traslación en energía interna, colocando así el ión en un estado activado (excitado), seguido de una disociación unimolecular del ión activado para producir productos (McLuckey, 1992). Se pueden observar procesos de activación colisional de baja energía (Hunt et al., 1986) y de alta energía (Biemann, 1990), dependiendo de la instrumentación utilizada. Como se ilustra en la figura 11, los iones producto que surgen de la escisión de la cadena principal de α C-C, el enlace amida CN o el enlace N- α C se denominan iones de tipo a, b y c respectivamente, si la carga se mantiene en el fragmento amino-terminal, o iones de tipo x, y, z, respectivamente, si la carga se retiene en el fragmento

carboxi-terminal. Los iones de productos observados más comúnmente como resultado del CID de baja energía pertenecen a las series de iones de tipo b y y.

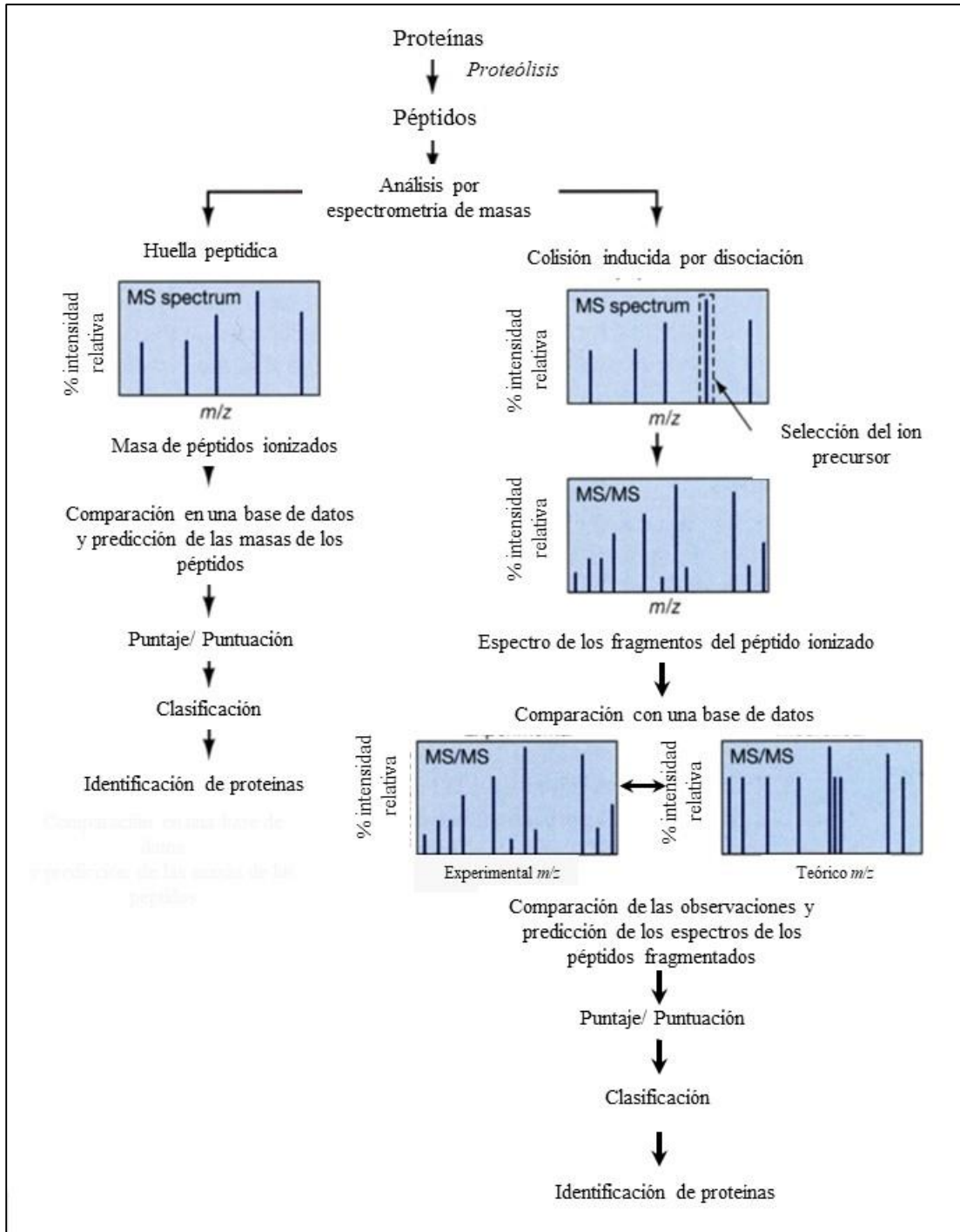


FIGURA 10. Comparación esquemática entre la huella de masa del péptido y la identificación de la proteína MS / MS.

*Editado y tomado de Simpson, 2003b.

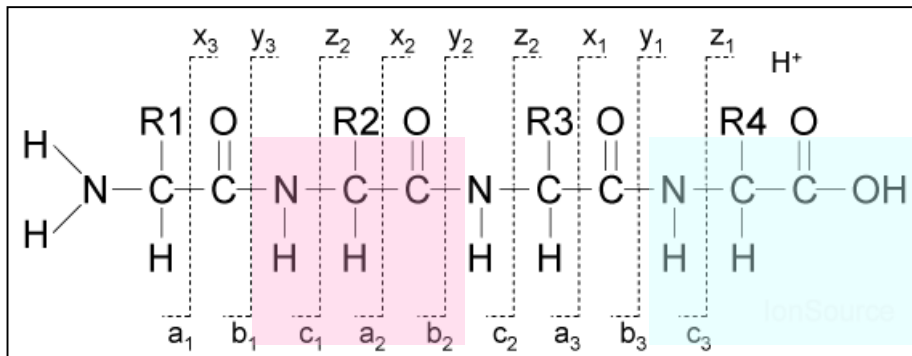


FIGURA 11. Fragmentación de la nomenclatura iónica de un tetrapéptido protonado. En la notación, los números indican el número de residuos de aminoácidos contenidos en el ión al que se hace referencia, la letra (a,b,c) o (x,y,z) indica el sitio de fragmentación del ión: las letras b e y corresponden a la ruptura en el enlace peptídico, la ruptura más habitual en activación CID. Por simplicidad se muestra un péptido de sólo cuatro aminoácidos.

*Tomado de Roepstorff y Fohlma, 1984.

Como alternativa a la fragmentación CID ampliamente utilizada, otros tipos de fragmentación emergentes también utilizados en los análisis LC-MS/MS incluyen: disociación de captura de electrones (ECD) (Zubarev et al., 1998), disociación de transferencia de electrones (ETD) (Coon et al., 2005; Syka et al., 2004) o disociación colisional de mayor energía (HCD) (Olsen et al., 2007).

ETD y ECD conducen a la generación de diferentes iones de fragmentos, denominados c y z, y han demostrado ser ventajosos para la identificación de PTM y polipéptidos caracterizados por alta masa y carga.

La disociación colisional de mayor energía (HCD) es una técnica de fragmentación adicional que se puede utilizar exclusivamente en el LTQ-Orbitrap. La fragmentación de HCD es similar a la fragmentación en instrumentos de triple cuadrupolo o cuadrupolo-TOF, y supera el problema del corte de baja masa de la fragmentación de la trampa de iones.

II.17. Motores de búsqueda para la identificación de proteínas.

Miles de espectros MS/MS se adquieren en un espectrómetro de masas, estos espectros MS se conocen como datos sin procesar, por lo que hay una serie de herramientas

computacionales disponibles para interpretar esos espectros y asignar una secuencia peptídica que permita la identificación de proteínas.

Hay varias herramientas de búsqueda de bases de datos de MS/MS disponibles actualmente, incluidas aplicaciones comerciales establecidas y ampliamente utilizadas como SEQUEST (Eng et al., 1994), MASCOT (Perkins et al., 1999) y programas integrados tales como SpectrumMill (Zhang et al., 2002), Phenyx (Allet et al., 2004), Proteome Discoverer (Thermo Scientific) o Protein Pilot (Applied Biosystems) y otras herramientas de código abierto como X! Tandem (Craig et al., 2004) o OMSSA (Geer et al., 2004).

Los datos sin procesar generados por el espectro de masas deben convertirse en una lista procesada simplificada de picos (archivos.pkl) adecuados para la búsqueda en la base de datos. La lista de picos incluye la masa del pico centrada, estado de la carga precursora calculada (z), el valor m/z y otras características espectrales (por ejemplo, la intensidad pico). Generalmente la identificación de proteínas es realizada correlacionando los espectros experimentales de MS/MS con los espectros teóricos predichos para cada péptido contenido en una base de datos de proteínas (Eng et al., 1994; Geer et al., 2004) de acuerdo con las reglas comunes de fragmentación de péptidos. La búsqueda está restringida a un subconjunto de espectros de acuerdo con diferentes parámetros de búsqueda seleccionados. El archivo de salida de las herramientas de búsqueda de la base de datos es una lista de coincidencias clasificadas de acuerdo con su puntaje asignado de las características específicas de cada algoritmo de búsqueda.

Un paso crítico para la identificación de proteínas es la selección de los parámetros de búsqueda tales como: el tipo de masas a seleccionar (ya sea monoisotópica o masa promedio), tolerancia de la masa del ión parental (para las tolerancias de la masa del producto y precursor que depende del espectrómetro de masas), tipo de digestión enzimática utilizada (generalmente tripsina), supuestas escisiones erróneas (generalmente 2), modificaciones químicas (modificaciones que dependen de la estrategia experimental, al hacer uso de agentes alquilantes como la iodacetamida) y modificaciones bioquímicas (modificaciones variables establecidas por los criterios y objetivos del usuario). Por lo general, la mayoría de los motores de búsqueda realizan una búsqueda inversa en la base de datos para disminuir el número de falsos positivos y esta tasa de descubrimiento de falsos positivos (False discovery

rate, FDR) puede alcanzar hasta el 1%. Otro parámetro importante seleccionado es la base de datos en la que se realizan las búsquedas teniendo en cuenta el organismo en estudio, la calidad, el grado de redundancia y el tamaño de la base de datos. Por lo general, se usan bases de datos de proteínas como SwissProt, NCBIInr entre otras, aunque las bases de datos genómicas se podrían usar como última opción debido a un mayor número de falsos positivos.

II.18. Cuantificación basada en gel.

II.18.1. Método 2-DE-semicuantitativo.

Como se describió anteriormente, 2-DE es una técnica poderosa para la separación de mezclas de proteínas complejas de acuerdo con el punto isoeléctrico (pI) en un primer paso de isoelectroenfoque (IEF) y masa molecular (Mw) resuelto en un segundo paso de electroforesis SDS-PAGE. 2-DE permite la comparación de las especies de proteínas de diferentes muestras a través de sus mapas de proteomas separados. El análisis de puntos diferenciales se basa en una comparación densitométrica de la abundancia de proteínas determinada como el volumen de puntos en una imagen digital de gel. Este análisis se realiza con la ayuda de sofisticadas aplicaciones de software disponibles comercialmente que permiten, de manera relativamente automática, conocer la abundancia relativa de cada punto en el mapa entre las muestras analizadas en el experimento. Después de la separación 2-DE, el flujo de trabajo básico para un análisis de expresión diferencial basado en la tinción única 2-DE implica, tinción de gel, adquisición de exploración de imagen, análisis de punto diferencial de imágenes digitalizadas y finalmente, identificación de proteínas de esos puntos de interés. La tinción de proteínas se realiza generalmente utilizando azul de coomassie coloidal (CBB) (Neuhoff et al., 1988) para su rango dinámico razonable o alternativamente, con tintes fluorescentes como SyproRuby (Molecular Probes) que presentan mayor sensibilidad que CBB y un rango dinámico lineal más amplio (Chevalier et al., 2004) pero necesita de un escáner de fluorescencia. Un método alternativo de tinción de proteínas es la tinción con plata (Shevchenko et al., 1996), que es muy sensible, por lo que es útil para detectar proteínas poco abundantes y también es compatible con espectrometría de masas, sin embargo, no es un método recomendable para estudios de análisis de expresión para su pobre rango dinámico y reproducibilidad (López et al., 2000). Después de la adquisición de la imagen, se realiza una cuantificación relativa mediante la comparación de las imágenes de

2-DE que siguen algunos pasos básicos, a saber, detección de puntos, coincidencia y normalización, comunes entre las marcas de software disponibles. La comparación se realiza contra una referencia entre todas las imágenes en el experimento que se conoce como la imagen de referencia. Una vez analizados, los puntos de interés se escinden del gel y se analizan por MS como se describió anteriormente.

Este método se considera semicuantitativo debido a la falta de una alta reproducibilidad técnica de los geles 2-DE que afecta a la superposición de las imágenes y debido a una cuantificación basada en un gel sintético promedio. Estas dificultades se superan con la introducción de la técnica de electroforesis en gel diferencial (DIGE).

II.18.2. Electroforesis en gel de expresión diferencial (DIGE).

La tecnología DIGE (descrita por primera vez por Ünlu et al., 1997) implica el marcaje covalente de muestras antes de la electroforesis con colorantes fluorescentes de resolución espectral, Cy2, Cy3 y Cy5. Las muestras se mezclan antes del IEF y se resuelven en el mismo gel 2-DE, minimizando así los artefactos técnicos. La principal diferencia con respecto al método de tinción única es que la variación técnica se minimiza mediante la introducción de un estándar interno, cuya imagen sirve como referencia experimental.

Esta es una técnica bastante costosa, no solo porque los colorantes fluorescentes son caros sino porque la etapa de identificación de espectrometría de masas punto por punto necesaria está separada de la cuantificación a nivel de gel. En muchos casos, las manchas individuales pueden consistir de más de una proteína, que puede reconocerse cuando se analiza por espectrometría de masas (Gygi et al., 2000). Esta circunstancia interfiere con la cuantificación y se han desarrollado algunos enfoques para abordar el problema (Ishihama et al., 2005). Una de las principales ventajas de la cuantificación basada en gel es la detección de isoformas y PTM. La cuantificación basada en gel 2-DE es una herramienta poderosa para cuantificar proteomas y a pesar de sus desventajas, todavía se usa ampliamente y se ha aplicado en el pasado a una variedad de preguntas en biología vegetal (Casati et al., 2005; Amme et al., 2006; Mooney et al., 2006; Hjerno et al., 2006; Hajduch et al., 2007; Keeler et al., 2007; Bohler et al., 2007).

II.19. Cuantificación basada en espectrometría de masas.

II.19.1. Técnicas sin etiqueta.

Los estudios proteómicos diferenciales que emplean la cuantificación sin etiquetas comparan dos o más muestras en función de las intensidades de iones de péptidos idénticos o en función del número de espectros adquiridos para cada proteína. Idealmente, las muestras para comparaciones sin etiquetas se ejecutan consecutivamente utilizando la misma configuración LC-MS/MS para evitar variaciones en las intensidades de iones debido a las diferencias en la configuración del sistema y por lo tanto, permitir una reproducción precisa de los tiempos de retención. Este es un análisis cuantitativo tipo escopeta que requiere un gran número de réplicas y tiene la limitación de una reproducibilidad pobre entre muestras, por lo que debe optimizarse a través de la etapa cromatográfica. Se puede aplicar a cualquier material biológico y la cobertura del proteoma respecto de las proteínas cuantificadas es alta porque básicamente se puede cuantificar cada proteína que se identifica por uno o más espectros peptídicos. Por lo tanto, los métodos sin etiquetas generalmente tienen una gran profundidad analítica y un rango dinámico, lo que brinda una ventaja a este método cuando se esperan grandes cambios globales de proteínas entre los tratamientos. Sin embargo, especialmente con los recuentos espectrales, la ventaja de una alta cobertura del proteoma puede tener el costo de una precisión relativamente pobre (Old et al., 2005).

Actualmente, los enfoques basados en MS sin marcaje (*'gel-free/label-free approaches'*) para la cuantificación de proteínas son bastante populares debido a la mayor resolución de los instrumentos utilizados para la identificación (como el espectrómetro de masas Orbitrap y otros equipos) y el desarrollo de nuevos algoritmos para la cuantificación, apoyado por un riguroso diseño de investigación y validación estadística (Porteus et al., 2011; Nahnsen et al., 2013; Holm et al., 2012). La cuantificación de proteínas por este método se basa generalmente en dos categorías: en las mediciones de los cambios de intensidad de iones, tales como áreas de los picos de péptidos o alturas de los picos en la cromatografía, y en base al conteo de los espectros generados de las proteínas identificadas en el análisis MS/MS (en inglés *'spectral counting'*). La cuantificación basada en MS sin marcaje presenta una menor precisión (incertidumbre > 30%) en comparación con los métodos basados en marcaje. Sin embargo, la cuantificación tiene un rango dinámico mayor y, en general, es una metodología mucho más sencilla y más barata (De Olivera et al., 2011).

La Proteómica libre de geles y libre de marcaje (*gel free, label free*) se basa en la cuantificación de la intensidad o del número de espectros de masas que están asignados a los péptidos empleados para identificar cada proteína particular. Se parte de la base que las proteínas más abundantes generarán también los péptidos más abundantes en la muestra. Un aumento del número de péptidos generará un aumento del spectral counting y también de su intensidad. La aproximación más sencilla para el procesamiento de estos datos consiste en el empleo del factor normalizado de abundancia espectral (NSAF, normalized spectral abundance factor, Paoletti et al. 2007) la cual es calculada de la siguiente forma:

$$(\text{NSAF})_k = \frac{(\text{PSM}/L)_k}{\sum_{i=1}^N (\text{PSM}/L)_i}$$

dónde el número total de espectros asignados a la proteína k (PSM) es dividido por la longitud de la proteína en aminoácidos (L), y posteriormente este valor es dividido por la suma de los PSM/L para todas las proteínas.

Esta aproximación es altamente dependiente de las bases de datos empleadas, ya que solo se cuantificarán aquellos iones que se hayan asignado a un péptido conocido que forme parte de una proteína (Valledor et al. 2012). El análisis ideal requeriría de la disposición del genoma del individuo a analizar, así como de todos sus modelos génicos. En caso de no estar disponible una base suficientemente amplia del genoma, o en caso de buscar una cuantificación más precisa y pese a que existen formas más complejas y sofisticadas de solventar este problema, una aproximación sencilla consiste en analizar la intensidad de los distintos iones presentes en la muestra.

Una vez estos iones estén cuantificados, podrán asignarse posteriormente a una proteína específica. Protmax es un software sencillo que permite alinear los espectros correspondientes a múltiples muestras y realizar tanto contabilizar los espectros como determinar su intensidad.

Esta aproximación se aleja del objeto de este capítulo, pero desde aquí podemos recomendar su descarga y la realización de los distintos tutoriales si se desea explorar esta alternativa de cuantificación (Egelhofer et al. 2013).

II.19.2. Etiquetado de isótopos estables.

Los enfoques de marcaje con isótopos estables se basan en el supuesto de que un péptido marcado con isótopos estables difiere del péptido no marcado solo en términos de su masa, pero muestra las mismas propiedades químicas durante la cromatografía. La etiqueta se puede introducir en varios pasos durante la preparación de la muestra. En el etiquetado metabólico, la etiqueta se introduce en toda la célula u organismo a través del medio de crecimiento; En el etiquetado químico se agrega a las proteínas o péptidos trópticos a través de una reacción química. La limitación de la cuantificación de proteínas en muestras complejas por métodos basados en isótopos estables radica principalmente en las interferencias de señal causadas por componentes de coelución de masa similar. Por lo tanto, la forma más eficiente de optimizar el análisis cuantitativo es disminuir la complejidad de la muestra aumentando los tiempos del gradiente cromatográfico o mediante el fraccionamiento cromatográfico multidimensional antes del análisis MS/MS.

Etiquetado metabólico: el marcaje metabólico explota la incorporación *in vivo* de isótopos estables durante la biosíntesis de proteínas y generalmente se logra mediante el crecimiento de células u organismos completos en presencia de sales de amonio enriquecidas con N^{15} o aminoácidos codificados isotópicamente, H^2 -leucina, C^{13} -lisina y C^{13}/N^{15} -arginina (Ong et al., 2002, 2003). El etiquetado con N^{15} (Wang et al., 2002) fue el primer tipo de marcaje metabólico en proteómica. La interpretación de la cuantificación es bastante compleja (Bachi y Bonaldi, 2008) y la dificultad añadida de las herramientas de software específicas hace que el etiquetado metabólico tenga varios inconvenientes para su uso general. El análisis cuantitativo de proteínas de alto rendimiento basado en el marcado metabólico se estableció utilizando aminoácidos esenciales marcados con isótopos estables en cultivos de células de mamíferos (SILAC del Inglés: Stable Isotope Labeling by Aminoacids in Cell culture) (Ong et al., 2002). SILAC se utiliza principalmente en cultivos de células de mamíferos debido a su incapacidad para sintetizar todos los aminoácidos, por lo tanto, la incorporación del etiquetado en las proteínas es completa. En contraste, como las plantas son capaces de sintetizar todos los aminoácidos, el etiquetado es incompleto, lo que dificulta la interpretación del resultado.

SILAC se basa en el marcaje durante el cultivo de una línea celular crecida en dos condiciones diferentes donde en una de las condiciones uno o más aminoácidos del medio de cultivo son sustituidos por versiones marcadas con isótopos pesados estables (por ejemplo ^{13}C en lisina o arginina). El uso de distintos isótopos para el marcaje hará que dos péptidos con la misma secuencia proporcionen dos valores diferentes de masa molecular y por tanto una relación m/z . Esta diferencia varía según el marcaje utilizado y recibe el nombre de desplazamiento de masa o *mass shift* y permite alinear y comparar los péptidos de la misma proteína obtenida en dos condiciones diferentes. Estos péptidos, no obstante, serán indistinguibles de sus equivalentes ligeros en el resto de sus propiedades físico-químicas incluyendo su comportamiento cromatográfico. La intensidad relativa de las señales ligera y pesada de un mismo péptido permite estimar la diferencia de concentración de dicho péptido y, por extensión, de su proteína de origen en las dos condiciones experimentales.

Una vez analizada la mezcla de péptidos por espectrometría de masas, los elementos derivados de la señal utilizados para calcular la abundancia de los péptidos identificados en las diferentes muestras, denominados *features*, son los picos MS^1 o relación m/z de los iones precursores de los péptidos. Para poder calcular la variación relativa de intensidad entre las versiones ligera y pesada del mismo péptido, basta localizar un par de picos MS^1 asociados al mismo péptido y comprobar que experimentan una diferencia de relación m/z que consistente con el desplazamiento de masa introducido. La diferencia entre las intensidades de ambas señales MS^1 proporcionará la variación relativa entre ambos péptidos. Finalmente y una vez localizados e identificados las versiones ligera y/o pesada de todos los péptidos que pertenecen a una determinada proteína, se calculará el valor promedio de la variación relativa de abundancia de la proteína en cuestión.

En los espectros de masas, cada péptido del digerido de la muestra se detecta en forma de un multiplete de señales separadas entre sí en función del número de isótopos pesados que cada péptido tiene incorporados. Las intensidades relativas de las señales en estos multipectos determinan las cantidades relativas del péptido en cada una de las muestras.

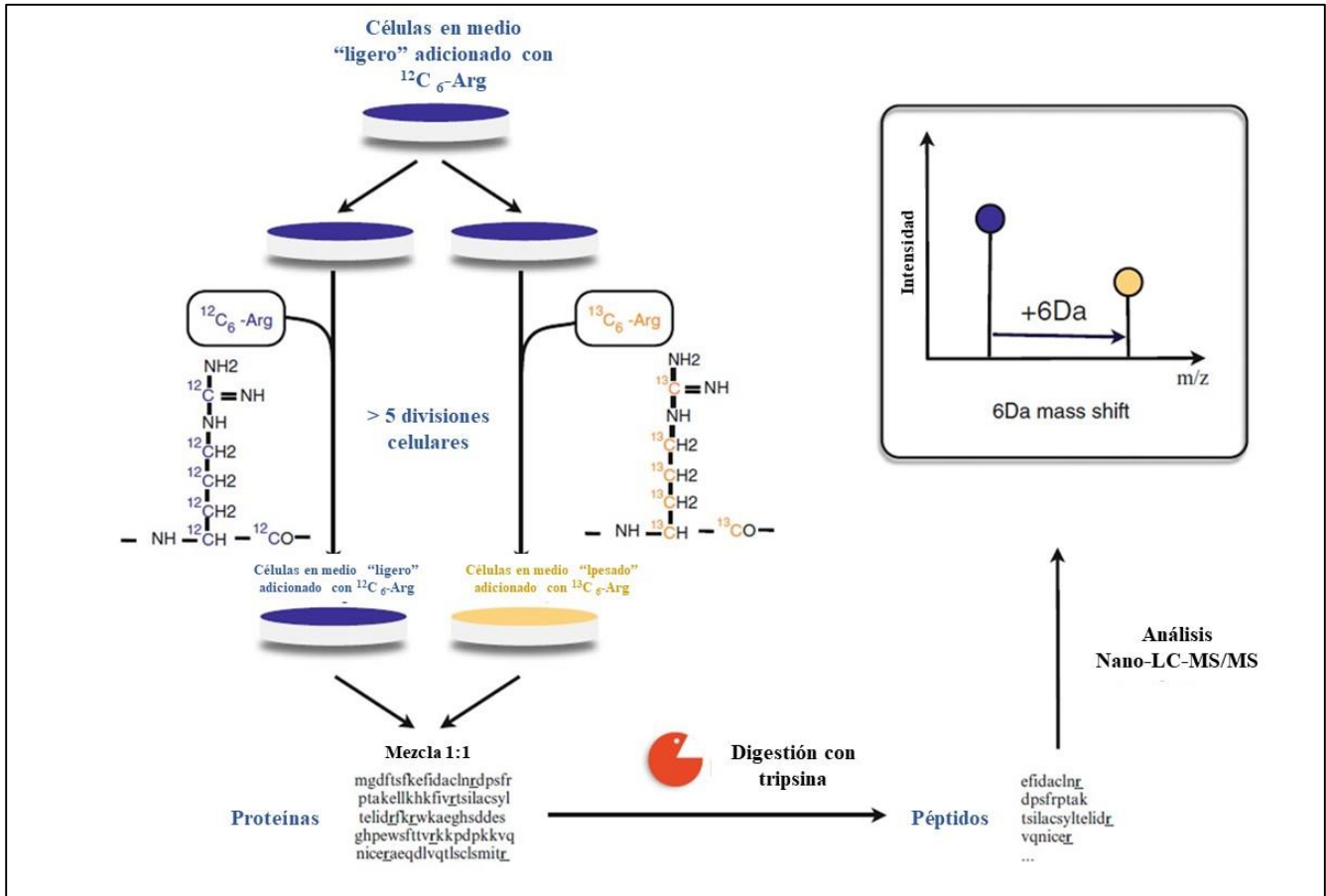


FIGURA 12. Cuantificación de proteínas con SILAC. Las células están marcadas diferencialmente en un medio con arginina "normal" o "mediana" con arginina "pesada". Después de al menos cinco divisiones celulares para asegurar virtualmente marcado completo, las dos poblaciones de células se combinan, se digieren con tripsina y se analizan por nano LC-MS / MS. La escisión triptica crea pares de péptidos que difieren en 6 Da debido a la diferencia de peso molecular entre $^{12}\text{C}_6$ -arginina y $^{13}\text{C}_6$ -arginina.
*Editado y tomado de Hoedt et al., 2019.

Etiquetado químico: el etiquetado químico funciona esencialmente de la misma manera que se describe para el marcaje metabólico con ^{15}N , excepto que el marcador se introduce en proteínas o péptidos aislados mediante una reacción química, por ejemplo, con grupos sulfhidrilo o grupos amino, o mediante acetilación o esterificación de residuos de aminoácidos (Ong y Mann, 2005). El marcado de isótopos también se puede introducir en la cadena peptídica durante la reacción enzimática de la digestión triptica mediante la adición de H_2^{18}O , lo cual introduce ^{18}O a los sitios de escisión del péptido (Yao et al., 2001; Sakai et al., 2005). La cuantificación se basa en exploraciones completas y en contraste con el

etiquetado ^{15}N , el cambio de masa entre los péptidos pesados y ligeros es constante. Una etiqueta de isótopos químicos comúnmente utilizada es la etiqueta de afinidad codificada por isótopos (ICAT, del inglés: Isotope Coded Affinity Tag), que se une a grupos sulfhidrilo de residuos de cisteína (Gygi et al., 1999b). Es una herramienta útil para estudiar el estado de oxidación o reducción de las proteínas, pero como la cisteína no es un aminoácido muy abundante, el número de péptidos que se pueden etiquetar y cuantificar por medio de marcadores de sulfhidrilo es bastante bajo.

Etiquetas de masas isobáricas: el etiquetado de masas isobáricas (TMT del inglés: Tandem Mass Tags) de Thompson et al., (2003) difiere de la estrategia descrita anteriormente en que la adición de etiquetas produce inicialmente péptidos marcados con etiquetas de la misma masa total que se eluye en la cromatografía de líquidos. Solo en la fragmentación de péptidos se pueden distinguir las diferentes masas etiquetadas. A medida que cada etiqueta agrega la misma masa total a un péptido dado, cada especie peptídica produce un solo pico durante la cromatografía líquida, incluso cuando se mezclan dos o más muestras. Por lo tanto, también habrá solo un pico en la exploración de MS de la muestra y solo se aislará una m/z para la fragmentación. Los fragmentos están en el rango de masa baja, que generalmente no es cubierto por iones de fragmentos peptídicos típicos. La relación de intensidad de los diferentes iones indicadores se utiliza para la cuantificación. Por lo tanto, la cuantificación para el etiquetado de masas isobáricas se basa en los espectros de fragmentos en lugar de en las exploraciones de la muestra. iTRAQ (Isobaric Tags for Relative and Absolute Quantification desarrollada por Ross et al., (2004) y TMT (Thompson et al., 2003) son etiquetas de masa isobárica disponibles comercialmente que se introducen en las proteínas de interés a nivel de péptidos trípticos.

II.20. Análisis funcional.

Un estudio de proteómica a gran escala es un gran avance en la biología experimental. Trae una gran cantidad de proteínas que presentan un desafío importante para los biólogos para explorar sus funciones biológicas. El método tradicional de "literature mining" a menudo consume mucho tiempo y es poco eficiente, por lo que se deben hacer grandes esfuerzos para desarrollar estrategias bioinformáticas más efectivas (Gehlenborg et al., 2010).

En los últimos años, se ha construido un vocabulario jerárquico, controlado dinámicamente Gene Ontology (GO) para describir la función molecular conocida, la ubicación subcelular y el papel biológico de las proteínas, y obtuvo un éxito significativo en la investigación biológica, especialmente en la investigación experimental a gran escala (Ashburner et al., 2000). Actualmente, GO se utiliza ampliamente para anotar la función de varios organismos y convertirse en el estándar para la anotación funcional.

La agrupación de datos ayuda a analizar las tendencias globales, hay varios algoritmos de agrupación disponibles, sin embargo, la agrupación jerárquica o k-medias se utiliza a menudo en muchas herramientas de software disponibles (Sturn et al., 2002).

III. JUSTIFICACIÓN.

Acourtia cordata es una de las especies del género *Acourtia* en la cual se han detectado altas concentraciones del sesquiterpeno conocido como perezona. Este compuesto y algunos de sus derivados han demostrado poseer diversas actividades de importancia farmacológica, tales como: la actividad citotóxica, hipoglucemiante, vasorelajante, antioxidante, entre otras. Además de presentar algunas otras actividades como la de compuesto con actividad antibacteriana, antialimentaria, asimismo se ha propuesto a la perezona como inhibidor de la corrosión de conductos para petróleo.

También se ha demostrado que este compuesto puede ser utilizado como sustancia base para la obtención tanto de nuevos pigmentos con la ventaja de ser biodegradables y también como molécula base para la obtención de otros compuestos derivados sintéticos con potenciales similares o distintos a los que presenta la perezona.

Por otra parte, se ha observado que las poblaciones naturales de *A. cordata* son relativamente pequeñas y que utilizar la planta silvestre para la obtención de perezona no sería viable, ya que la perezona se acumula en las raíces de la planta y su extracción del medio natural implica su destrucción. Por ese motivo se han realizado estudios para la producción de perezona a través de estrategias biotecnológicas como es el cultivo de tejidos vegetales y la obtención de raíces transformadas, sin embargo los rendimientos obtenidos mediante estas metodologías son muy bajos en comparación con los rendimientos obtenidos de las plantas silvestres. Se desconoce si la producción de este compuesto en plantas silvestres está relacionado con otros factores como son los factores ambientales, la variación en la concentración de nutrientes en el suelo, estrés inducido por herbívoros, hongos, bacterias o algún otro microorganismo.

A pesar de las aportaciones anteriormente mencionadas sobre la importancia de la especie y de sus productos naturales biosintetizados e identificados, hasta el momento no se han realizado estudios para el género, ni para la especie en relación a la identificación, localización y caracterización de las proteínas involucradas en la biosíntesis de sesquiterpenos. Ni tampoco existen reportes sobre el efecto o relación que tienen los factores edáficos en la producción de compuestos como los sesquiterpenos en *A. cordata*.

Dado que mediante la utilización de herramientas como la proteómica aplicadas al estudio de tejidos y organismos vegetales, se ha podido obtener gran información no solo de la cantidad de proteínas presentes en una condición dada, sino también de sus diversas variaciones de expresión y síntesis en los tejidos vegetales, se propone realizar un estudio proteómico en raíces de *A. cordata* que permita la comparación del proteoma en raíces con distintas concentraciones de perezona y además la identificación de algunas de las proteínas presentes en estas condiciones. Por otro lado se plantea también identificar los nutrientes en el suelo que contribuyan de alguna manera a la producción de este compuesto y algunos otros factores como son los de talla.

Este estudio permitirá tener un panorama amplio sobre las proteínas que son expresadas diferencialmente en raíces de *A. cordata* con distintas concentraciones de perezona. Además, permitirá también sentar las bases para lograr la identificación de genes involucrados con la expresión de enzimas que participan en la biosíntesis de perezona para posteriormente plantear la sobreexpresión de algunos de estos genes en el intento de incrementar la producción de este sesquiterpeno en cultivos de raíces *in vitro*.

IV. HIPÓTESIS.

El estudio proteómico comparativo de las raíces de poblaciones silvestres de *Acourtia cordata* permitirá establecer diferencias y similitudes entre ellas, así como identificar algunas proteínas relacionadas con la biosíntesis de terpenos y su relación con factores edáficos y características fenotípicas en la acumulación y rendimiento del sesquiterpeno perezona.

V. OBJETIVOS.

V.1 OBJETIVO GENERAL.

Realizar un estudio proteómico comparativo de las raíces de *Acourtia cordata* y analizar el efecto de algunos componentes edáficos y características fenotípicas de las plantas, sobre la producción de perezona.

V.2 OBJETIVOS PARTICULARES.

1. Obtener los rendimientos de perezona y agrupar en tres categorías (alta, media y baja) las raíces de *A. cordata* de acuerdo a su producción de perezona.

2. Obtener los patrones de expresión proteica de las raíces productoras de perezona de acuerdo a su categoría.

3. Identificar y comparar el proteoma de cada grupo de raíces productoras de perezona.

4. Analizar mediante herramientas bioinformáticas las proteínas sobreexpresadas y subexpresadas.

5. Analizar y establecer la probable influencia de factores edáficos y talla de las plantas en la producción de perezona.

CAPÍTULO 1. CUANTIFICACIÓN DE PEREZONA Y ANÁLISIS PROTEÓMICO.

1.1 INTRODUCCIÓN.

En el primer capítulo de esta tesis se aborda el estudio proteómico de las raíces silvestres de *A. cordata* con distintos niveles de producción de perezona, para lo cual se analizó el rendimiento de la perezona en 60 plantas colectadas dentro del área del corredor biológico e "Chichinautzin" en el Estado de Morelos, se agruparon los individuos de acuerdo a su rendimiento (alto, medio y bajo) para posteriormente realizar el análisis comparativo del proteoma apoyándonos de las herramientas proporcionadas por la proteómica. Puesto que la proteómica está considerada como una disciplina que representa una poderosa herramienta para describir distintos procesos dentro del reino vegetal tales como, cambios a nivel morfológico, fisiológico y bioquímico durante el desarrollo de los órganos y tejidos (Takáč et al., 2011; Di Michele et al., 2006; Tian et al., 2009; Mohamad et al., 2011), respuesta a distintos estímulos ambientales y diferentes tipos de estrés biótico y abiótico (Chmielewska et al., 2016; Trupiano et al., 2012; Aghaei and Komatsu, 2013; Ghosh and Xu, 2014; Fang et al., 2015; Sergeant and Renaut, 2010), así como el estudio de las proteínas de rutas biosintéticas que conducen a la producción de metabolitos secundarios de interés farmacológico (Sud et al., 2014; Senthil et al., 2011; Bhattacharyya et al., 2012; Ma et al., 2013; Champagne et al., 2012; Bryant et al., 2015) entre otros procesos.

El análisis de las proteínas se llevó a cabo a través de dos métodos: el método 1 en el cual las muestras proteicas fueron separadas por electroforesis bidimensional (2DE) y después analizadas por MALDI-TOF por sus siglas en inglés Matrix-Assisted Laser Desorption/Ionization (desorción/ionización de matriz asistida por láser) y TOF por el detector de iones que se acopla al MALDI y cuyo nombre procede también de sus siglas en inglés Time-Of-Flight (tiempo de vuelo). El método 2 en el cual las proteínas fueron separadas y analizadas por cromatografía líquida de ultra alto rendimiento acoplada al cuadrupolo de tiempo de vuelo conocida por sus siglas en inglés como UHPLC-QTOF (Ultra-High-Performance Liquid Chromatography coupled to Quadrupole Time-Of-Flight).

1.2. MATERIAL Y MÉTODOS.

1.2.1 Material vegetal.

Para realizar el estudio proteómico, se recolectaron raíces de 60 plantas silvestres de *A. cordata* en tres lugares dentro del área ecológicamente protegida el corredor del “Chichinautzin” en el Estado de Morelos a 2309 m a.s.l. y 18 ° 45 ' 0 " N, 99 ° 4 ' 0 " W. Se depositaron ejemplares con los números 34664, 34665 y 34666 en el Herbario HUMO del Centro de Investigación de Biodiversidad y Conservación de la Universidad Autónoma del Estado de Morelos, Mexico (figura 13). Cada muestra de raíces se guardó inmediatamente en bolsas de plástico y se almacenó en hielo para transportarlas al laboratorio. Una vez en el laboratorio, cada muestra se dividió y en dos partes: una parte se lavó cuidadosamente y se almacenó a -80 ° C hasta que se procesaron para la extracción de proteínas y la otra parte se usó para la cuantificación de perezona (figura 14).

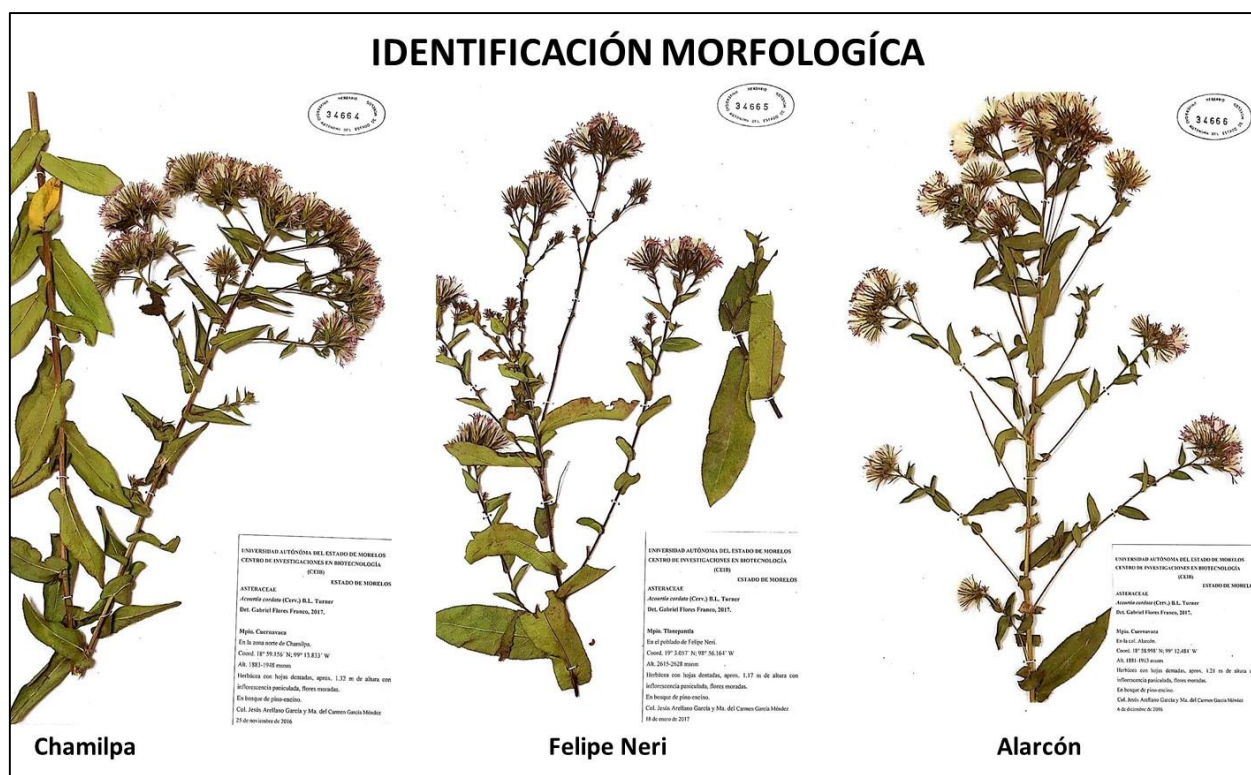


FIGURA 13. Ejemplares de *A. cordata* depositados en HUMO-Herbarium del Centro de Investigación de Biodiversidad y Conservación.

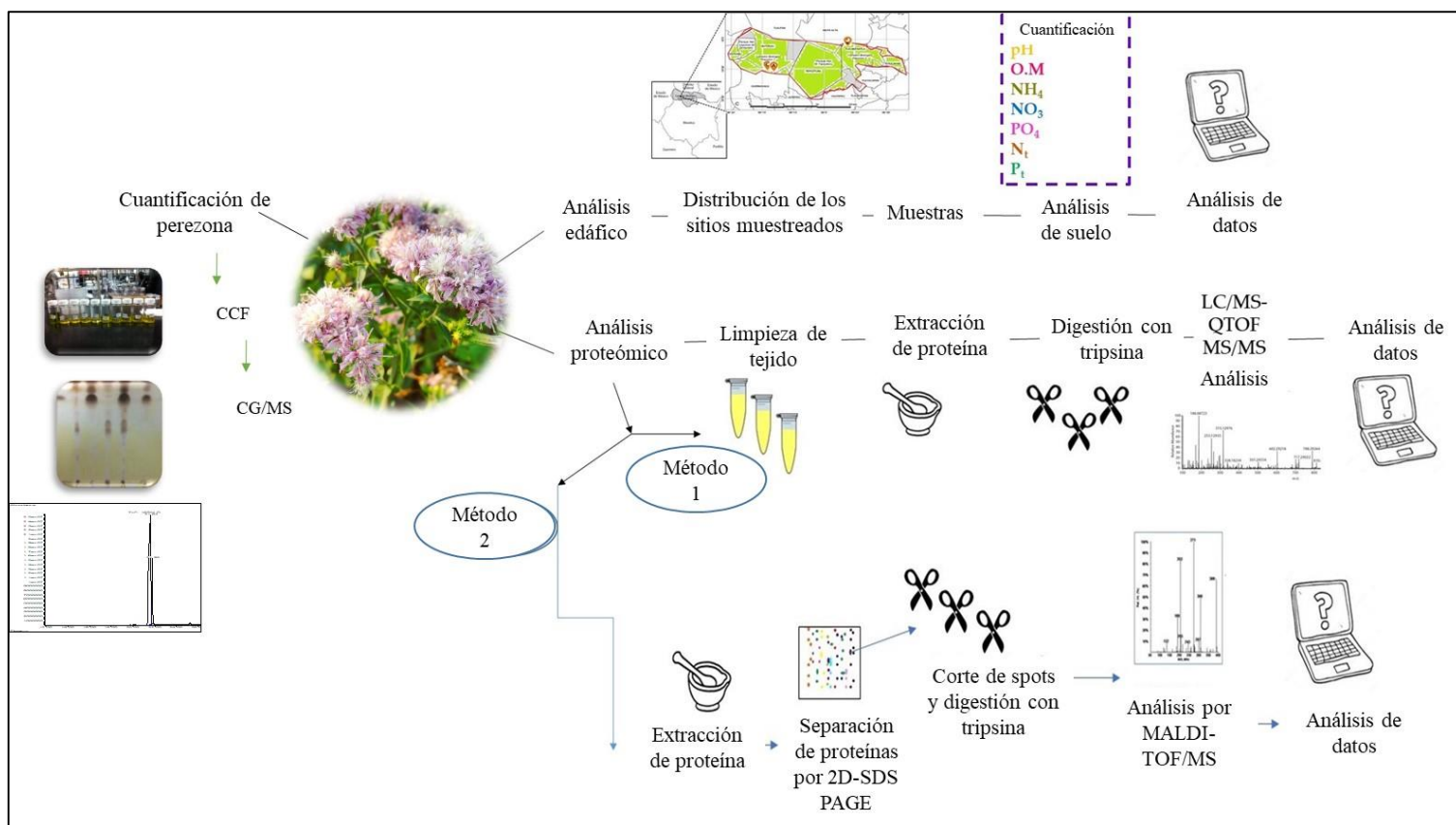


FIGURA 14. Flujo de trabajo que se realizó para el análisis de los extractos proteicos de las raíces de *A. cordata*.

1.2.2 Cuantificación de perezona.

1.2.2.1 Cromatografía de gases acoplado a espectrometría de masas.

La perezona se extrajo de las raíces de las plantas de *A. cordata*, para lo cual se utilizaron 2 g de raíz para la extracción con n-hexano por triplicado. Posteriormente, los extractos obtenidos se disolvieron en metanol para analizarlos mediante cromatografía de gases acoplada a espectrometría de masas. Para confirmar la presencia, así como para la cuantificación de la perezona, se usó un estándar de perezona como referencia y se empleó un cromatógrafo de gases Agilent Technologies 6890 acoplado a un detector de masas tipo cuadrupolo MSD 5973 (HP Agilent). Los compuestos se separaron en una columna capilar HP-SMS de 30 mm de largo, con un diámetro interno de 0.25 mm y una película de 0.25 μm . El gas portador fue helio, administrado a una velocidad de inyección lineal de 1 ml/min a flujo constante. La temperatura de entrada del inyector fue de 250 °C y la temperatura de la

columna, inicialmente a 40 °C, aumentó gradualmente a una velocidad de 10 °C/min hasta alcanzar 250 °C y manteniéndola durante 20 minutos. (Sánchez-Ramos et al., 2018).

1.2.3 Análisis proteómico.

1.2.3.1 Limpieza del tejido de la raíz.

Antes de la extracción de proteínas, cada muestra de raíz se limpio de acuerdo con el protocolo utilizado por Wang y colaboradores (2003) con algunas modificaciones (Martínez-Esteso et al. 2011). Para la extracción de proteínas, se trituraron 4.0 g de tejido en un mortero y nitrógeno líquido. Una vez triturado, el tejido se transfirió a tubos de centrífuga de 50 ml, se suspendió en 10 ml de hexano puro, se agitó en un agitador vortex durante 30 segundos y se centrifugó a 10, 000 x g por 5 minutos a 4 °C. Luego, el sobrenadante se desechó y el sedimento se limpió nuevamente 3 a 4 veces hasta que el sobrenadante fue incoloro. Después, el sedimento se resuspendió en 10 ml de una mezcla de acetato de etilo:etanol (1: 2) durante 1 minuto, se centrifugó 5 minutos a 4 °C y luego se descartó el sobrenadante y se resuspendió adicionando de 10 ml de acetona pura. Este procedimiento de resuspensión, centrifugación y descarte del sobrenadante, se repitió durante los siguientes pasos cambiando solo la solución de resuspensión a ácido tricloroacético (TCA) al 10% en acetona fría y finalmente se cambió a TCA al 10% en agua fría y acetona al 80%. En cada uno de estos pasos, la muestra se centrifugó a 10, 000 x g a 4 °C. Finalmente, el sedimento se secó a temperatura ambiente y luego se almacenó a -80 °C para uso futuro.

1.2.4 Extracción de proteína.

Para la extracción de proteínas y el análisis adicional de proteínas, todas las muestras de raíces individuales se agruparon de acuerdo con su rendimiento de perezona, utilizando una prueba de media estadística (k-medias) para la agrupación. De acuerdo con esta prueba, se generaron 3 grupos: productores de perezona altos, medios y bajos. Cada grupo se dividió para tener tres grupos de muestras como réplicas biológicas de cada grupo.

El polvo seco y limpio del tejido (aproximadamente 0.25 g) se homogeneizó en microtubos de 2 ml utilizando 0.7 ml de buffer extracción que contenía sacarosa 0.7 M, Tris-HCl 0.5 M, pH 8.0, HCl 0.1 M, EDTA 50 mM, DTT al 1%, PVPP al 1% y un cóctel de inhibidores de proteasas (4-(2-aminoetil) bencenosulfonil fluoruro, E-64, bestatina, leupeptina, aprotinina y EDTA sódico (Sigma-Aldrich). El homogenizado se añadió con 0.7

ml de fenol saturado con Tris, pH 8.0 (Sigma, St. Louis, MO, EE. UU.). La mezcla se agitó a máxima velocidad en vortex durante 30 segundos y se incubó con agitación orbital en hielo durante 1 hora. La fase de fenol se separó por centrifugación a 10,000 xg durante 20 minutos a 4 °C. La fase de fenol se recuperó y se pasó a nuevos microtubos de 2 ml. La fase acuosa restante se volvió a extraer con 0.7 ml de fenol saturado con Tris, pH 8.0 y 0.7 ml de tampón de extracción. Las proteínas se precipitaron de las fases de fenol reunidas agregando 5 volúmenes de Acetato de amonio 0.1 M en metanol en frío, se incubó a -20 ° C durante toda la noche. El pellet de proteínas se lavó dos veces con acetato de amonio 0.1 M en metanol y dos veces con acetona al 80% fría. El pellet se resuspendió en urea 6 M. Finalmente, la concentración de proteínas se determinó mediante el ensayo de proteínas RCDC (BIO-RAD, Madrid, España) basado en el ensayo de Lowry modificado (Raghupathi y Diwan, 1994).

1.2.5. Separación de proteínas por 2-DE, adquisición y análisis de imágenes.

Las proteínas fueron solubilizadas en 125 µL de buffer de rehidratación (Urea 7M, Thiourea 2 M, CHAPS 4%, IPG buffer 2%, azul de bromofenol 0.0002%), la mezcla se colocó en una tira de 20 cm con gradiente de pH de 3-10 y 4-7 y se rehidrataron por 16 h continuas, después de este tiempo, se corrió la primera dimensión o isoelectroenfoque (IEF) a 12 kVh. Para el SDS-PAGE, las tiras fueron equilibradas 15 min con buffer de equilibrio (6 M urea, 75 mM Tris-HCl pH 8.8, 29.3% glicerol, 2% SDS, 0.002% azul de bromofenol y DTT 60 mM) y otros 15 min con el mismo buffer, pero ahora con iodoacetamida 18.4 mM, posterior a esto, las tiras fueron corridas en un SDS-PAGE al 13%. Los geles fueron teñidos con azul de coomasie y escaneados con un scanner Bio-Rad GS-800 Calibrated Densitometer (Ramírez-Ramírez y col., 2008).

El análisis de las imágenes de los geles 2-DE se llevó a cabo con el software Progenesis SameSpots v 4.0 (Nonlinear Dynamics). Se realizaron 3 réplicas biológicas de los 3 grupos experimentales: El grupo de las altas (A) productoras de perezona etiquetadas como A1, A2 y A3, grupo de las medias (M) productoras etiquetadas como M1, M2 y M3 y finalmente las bajas (B) productoras etiquetadas como B1, B2 y B3. Una vez realizada la detección y comparación de los spots en cada muestra, las imágenes fueron alineadas en relación con el gel de referencia. Se aplicó una prueba de ANOVA a los spots cuyo cociente normalizado varió.

1.2.6 Identificación de proteínas por MS y comparación en la base de datos.

1.2.6.1 Digestión de proteínas en gel y análisis por MALDI/TOF.

Los spots significativamente diferentes fueron cortados y extraídos del gel 2-DE y para desteñirse se colocaron en una solución con 50 mM bicarbonato de amonio (NH_4HCO_3) en 50% de acetonitrilo, reducido con 10 mM de dithiothreitol (DTT) en 50 mM NH_4HCO_3 , alquilado con 40 mM de iodoacetamida en 50 mM NH_4HCO_3 , posteriormente fueron colocados en 100% de acetonitrilo y se dejaron toda la noche a 37 °C en presencia de tripsina. Los péptidos digeridos fueron extraídos con 0.1% de ácido trifluoroacético (TFA) en 50% de acetonitrilo. Las muestras fueron liofilizadas y posteriormente resuspendidas en 5mg/ml α -Cyano-4-hydroxycinnamic acid (CHCA), 0.1% TFA y 50% de acetonitrilo de acuerdo con el protocolo propuesto por Yan., *et al* 2005.

Los péptidos digeridos fueron analizados por espectrometría de masas MALDI-TOF, para lo cual se generaron listas de masas de péptidos tripticos y se compararon con la base de datos Viridiplantae en NCBI (<http://www.ncbi.nlm.nih.gov/>), utilizando el motor de búsqueda MASCOT (Matrix Science, Ltd., Reino Unido) con los siguientes parámetros: una escisión perdida, carbamidometil de cisteínas como modificación química fija, modificación y oxidación de la metionina como variable y tolerancia de masas 200 ppm.

1.2.7. Digestión de proteínas en líquido y análisis por cromatografía líquida de ultra alto rendimiento acoplada al cuadrupolo de tiempo de vuelo (UHPLC-QTOF).

Los extractos de proteína se redujeron con 5 μl de DTT 0.2 M, la reacción se mantuvo a 37 °C durante 60 min, posteriormente las muestras se alquilaron con 20 μl de yodoacetamida 0.2 M (IAM) a temperatura ambiente durante 60 minutos en la oscuridad. La concentración de urea durante la digestión se redujo diez veces con bicarbonato de amonio (BCA) 25 mM que es compatible con la actividad de la tripsina. La tripsina modificada de grado de secuenciación se agregó a una relación proteína/enzima de 30:1 y las muestras se digirieron a 37 °C durante la noche y finalmente, se agregó la proporción de proteína/enzima tripsina 60:1 y se permitió la digestión durante 4 -5 h adicionales a 37 °C. La digestión triptica se detuvo acidificando la muestra con ácido fórmico. Una vez digeridas, cada muestra se evaporó y resuspendió en TCA al 0.5%, Acetonitrilo al 5% (ACN) para desalarla. Los péptidos se pasaron a través de una columna C18 (Pierce® C18 Spin Columns, Thermo

Scientific), previamente preparados para activarse con buffer de ACN al 50% y equilibrados con una solución de TCA al 0.5% en ACN al 5%. Una vez que los péptidos estuvieron dentro de la columna, se lavaron con un buffer de activación y se eluyeron con 70% de ACN y 0.1% de ácido fórmico, para luego evaporarlos. Los péptidos se resuspendieron en acetonitrilo al 3%, ácido fórmico al 0,1% antes del análisis por LC-MS / MS.

Las digestiones de péptidos desalados se analizaron en un espectrómetro de masas Agilent 6550 iFunnel Q-TOF (Agilent Technologies) acoplado a un sistema Agilent 1290 UHPLC. Las muestras de péptidos se cargaron en una columna de mapeo de péptidos Agilent AdvanceBio (2.1 mm x 250 mm, tamaño de partícula de 2.7 μm , operado a 50 °C) a través de un inyector automático Infinity (Agilent Technologies) con un buffer A (agua, ácido fórmico al 0.1%) que fluye a 0.400 mL/min. Los péptidos se eluyeron en el espectrómetro de masas utilizando un gradiente lineal de 140 minutos de ACN al 3% - 40% en ácido fórmico al 0.1%. Los péptidos se introdujeron en el espectrómetro de masas desde la LC utilizando una fuente Jet Stream (Agilent Technologies) que funciona en modo de ion positivo (3500 V) y en modo de alta sensibilidad. Los parámetros de la fuente emplean la temperatura del gas (250 °C), el gas de secado (14 L/min), el nebulizador a 35 psi, la temperatura del gas de la envoltura (250 °C), el flujo de gas de la envoltura (11 L/min), el VCap (3500 V), fragmentador (360 V), OCT 1 RF Vpp (750 V). Los datos se adquirieron con el software de estación de trabajo Agilent MassHunter, adquisición de datos LC/MS B.08.00 (compilación 8.00.8058.0) operando en modo Auto MS/MS, por lo que los 20 iones más intensos (estados de carga, 2-5) dentro de 300-1700 m/z rango de masa por encima de un umbral de 1000 cuentas se seleccionaron para análisis de MS/MS. Los espectros MS/MS (50-1700 m/z) se recolectaron con el cuadrupolo configurado en resolución "estrecha" y se adquirieron hasta que se recolectaron 25000 conteos totales o para un tiempo de acumulación máximo de 333 ms.

Cada espectro MS/MS se preprocesó con la herramienta de extracción de Spectrum Mill Proteomics Workbench (Agilent) para obtener una lista de picos y mejorar la calidad espectral combinando los espectros MS/MS con el mismo precursor ($\Delta m/z < 1.4$ Da y cromatografía $\Delta t < 15$ s). Se realizó una búsqueda en el conjunto de datos reducido contra un subconjunto de la base de datos de proteínas NCBIInr compuesta por secuencias de la familia Asteraceae y proteínas contaminantes sin restricciones taxonómicas en el modo de identidad

con la herramienta de búsqueda MS/MS de Spectrum Mill Proteomics Workbench y con las siguientes configuraciones: tripsina, hasta 2 escisiones perdidas, carbamidometilación de cisteínas como modificaciones fijas, oxidación de metionina como variable, tolerancia de masa de 20 ppm para el precursor y 50 ppm para los iones del producto. Los hits de péptidos se validaron en el modo péptido (FDR <1.2%) y luego en el modo de proteína de acuerdo con la configuración de puntaje recomendada por el fabricante. Las identificaciones positivas se consideraron solo cuando se emparejaron dos o más péptidos, y su puntuación total fue > 20.

1.2.8 Análisis cuantitativo sin etiqueta (LC-MS).

El análisis LC-MS sin etiquetas se llevó a cabo utilizando Progenesis QI para el software de proteómica versión 4.1 (NonLinear Dynamics, Reino Unido), según lo recomendado por el fabricante (www.nonlinear.com para obtener más información sobre la alineación, la normalización, el cálculo de la abundancia de péptidos, etc.). Cada ejecución de la muestra se sometió a una alineación que involucró alinear los datos de acuerdo con el tiempo de retención por Cromatografía líquida (CL) de cada una de las muestras, esto permite que cualquier desviación en el tiempo de retención quede en un tiempo de retención ajustado para todas las ejecuciones en el análisis. Posteriormente, la ejecución de la muestra que produjo la mayoría de los iones peptídicos se usó como ejecución de referencia, a esta ejecución se le ajustó el tiempo de retención de todas las demás ejecuciones y se normalizaron las intensidades máximas. El algoritmo de cuantificación de péptidos, Progenesis lo calculó de acuerdo a la abundancia de los péptidos y como la suma de las áreas de los picos dentro de sus límites isotópicos. Cada valor de abundancia se transformó en un valor de abundancia normalizado aplicando un factor de escala global. La abundancia de proteínas se calculó como la suma de las abundancias de todos los iones peptídicos que se identificaron como provenientes de la misma proteína (Meleady et al 2012).

1.2.9 Categorización y anotación funcional.

Los cambios significativos en la abundancia relativa de proteínas para todas las comparaciones fueron al menos de 2.0 veces ($\text{foldlog}_2 >1$ o $\text{foldlog}_2 <-1$), las proteínas expresadas diferencialmente se clasificaron por su ontología genética (GO) utilizando el software Blast2GO v5 (<https://www.blast2go.com/>) (Conesa y Götz, 2008). Se asignaron

tres vocabularios GO: el proceso biológico, la función molecular y el componente celular. Se generó un archivo de secuencias FASTA a partir del grupo de proteínas previamente identificadas y cuantificadas utilizando el sitio web de NCBI. Blast2GO fue alimentado con un archivo FASTA; primero se introdujo la descripción de la secuencia mediante una búsqueda BLASTp frente a NCBItr (corte de valor $e^{-1} \times 10^{10}$), luego se mapearon los términos GO, IC e Interpro, seguidos de una anotación de secuencia (1 x 10⁶ E-value Hit-Filter, corte la cobertura de 0 Hsp-Hit, el límite de anotación de 55 y el peso de 5 GO). El análisis de las vías metabólicas en las que se clasificaron las proteínas identificadas se llevó a cabo mediante el uso de la base de datos de la Enciclopedia de genes y genomas de Kyoto (KEGG) (<http://www.genome.jp/kegg/>) (Kanehisa et al. 2017). Finalmente, se realizó la interacción proteína-proteína para las proteínas identificadas utilizando el software Cytoscape 3.6.1 (Shannon et al. 2003). La red de interacción proteína-proteína se obtuvo de la base de datos STRING; utilizamos las proteínas homólogas correspondientes de la puntuación más alta de *Arabidopsis thaliana* (<http://string-db.org/>) (Szklarczyk et al. 2017). El formulario de red de interacción STRING se visualizó en Cytoscape (<http://www.cytoscape.org/>) y la agrupación de redes se realizó empleando el algoritmo de expansión de “vecindad” de superposición (ClusterONE) para identificar las regiones conectadas dentro de las redes con posible superposición, para lo cual se identificaron solo los clusters con un valor significativo para nuestra comparación con un valor de $p < 0.05$ (Nepusz y Paccanaro, 2012).

1.3. RESULTADOS.

1.3.1 Cuantificación de perezona en raíces de *A. cordata*.

El análisis de cromatografía de gases acoplada a la espectrometría de masas permitió la identificación y cuantificación de la perezona con un tiempo de retención de entre 17.62 min y 17.75 min, con un patrón de fragmentación $m/z = 166, 191, 205$ y 248. Nuestros resultados mostraron que la perezona no varió significativamente entre los sitios de estudio, pero que hubo una gran variabilidad en la producción de perezona entre las plantas de un mismo sitio, ya que el rendimiento del compuesto varía de 0.329 ± 0.004 a 7.757 ± 0.170 mg / g de raíz en peso seco. Esta observación fue el resultado de tres réplicas biológicas. De esta manera,

encontramos altos productores de perezona cuyos rendimientos están entre 7.757 ± 0.170 y $4.474 \pm 2.9405 \times 10^{-5}$ mg/g de raíz, productores medios que van de 4.169 ± 0.4868 a 2.881 ± 0.0608 mg / g de raíz y los productores bajos de que producen entre 0.990 ± 0.017 y 0.329 ± 0.004 mg / g de raíz (Figura S1).

1.3.2 Comparación de perfiles proteicos 2-DE de raíces de *A. cordata*.

Para el análisis de los extractos totales de proteínas se formaron 3 grupos experimentales que fueron creados utilizando una prueba de cluster por k-medias de acuerdo con la prueba los tres grupos experimentales fueron: altos (A), medios (M) y bajos (B) productores de perezona. Partiendo de este análisis, el siguiente manejo de las muestras fue realizar pooles para cada grupo experimental tomando en cuenta solo el rendimiento del compuesto en cada muestra y no el sitio de colecta, debido a que no existen diferencias significativas en el rendimiento del compuesto al comparar por sitio de colecta. El análisis comparativo entre las imágenes de los geles de 2-DE entre los tres grupos experimentales demostró que existe una gran similitud entre los patrones proteicos de las raíces altamente, medianamente y bajas productoras de perezona. En el análisis de las imágenes de los geles de electroforesis bidimensional 2-DE se lograron detectar 360 spots de los cuales solo 6 son spots diferenciales entre los 3 grupos experimentales (ANOVA $p < 0.02$ y fold change ≥ 1.5). En la figura 15A se muestra la imagen del gel 2-DE de referencia utilizado para mostrar los spots seleccionados. El análisis de componentes principales (PCA) muestra un agrupamiento claro de los spots detectados lo cual coincide con los 3 grupos experimentales formados, el porcentaje de variabilidad explicada por el PCA fue de 86.84%, la calidad estadística de este análisis empleando tres replicas biológicas fue de 83.3% (> 0.8), lo que significa que el poder de la prueba estadística esta dentro del rango aceptable de acuerdo a los parámetros manejados por el software, ya que un umbral de potencia generalmente aceptado es del 80% (figura 15B). Los spots seleccionados se distribuyen en tres clusters como se observa en la figura 15C: en el cluster 1 se encuentran representados los spots que incrementan su abundancia en el grupo de mediana (M) producción, comparado con los grupos altos y bajos productores, el cluster 2 está representado por un spot que se encuentra menos incrementado en cuanto a su abundancia en el grupo M y finalmente el cluster 3 que engloba a spots más incrementados en el grupo B en comparación con el grupo A y M.

En total se encontraron 6 spots diferencialmente expresados de los cuales solo 4 fueron identificados por MALDI-TOF, estos spots fueron 207, 271, 277 y 386 que se identificaron como CFI_RAPSA, PP134_ARATH, XP_017604040.1 y XP_022021167.1 respectivamente como se muestra en la tabla 4.

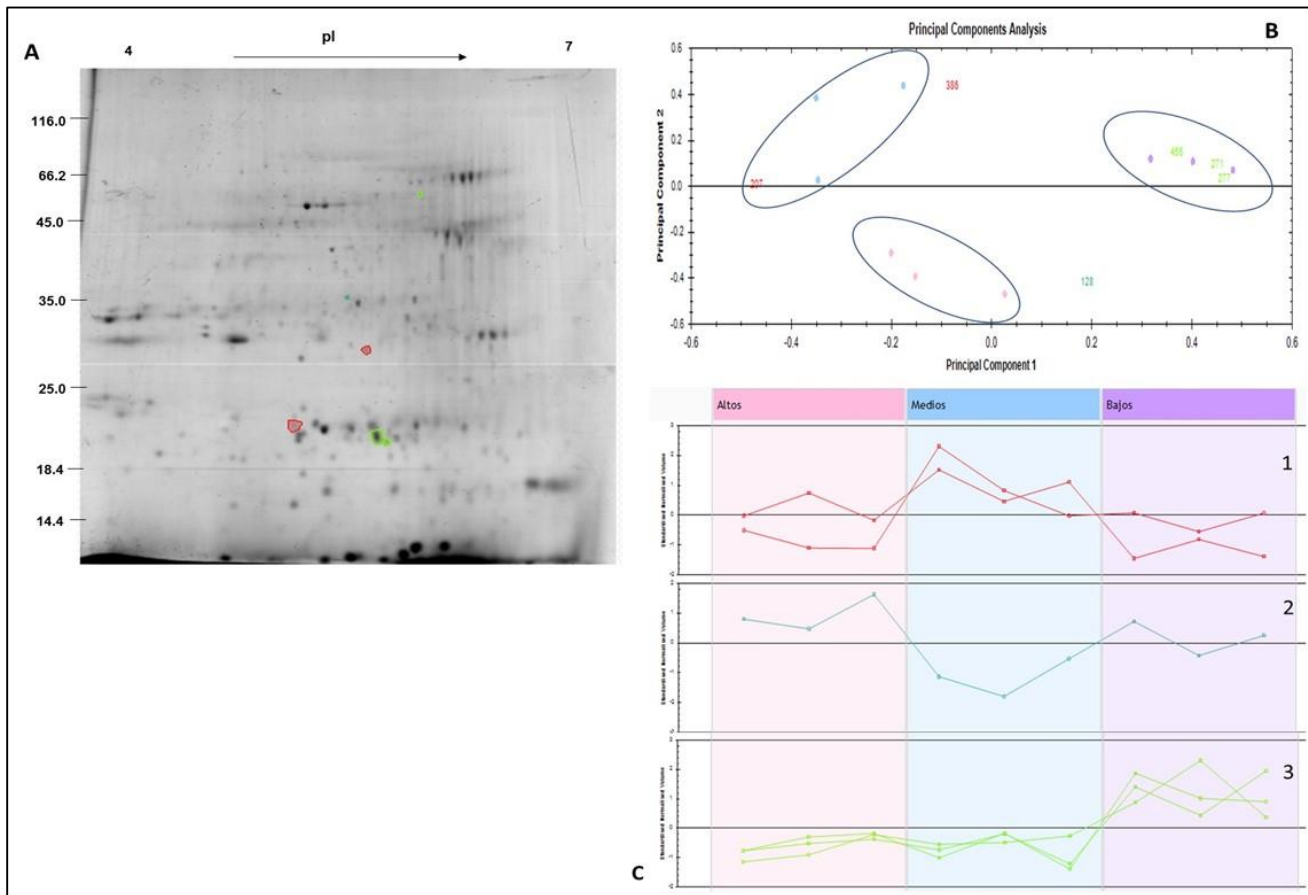


FIGURA 15. Análisis de las imágenes de 2-DE de los extractos proteicos de las raíces productoras de perezona. **A)** Imagen del gel 2-DE de referencia que señala los spots con diferencias significativas en abundancia entre los tres grupos experimentales. Cada spot seleccionado pertenece a un cluster representado por un código de color. Rojo (cluster 1), azul (cluster 2), verde (cluster 3). **B)** Imagen del análisis de PCA que muestra el primer y segundo componente los cuales en conjunto explican el 86.84% de la variabilidad de los spots seleccionados. Spots representados por código de color. Puntos rosas representa al grupo altamente productor, puntos azules al grupo de medios productores y puntos morados al grupo de bajos productores. **C)** Distribución de los spots seleccionados en tres clusters. Cada línea representa la abundancia estandarizada de cada spot en todos los geles analizados.

Se realizó una búsqueda en las bases de datos de UniProt, NCBIprot y SwissProt para describir las identidades de las proteínas antes mencionadas, los resultados mostraron que CFI_RAPSA corresponde a una proteína que participa en la biosíntesis de metabolitos secundarios, específicamente en la biosíntesis de flavonoides. PP134_ARATH se sugiere que pertenece al grupo de proteínas mitocondriales. XP_017604040.1 se identificó como una proteína de nodulación. Finalmente, XP_022021167.1 se sugiere que pertenece a la familia de las proteínas F-box que son conocidas por asociarse a funciones celulares como la transducción de señales y la regulación del ciclo celular. Las proteínas F-box están involucradas en el crecimiento y desarrollo vegetativo y reproductivo de muchas plantas.

Tabla 4. Proteínas expresadas diferencialmente identificadas por MALDI-TOF MS en *A. cordata*.

No. Spot ^a	No. acceso ^b	Proteína identificada	Organismo	MASCOT Score ^c	% Covertura ^d	Matches (Peptidos)	pI Teorico	Mr Teorico
207	CFI_RAPSA	Chalcone--flavonone isomerase	<i>Raphanus sativus</i>	62	27	8	4.82	26900
271	PP134_ARATH	Pentatricopeptide repeat-containing protein At1g80150, mitochondrial	<i>Arabidopsis thaliana</i>	61	25	9	9.40	45933
277	XP_017604040.1	Nodulation protein H-like	<i>Gossypium arboreum</i>	86	14	7	9.00	44836
386	XP_022021167.1	F-box/FBD/LRR-repeat protein At1g13570-like	<i>Helianthus annuus</i>	90	16	11	5.1	73832

^a Número de spot correspondiente en el gel 2-D, se muestra en la figura 15A

^b Número de accession

^c Mascot score.

^d Porcentaje de secuencia proteica predicha cubierta por péptidos emparejados

1.3.3 Identificación y clasificación funcional de las proteínas diferencialmente expresadas.

Los datos obtenidos por espectrometría de masas (MS) de cada grupo de plantas de *A. cordata* se compararon con la base de datos NCBI para la familia Asteraceae, utilizando el motor de búsqueda Spectrum Mill Proteomics Workbench. De esta manera se identificaron un total de 2122 proteínas.

Los cambios significativos en la abundancia relativa de proteínas para todas las comparaciones fueron al menos de 2.0 veces ($\text{foldlog}^2 > 1$ o $\text{foldlog}^2 < -1$). El software Blast2Go (Götz et al. 2008) se utilizó para la asignación automática de la descripción de proteínas, también se usó para tener la anotación de las secuencias comparadas por homología con la base de datos de NCBI. La comparación de las proteínas sobreexpresadas y subexpresadas en los tres grupos se analizó y clasificó según su ontología genética (GO) de la siguiente manera: proceso biológico (GOBP), función molecular (GOMF) y componente celular (GOCC). Es importante tener en cuenta que algunas proteínas se identificaron y clasificaron en más de una categoría GO, de tal manera que el número total de proteínas en cada gráfico puede ser mayor que el número inicial.

1.3.6. Comparación proteómica entre plantas altamente productoras y de baja producción de perezona.

Se realizó el análisis de enriquecimiento de proteínas sobre y subexpresadas; derivado de este análisis los principales términos GO enriquecidos significativamente y las vías se muestran en la figura 16. La categoría de GOBP más enriquecida en las proteínas sobreexpresadas fueron: el proceso del metabolismo secundario, regulación del proceso biosintético de macromoléculas, desenrollado de ADN dúplex, regulación de la proteólisis, regulación de la biosíntesis de compuestos nitrogenados, entre otros. Para la categoría GOCC fueron: la cobertura o envoltura membranal plástidica y cloroplástidica, finalmente para la categoría GOMF fueron: la unión a enzimas, unión a proteínas, actividad de oxidoreductasa, actividad de helicasa, actividad de pirofosfatasa, entre otros. (Figura 16A).

Para el caso de las subexpresadas, las categorías enriquecidas para GOBP fueron: la transducción de señales, la comunicación celular, la cascada MAPK, la regulación de la actividad de transferasa. Para la categoría GOMF fueron: la transferencia de grupos que contienen fósforo, la actividad de quinasas para las proteínas serina/treonina y la actividad oxidoreductasa (Figura 16B).

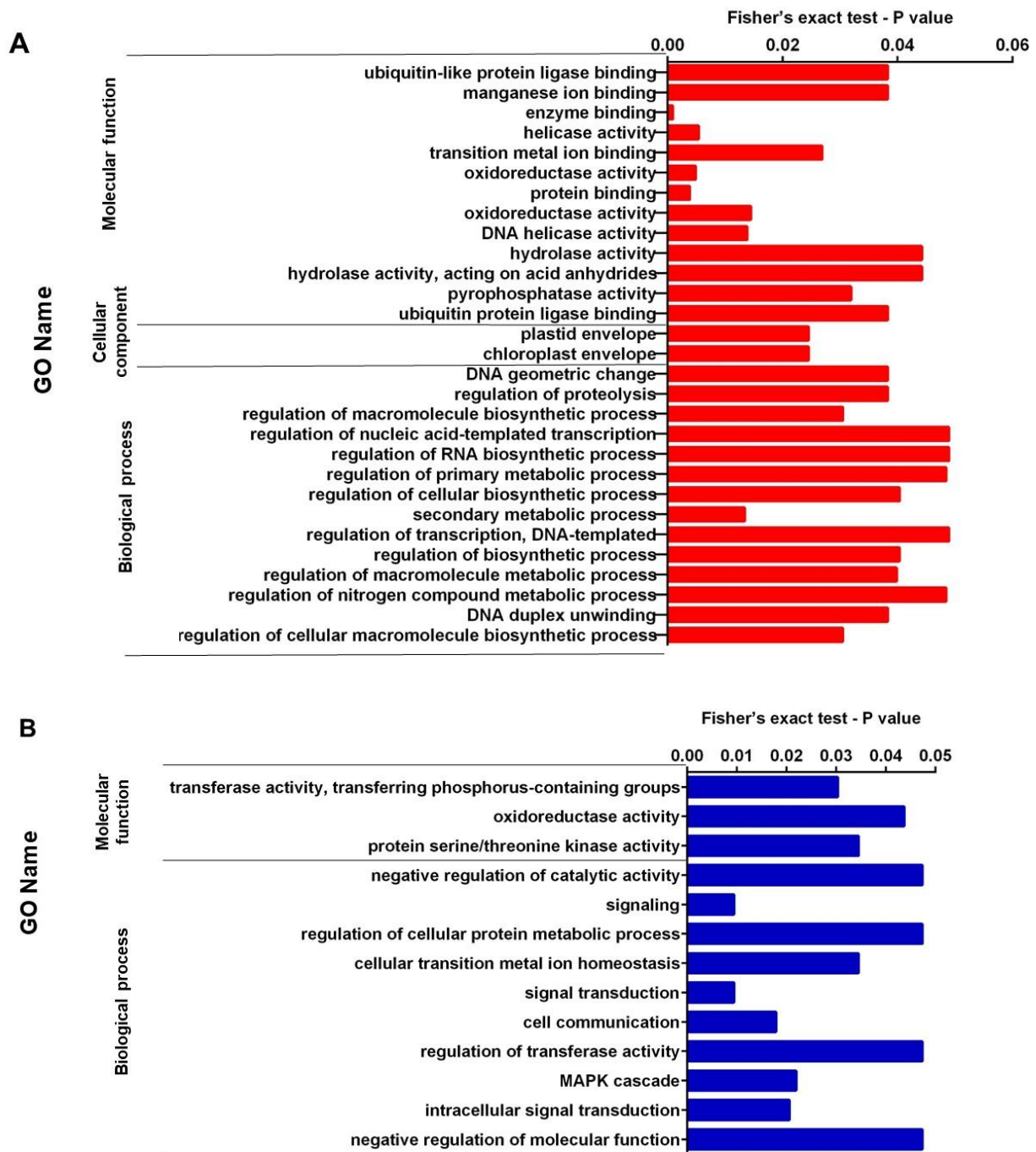
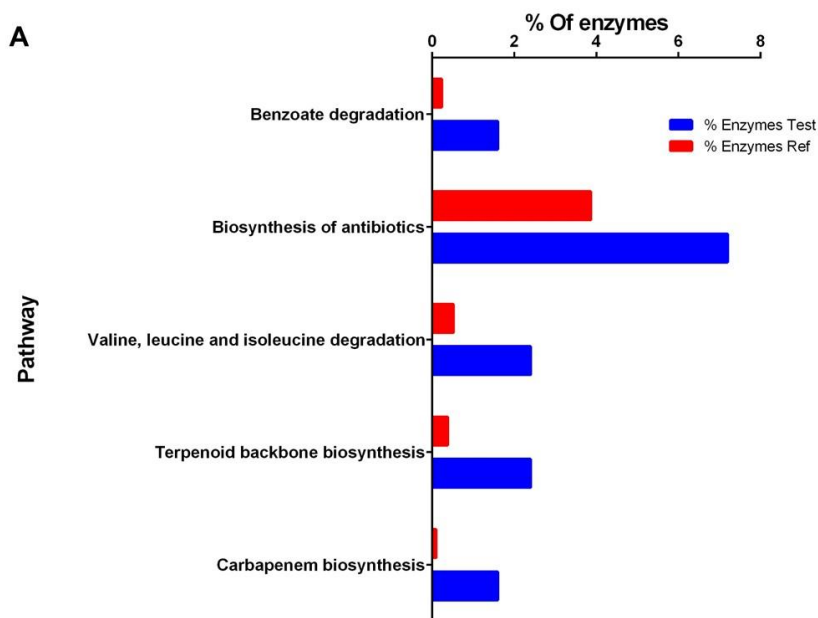


FIGURA 16. Análisis de enriquecimiento de la anotación funcional de las proteínas sobreexpresadas en plantas con alta producción de perezona (A) y proteínas subexpresadas en plantas con baja producción de reperezona (B). Comparación entre los grupos de alta y baja producción de perezona.

El conjunto de proteínas sobreexpresadas y subexpresadas también fue analizado por Kyoto Encyclopedia of Genes and Genomes (KEGG). Se realizó un análisis de enriquecimiento utilizando la prueba exacta de Fisher (p -value <0.05). Las rutas enriquecidas KEGG para las proteínas sobreexpresadas de la comparación entre el grupo alto contra el bajo productor de perezona fue la biosíntesis de carbapeneno, la biosíntesis de terpenoides, degradación de valina, leucina e isoleucina, la biosíntesis de antibióticos y la degradación del benzoato (Figura 17A). Por otro lado, las principales rutas KEEG enriquecidas para las proteínas subexpresadas de esta misma comparación fueron agrupadas en rutas del metabolismo del piruvato, metabolismo de las purinas, glucólisis/gluconeogénesis, metabolismo de la tiamina, metabolismo del glioxilato y dicarboxilato, fosforilación oxidativa, biosíntesis de hormonas esteroideas, metabolismo de las galactosas, metabolismo de drogas, metabolismo de las pirimidinas, vía de pentosas fosfato, metabolismo de aminoazúcares y azúcar nucleótidos, biosíntesis de antibióticos, metabolismo del almidón y sacarosa y metabolismo del nitrógeno (figura 17B).



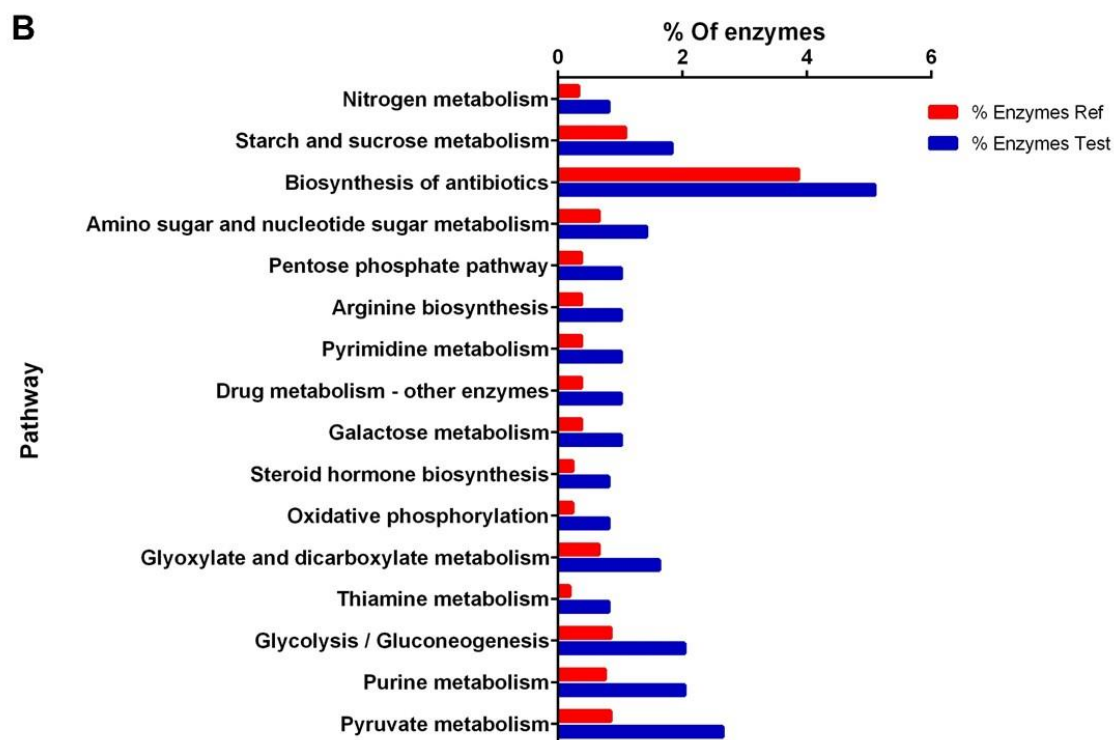


FIGURA 17. Análisis de enriquecimiento de las vías KEGG de las proteínas diferencialmente expresadas entre los grupos de alta y baja producción de perezona. **A)** Conjunto de proteínas sobreexpresadas en la comparación de los grupos altos contra bajos productores de perezona; **B)** Conjunto de proteínas subexpresadas en la comparación de los grupos de altos contra bajos productores de perezona.

1.3.4 Análisis de interacción proteína-proteína.

Además, buscamos las interacciones predichas para las proteínas diferencialmente expresadas. El análisis de la red de proteínas mostró que, en el grupo de proteínas sobreexpresadas, se crearon cinco clusters. La red predijo una interacción entre las proteínas AT1G11660, FPS1, HSC70-1, P5CS2, AT5G47720, MFP2, AIM1, GAPC2, PKT4, PKT3, AT4G13010, ACAT2, AOR, KAT5 (cluster uno); este grupo de proteínas codifica para proteínas que están involucradas en procesos de defensa como respuesta para contender con el estrés, la biosíntesis de metabolitos secundarios y ácidos grasos. La segunda interacción se observó entre PHYA, FUS5, COP8, COP13, FUS12, COP9, FUS6; este grupo de proteínas que están involucradas en procesos tales como la respuesta a estímulos por estrés abiótico.

Además, se observó la formación del cluster tres con la interacción entre ESP3, At1g80070, MAC3A, MAC3B, CDC5, AT5G46840, MOS4, AT3G19810; proteínas implicadas en el proceso de “splaiçosoma”. Y finalmente, una interacción entre UPF3, DCP2, LBA1, DCP1, AT2G39260 y AT2G40770, AT1G30680, SRS2 que forman el cluster cuatro y cinco respectivamente, e involucran a proteínas involucradas en el desarrollo y adaptación de la planta (Figura 18A).

Por otro lado, en el grupo de proteínas subexpresadas se crearon nueve agrupaciones o clusters, la primera agrupación fue entre las proteínas LOS1, AT2G45030, RPS11, RPS5A, AT1G67430, RPS5B, AT1G07210, AT3G09630, AT2G09990, AT3G13580, AT1G80750, RPL23AB, AT4G36130, AT3G49910, AT4G15000, RPL3B, P40, AT4G34670, AT1G70600, AT2G19750, AT1G04480, AT3G02560, EMB3010, AT5G04800, AT4G11120, BBC1, RPL5B, AT5G48760, rps15, AtCg00790, AT5G59240, AT5G39850, AT5G65220, este grupo de proteínas están involucradas en procesos de traducción en el ribosoma, algunas se describen como constituyentes estructurales del ribosoma, además de expresarse en otras ocasiones como proteínas involucradas en la respuesta para contener al estrés causado por frío.

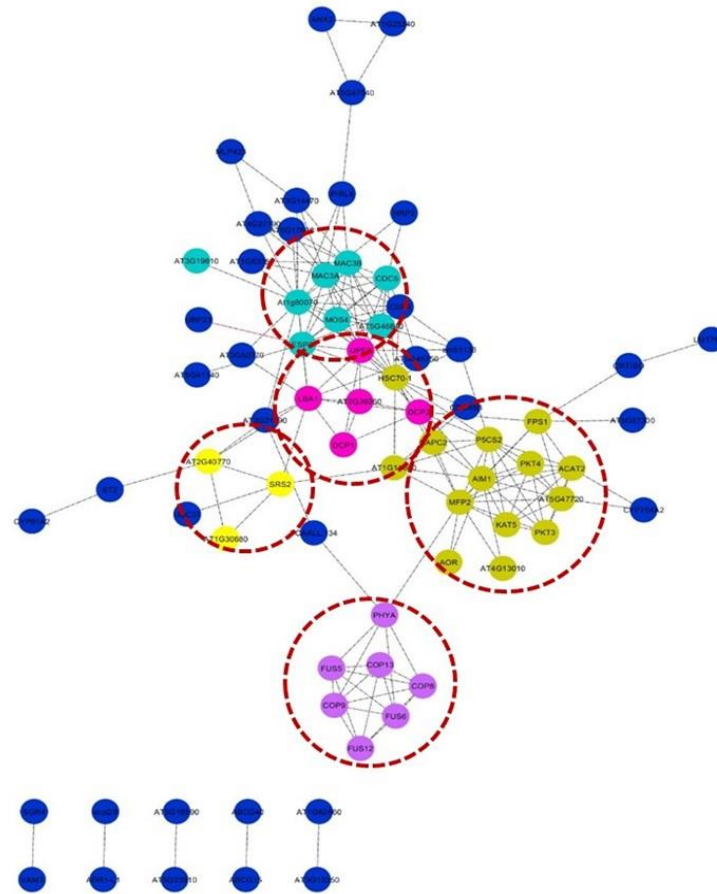
El segundo cluster fue creado por la interacción entre AT5G51570, VHA-A3, AT3G42050, AT3G28715, VHA-A, TUF, las cuales son proteínas vacuolares. También se observó una interacción entre proteínas como NUP155, NUP1, NTF2B, EMB3142, RAN3, las cuales conformaron el cluster tres; este grupo de proteínas está involucrado en transporte, específicamente en el transporte nucleocitoplásmico. El cluster cuatro fue formado por la interacción entre EMB1467, NDHF, MATK y NAD7, estas proteínas son proteínas transportadoras que forman parte del transporte de electrones acoplado a la biosíntesis de ATP, la respiración celular, la fotorrespiración y la respuesta al estrés oxidativo. En el cluster cinco se observó la interacción entre AT3G53980, AGP30 y AT1G17860; en esta agrupación se encuentran proteínas involucradas en procesos como la resistencia sistémica adquirida, la regulación del desarrollo de la raíz, la regulación del proceso de latencia de las semillas y la actividad inhibidora de la endopeptidasa. También se encontró una interacción entre las proteínas MDH, GDH2, PGMP, ALDH11A3, PKP-ALPHA, NADP-ME4, HXK2, NADP-ME1, PPC3, CSY2, PMDH1, TPI, GLN1-1, IDH-V las cuales conforman el cluster seis, en este grupo se encuentran proteínas involucradas en procesos como metabolismo de malato,

respuesta al frío, metabolismo celular de aminoácidos, respuesta al cadmio, respuesta al estrés salino, biosíntesis de ácidos grasos, señalización dependiente de hexocinasa, muerte celular programada, desarrollo foliar y fotosíntesis, entre otros. El cluster siete fue formado por la interacción entre LD, FLK y AT2G47820, este grupo de proteínas está involucrado en procesos de la diferenciación celular, procesos del desarrollo y regulación del desarrollo de las flores, la transición vegetativa a la fase reproductiva del meristemo, entre otros procesos.

Además se observó la interacción entre AT5G46940, UGD3, RGP5 la cual formó el cluster ocho, en este cluster se encuentran proteínas involucradas en procesos tales como la actividad del receptor de glutamato ionotrópico, el metabolismo de carbohidratos, constituyentes de la pared celular, y la respuesta al estrés salino. Finalmente, el cluster nueve fue formado por la interacción entre ABCF4, MSR2, AT2G4629; estas son proteínas involucradas en procesos de transporte, metabolismo de manosa y la formación del complejo de iniciación de la traducción citoplásmica (Figura 18B).

A

Cluster	Members
1	AT1G11660, FPS1, HSC70-1, P5CS2, AT5G47720, MFP2, AIM1, GAPC2, PKT4, PKT3, AT4G13010, ACAT2, AOR, KAT5
2	PHYA, FUS5, COP8, COP13, FUS12, COP9, FUS6
3	ESP3, At1g80070, MAC3A, MAC3B, CDC5, AT5G46840, MOS4, AT3G19810
4	UPF3, DCP2, LBA1, DCP1, AT2G39260
5	AT2G40770, AT1G30680, SRS2



B

Cluster	Members
1	LOS1, AT2G45030, RPS11, RPS5A, AT1G67430, RPS5B, AT1G07210, AT3G09630, AT2G09990, AT3G13580, AT1G80750, RPL23AB, AT4G36130, AT3G49910, AT4G15000, RPL3B, P40, AT4G34670, AT1G70600, AT2G19750, AT1G04480, AT3G02560, EMB3010, AT5G04800, AT4G11120, BBC1, RPL5B, AT5G48760, rps15, AtCg00790, AT5G59240, AT5G39850, AT5G65220
2	AT5G51570, VHA-A3, AT3G42050, AT3G28715, VHA-A, TUF
3	NUP155 NUP1 NTF2B EMB3142 RAN3
4	EMB1467, NDHF, MATK, NAD7
5	AT3G53980, AGP30, AT1G17860
6	MDH, GDH2, PGMP, ALDH11A3, PKP-ALPHA, NADP-ME4, HXK2, NADP-ME1, PPC3, CSY2, PMDH1, TPI, GLN1-1, IDH-V
7	LD, AT2G47820, FLK
8	AT5G46940, UGD3, RGP5
9	ABCF4, MSR2, AT2G46290

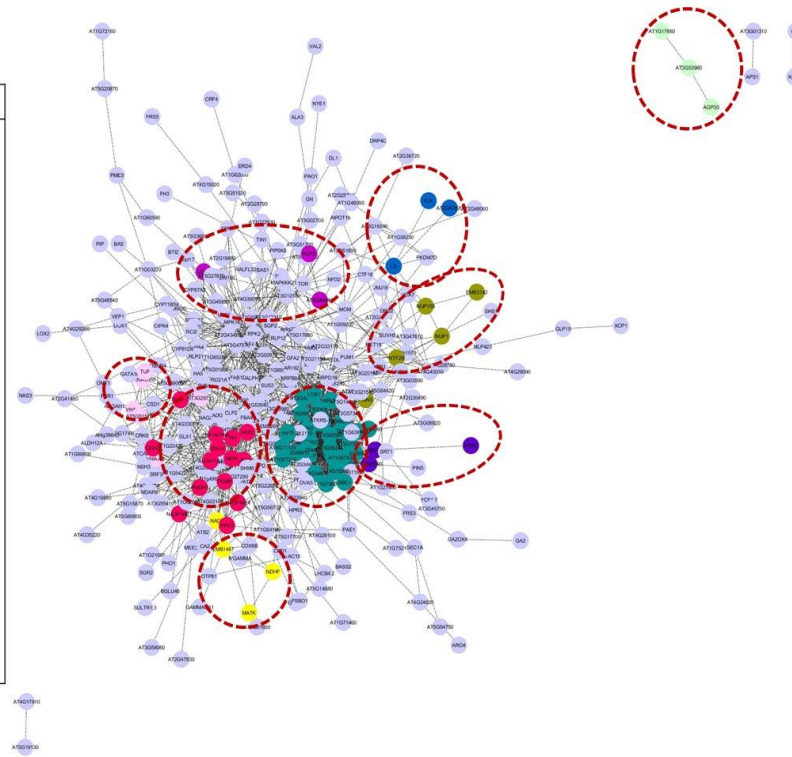


FIGURA 18. A) Redes de interacción proteína-proteína de las proteínas diferencialmente sobreexpresadas. Los puntos verdes representan el grupo uno (AT1G11660, FPS1, HSC70-1, P5CS2, AT5G47720, MFP2, AIM1, GACP2, PKT4, PKT3, AT4G13010, ACAT2, AOR, KAT5), los puntos púrpuras representan el grupo dos (PHYA, FUS5, 8, 8), FUS12, COP9, FUS6), los puntos azules representan el grupo tres (ESP3, At1g80070, MAC3A, MAC3B, CDC5, AT5G46840, MOS4, AT3G19810), los puntos de color rosa representan el grupo cuatro (UPF3, DCP2, LBA1, DCP1, AT2G39260) los puntos de color amarillo representan el grupo cinco (AT2G40770, AT1G30680, SRS2). **B)** Redes de interacción proteína-proteína de las proteínas diferencialmente subexpresadas. Los puntos verdes representan el grupo uno (LOS1, AT2G45030, RPS11, RPS5A, AT1G67430, RPS5B, AT1G07210, AT3G09630, AT2G09990, AT3G13580, AT1G80750, RPL23AB, AT4G36130, AT3G49910, AT4G15000, RPL3B, P40, AT4G34670, AT1G70600, AT2G19750, AT1G04480, AT3G02560, EMB3010, AT5G04800, AT4G11120, BBC1, RPL5B, AT5G48760, rps15, AtCg00790, AT5G59240, AT5G39850, AT5G65220), los puntos rosas representan el grupo dos (AT5G51570, VHA-A3, AT3G42050, AT3G28715, VHA-A, TUF), puntos verde oscuro representan el grupo tres (NUP155 NUP1 NTF2B EMB3142 RAN3), puntos amarillos representan el grupo cuatro (EMB1467, NDHF, MATK, NAD7), puntos verde claro representan el grupo cinco (AT3G53980, AGP30, AT1G17860), los puntos rosa oscuro representan el grupo seis (MDH, GDH2, PGMP, ALDH11A3, PKP-ALPHA, NADP-ME4, HXK2, NADP-ME1, PPC3, CSY2, PMDH1, TPI, GLN1-1, IDH-V), los puntos azul rey representan el grupo siete (LD, AT2G47820, FLK), los puntos morados representan el grupo ocho (AT5G46940, UGD3, RGP5) y los puntos morado oscuro representan el grupo nueve (ABCF4, MSR2, AT2G46290).

1.4. DISCUSIÓN.

A. cordata es una planta medicinal que se ha utilizado como fuente para la producción de perezona y otros sesquiterpenos con varias aplicaciones de importancia farmacológica, a pesar de esto es una especie sin secuencia ni anotación de su genoma, por lo tanto, se considerada una "planta no modelo" o lo que se conoce como una "especie huérfana". La investigación proteómica con un modelo como este es muy desafiante. Sin embargo, en esta investigación, fue posible realizar la exploración y el análisis del proteoma de las raíces productoras de perezona gracias a los avances en las herramientas proteómicas y las bases de datos utilizadas en este estudio. En este sentido, podemos mencionar algunos ejemplos de organismos huérfanos en los que se desarrollaron estudios proteómicos para el análisis del metabolismo secundario, algunas de estas especies son: *Picrorhiza kurroa*, *Podophyllum hexandrum*, *Panax ginseng*, *Catharanthus roseus*, *Withania somnifera*, *Euphorbia kansui*, *Silybum marianum*, entre otras especies (Senthil et al., 2011; Bhattacharyya et al., 2012; Champagne et al., 2012; Corchete y Bru 2013; Gharechahi et al., 2013; Ma et al., 2013; Sud et al., 2014 ; Zhao et al, 2014; Romero-Sandoval et al., 2017).

De manera general se clasificaron funcionalmente a proteínas que se encontraron sobreexpresadas o subexpresadas en el metabolismo de glicólisis y gluconeogénesis; en estos procesos se obtiene energía y esqueletos de carbono al oxidar o sintetizar la glucosa para el crecimiento y desarrollo de las plantas entre otros procesos. Del metabolismo de glicólisis se obtienen también moléculas precursoras como piruvato y acetil-CoA, estas moléculas son iniciadoras en otras rutas importantes como el metabolismo de lípidos, el ciclo de los ácidos tricarboxílicos, biosíntesis de terpenos y biosíntesis de aminoácidos.

Encontramos enzimas sobreexpresadas en el grupo de alta producción de perezona que participan en la vía metabólica de los terpenos; estas enzimas son acetil-CoA, C-acetiltransferasa, geranil-difosfato sintasa y farnesil difosfato sintasa, las tres participan en la biosíntesis de terpenos a través de la vía del mevalonato (MVA), una vía que tiene lugar en el citosol. La farnesil difosfato sintasa es una enzima clave en la biosíntesis de sesquiterpenos, cataliza las condensaciones consecutivas de dimetilalifosfato con isopentenil pirofosfato para producir geranil difosfato, que a su vez, con la adición de isopentenil pirofosfato o dimetilalil difosfato produce farnesil difosfato. Esta molécula es un precursor de sesquiterpenos y otros terpenos como esteroides, brasinoesteroides, farnesol, nerolidol,

germacreno, valenceno, entre otros (Lombard y Moreira, 2011); y probablemente también participe en la biosíntesis de perezona. Los sesquiterpenos son en su mayoría compuestos volátiles liberados por las plantas después del daño causado por los herbívoros, además, ciertos sesquiterpenos son fitoalexinas involucradas en la defensa directa de las plantas contra los herbívoros y patógenos microbianos (Dudareva et al. 2004).

Además, se encontraron enzimas sobreexpresadas, como el componente 2-oxoisovalerato deshidrogenasa E2, malonato-semialdehído deshidrogenasa y metilmalonato-semialdehído deshidrogenasa que participan en la vía de degradación de valina, leucina e isoleucina. En plantas, el catabolismo de aminoácidos es especialmente relevante en situaciones de estrés metabólico; condiciones de estrés ambiental y de desarrollo. En caso de estrés, estos aminoácidos se degradan rápidamente. Además, el catabolismo de aminoácidos es esencial para la respiración en condiciones de baja intensidad luminosa o en la oscuridad prolongada. En caso de escasez de luz, la disponibilidad de carbohidratos para la respiración es limitada, y los aminoácidos se pueden usar como sustratos alternativos (Araújo et al. 2011).

De acuerdo con lo anterior, sugerimos que la degradación de los aminoácidos genera moléculas de acetil-coA, que pueden usarse para la biosíntesis de terpenos. Este proceso podría ocurrir en plantas de *A. cordata* con alto rendimiento de perezona, ya que se observó una sobreexpresión de la biosíntesis de terpeno junto con la degradación de valina, leucina e isoleucina.

Por otro lado, observamos que varias enzimas del metabolismo de los carbohidratos y el metabolismo de la energía estaban subexpresadas, como la piruvato descarboxilasa, trifosfato isomerasa, fructosa-bifosfato aldolasa, piruvato-deshidrogenasa, phosphoenolpiruvato carboxikinasa (GTP), fosfoglucomutasa y piruvato kinasa, entre otras.

Otra forma de generar energía en las plantas es mediante la degradación del almidón y sacarosa. El almidón se descompone en glucosa y fructosa, que pueden ingresar a la vía glucolítica (Fernie et al. 2002; Bahaji et al. 2014).

En resumen, se propone que las plantas con alta concentración de perezona podrían estar optimizando su metabolismo; regulando a la baja (subexpresando) las vías del metabolismo basal y regulando a la alza (sobreexpresando) algunas vías del metabolismo secundario, como es la vía de biosíntesis de terpenos.

Finalmente, se identificaron interacciones entre algunas proteínas con el análisis realizado en STRING, de estas proteínas, se indicó que AT1G11660, HSC70-1 y FPS1 están directamente relacionados. AT1G11660 y HSC70-1 representan proteínas involucradas en procesos que están relacionados con la respuesta celular al calor, la respuesta de defensa a bacterias, hongos y otros organismos; estos procesos están relacionados con la biosíntesis de metabolitos secundarios como es la biosíntesis de terpenos. Gupta y colaboradores en 2014 reportaron varios compuestos volátiles como el isopreno, 2-metil butanonitrilo, β ocimeno y sesquiterpenos producidos de forma diferencial por *Hymenaea courbaril* bajo estrés por calor. También Pazouki y colaboradores en 2016 demostraron que el estrés por calor resulta en una mejora importante de las emisiones de terpenoides en *Solanum lycopersicum*.

1.5. CONCLUSIONES.

Es importante destacar que el presente trabajo es el primer estudio proteómico que se reporta para plantas de *A. cordata* con distintos niveles de producción de perezona; todas las proteínas identificadas contribuyen a tener una visión general del perfil proteómico de la especie, concretamente de sus raíces.

En general la eficacia de la electroforesis bidimensional en términos del número de proteínas observadas fue siempre más baja comparada con las técnicas LC-MS/MS. En este estudio se lograron identificar aproximadamente 2112 proteínas mediante LCMS/MS, mientras que por electroforesis bidimensional, se identificaron cerca de 360 spots en geles bidimensionales; lo cual muestra que evidentemente la técnica LC-MS/MS resuelve de mejor manera el proteoma de *A.cordata*.

Por otro lado, de las proteínas identificadas por LC-MS/MS se identificaron 616 proteínas diferencialmente expresadas al comparar el grupo de alta producción contra el de baja producción de perezona, de las cuales 125 proteínas corresponden a proteínas sobreexpresadas y 491 corresponden a proteínas subexpresadas. Algunas de estas proteínas identificadas fueron clasificadas de acuerdo con su ontología génica en procesos biológicos, componente celular y función molecular. El análisis realizado con KEGG permitió clasificar las proteínas en varias rutas del metabolismo basal incluyendo glicólisis, gluconeogénesis, metabolismo de aminoácidos, de lípidos y otras vías metabólicas como la biosíntesis de

metabolitos secundarios, tales como: proteínas que participan en la ruta de biosíntesis de terpenos, sesquiterpenos, fenilpropanoides, flavonoides, etc.

La predicción realizada de las probables interacciones existentes entre algunas de las proteínas sobreexpresadas mostró una posible expresión de proteínas involucradas en la biosíntesis de compuestos como los sesquiterpenos junto con la expresión de proteínas relacionadas en procesos de estrés causado por calor.

Este trabajo contribuirá a la apertura de futuras líneas de investigación, especialmente en especies no-modelo que se contemplan también como especies de importancia en el grupo de plantas medicinales.

1.6. PERSPECTIVAS.

1. Iniciar la secuenciación del genoma de la especie *Acourtia cordata*.
2. Realizar un análisis comparativo del proteoma de raíces de plantúlas de *A. cordata* cultivadas *in vitro* en diferentes estadios de desarrollo.
3. Realizar un análisis comparativo del proteoma de raíces de plantúlas de *A. cordata* cultivadas *in vitro* vs las raíces de plantúlas aclimatadas a invernadero.
4. Establecer cultivos de células en suspensión de *A. cordata* productores de perezona y otros compuestos de interés y realizar un análisis proteómico de estos cultivos a lo largo de una cinética de crecimiento.
5. Establecer cultivos de raíces transformadas de *A.cordata* para la producción de perezona y otros compuestos de interés y realizar un análisis proteómico de las raíces transformadas elicitadas y sin elicitar.
6. Realizar un estudio del proteoma de las raíces de plantas de *A. cordata* cultivadas en sustratos con distintas concentraciones de macro y micronutrientes.
7. Realizar un análisis de RNAseq en plantas productoras y no productoras de perezona.

CAPÍTULO 2. CORRELACIÓN ENTRE FACTORES ECOLÓGICOS Y LA PRODUCCIÓN DE PEREZONA.

2.1. INTRODUCCIÓN.

En el capítulo dos de esta investigación se presentan los resultados de la influencia de factores edáficos y de medidas de talla sobre el rendimiento de la perezona en plantas silvestres de *A. cordata*.

Se sabe que los metabolitos secundarios (MS) también llamados productos naturales juegan un papel muy importante dentro del ciclo de vida de las plantas y que poseen funciones clave para la sobrevivencia de las plantas en su medio ambiente ayudándoles de alguna manera a contender contra factores bióticos y abióticos del medio ambiente en el que se desarrollan, además de que les sirven para establecer ciertas relaciones ecológicas con otras plantas y organismos (Musilova., et al, 2016).

Los metabolitos secundarios que se sintetizan y acumulan en plantas de la misma especie pueden variar en su rendimiento dependiendo de las diferencias ambientales del lugar en el cual crecen y se desarrollan, estos factores ecológicos incluyen la altitud, latitud, precipitación, temperatura, humedad, composición del suelo, etc. Respecto a la composición del suelo, parámetros tales como la materia orgánica, el pH y contenido de algunos nutrientes como el Nitrógeno (N) y Fosforo (P) destacan por su significativa correlación con el contenido de ciertos productos naturales (Liu, W. et al, 2016).

En ese sentido, los resultados que se presentan en este capítulo se relacionan con la influencia que podría tener la composición del suelo sobre la producción de los metabolitos secundarios específicamente de la perezona, compuesto que se acumula en las raíces de plantas de *A. cordata* que crecen en 3 localidades diferentes dentro del estado de Morelos.

2.2. MATERIAL Y MÉTODOS.

2.2.1 Análisis edáfico.

Para caracterizar el suelo, se obtuvieron 60 muestras de suelo pertenecientes a tres localidades (20/sitio) y se recogieron sistemáticamente cuatro muestras de suelo a lo largo de un transecto de 40 m (recolectamos una muestra cada 10 m). Para el análisis del suelo, se pesaron 200 g de suelo seco que previamente se tamizó con una malla de 2 mm. Para cada muestra, se cuantificaron los siguientes parámetros: pH; contenido de materia orgánica (M.O.), medido con un analizador de carbono orgánico total (el CO₂ emitido durante la combustión se midió mediante detección infrarroja no dispersiva); contenido de amonio (NH₄), cuantificado después de una extracción con KCl; contenido de nitrato (NO₃),

cuantificado mediante electroforesis capilar de iones; contenido de fosfato (PO_4); nitrógeno total (N_t ; siguiendo el método de Bremner (1965)) y fósforo total (P_t , medido a través del método de Bray y Kurtz).

2.2.2. Análisis de datos.

Con el fin de determinar el efecto del sitio de recolección (localidad) en los caracteres macromorfológicos medidos en *A. cordata* y los parámetros del suelo, realizamos el Análisis de Varianza (ANOVA). Todos los datos discontinuos se transformaron como $[(x) \frac{1}{2} + 0.5]$ (Zar, 2010). Además, se realizaron análisis de regresión simples entre la producción de perezona y los parámetros del suelo y los caracteres macromorfológicos medidos.

Utilizamos un enfoque de regresión múltiple para examinar si los parámetros del suelo (pH, M.O., NH_4 , NO_3 , PO_4 , N_t , P_t) y macromorfológicos (diámetro basal, número de ramas, altura individual total y cubierta de arbustos) medidas que probablemente influyen en el contenido de la perezona. Específicamente, este análisis fue útil para determinar la contribución relativa de cada factor en la producción de perezona en *A. cordata*. Usamos un modelo de mínimos cuadrados estándar con estructura de error parcial (sumas de cuadrados tipo III) de pH, M.O., NH_4 , NO_3 , PO_4 , N_t , P_t , diámetro basal, número de ramas, altura individual total y cubierta de arbustos como nuestros factores. El software utilizado para el análisis estadístico fue STATISTICA 8.0 (Statsoft, 2007).

2.3. RESULTADOS.

2.3.1 Correlación entre factores edáficos y la producción de perezona.

Nuestros resultados mostraron que la localidad no tuvo un efecto significativo en ninguno de los parámetros de suelo analizados, excepto para N_t y P_t . De manera similar, la localidad no tuvo un efecto significativo en ningún carácter macromorfológico, excepto en la cobertura de arbustos (tabla 5).

En general, el análisis de regresión múltiple mostró que P_t , M.O. y N_t , tuvo una influencia positiva y estadísticamente significativa en los contenidos de perezona, explicando el 27.9%, 9.4% y 4.4% de su variación, respectivamente.

En contraste, no hubo relación entre los contenidos de perezona y los caracteres morfológicos de la planta (tabla 6).

Tabla 5 Media \pm error estándar y resultados de ANOVA para cada parámetro de suelo y para cada carácter macromorfológico medido en tres localidades del corredor biológico "Chichinautzin", Morelos, México.

Variable	Localidad			$F_{2,57}$	Unidades
	Chamilpa (media \pm s.e.)	Alarcón (media \pm s.e.)	Felipe Neri (media \pm s.e.)		
Parametros de suelo					
pH	6.74 \pm 0.09	6.60 \pm 0.08	6.50 \pm 0.09	1.880	n.s.
O.M.	6.60 \pm 0.51	5.83 \pm 0.68	8.31 \pm 1.18	2.240	n.s.
NH ₄	16.06 \pm 0.84	14.57 \pm 2.26	19.34 \pm 2.06	1.763	n.s.
NO ₃	9.46 \pm 0.50	9.07 \pm 0.55	11.51 \pm 0.43	6.870	**
PO ₄	4.00 \pm 0.41	3.43 \pm 0.42	2.14 \pm 0.23	6.623	**
N _t	0.41 \pm 0.03	0.46 \pm 0.06	0.44 \pm 0.05	0.290	n.s.
P _t	0.08 \pm 0.01	0.07 \pm 0.01	0.11 \pm 0.02	2.587	n.s.
Características morfológicas					
Diametro basal	24.56 \pm 3.61	18.04 \pm 2.25	17.36 \pm 2.11	2.097	n.s.
Número de ramas	3.30 \pm 4.17	2.25 \pm 0.27	2.80 \pm 0.28	2.550	n.s.
Altura promedio	84.20 \pm 7.86	89.05 \pm 9.93	104.00 \pm 6.21	1.605	n.s.
Cobertura	8141.29 \pm 903.23	9070.02 \pm 3104.76	17096.69 \pm 3304.76	4.280	*

n.s.= diferencias no significativas, * = $p < 0.05$; ** = $p < 0.01$

Tabla 6. Análisis de regresión múltiple que prueban la influencia de diferentes parámetros del suelo y caracteres macromorfológicos en los contenidos de perezona en *Acourtia cordata* en tres localidades del corredor biológico el "Chichinautzin", Morelos, México.

	SS	$F_{1,48}$	p	% variación
Parametros de suelo				
pH	1.61	0.91	n.s.	
M.O..	18.79	10.60	**	9.4
NH ₄	1.26	0.71	n.s.	
NO ₃	1.53	0.86	n.s.	
PO ₄	0.95	0.54	n.s.	
N _t	7.13	4.02	*	4.4
P _t	28.71	16.19	***	27.9
Caracteres morfológicos				
Diametro basal	1.00	0.56	n.s.	
Número de ramas	0.30	0.17	n.s.	
Altura promedio	0.64	0.36	n.s.	
Cobertura	0.29	0.16	n.s.	

n.s.= diferencias no significativas, * = $p < 0.05$; ** = $p < 0.01$, ***= $p < 0.001$

2.4. DISCUSIÓN.

2.4.1. Influencia de factores edáficos sobre la producción de perezona.

La cuantificación de la producción de perezona en las raíces de las plantas silvestres mostró que el rendimiento fue constante en las tres localidades analizadas, este resultado probablemente se deba al hecho de que los tres sitios de estudio están ubicados en la misma área natural con la misma historia geológica, con similitudes de condiciones climáticas y altitudinales, además de estar asociado con una vegetación dominada por bosques de *Pinus-Quercus* (Jaimes-Viera et al. 2018). Otra posibilidad es que los individuos asociados con los tres sitios de estudio correspondan a la misma población de *A.cordata*. Por lo tanto, se sugiere para futuras investigaciones llevar a cabo un estudio sobre la estructura y la diversidad genética que aclare si los individuos de los tres sitios de estudio son la misma población.

No obstante, se observó que entre individuos existe una gran variación en el rendimiento del compuesto y que también existe una correlación positiva con el contenido de materia orgánica, nitrógeno y fósforo total. Blanch y colaboradores (2007), reportaron que el nitrógeno podría promover las emisiones de terpenoides al promover la tasa de transporte de electrones y la fotosíntesis de las hojas que proporcionan los requerimientos de ATP y la disponibilidad de sustratos de carbono para la síntesis de isopreno. Además, es de esperarse que el nitrógeno favorezca la producción de terpenoides, especialmente las emisiones de mono y sesquiterpenoides en plantas, lo cual ya ha sido reportado previamente (Muzika et al. 1989). Además, algunos estudios describen que los macronutrientes como el fósforo influyen en la producción de terpenoides ya que los precursores de terpenoides IPP: isopentenildifosfato y DMAPP: dimetilalil difosfato, contienen enlaces de fosfato de alta energía y además el fósforo es un componente clave del ATP y NADPH que se requieren para síntesis de terpenoides (Niinemets et al. 2002).

Por otra parte, Liu y colaboradores mencionan que la acumulación de MS está fuertemente afectada por factores ambientales, por lo que las plantas regulan el tipo y la cantidad de MS de acuerdo con las variaciones ambientales. Estos autores concluyen que el contenido de materia orgánica, el pH del suelo y el potasio disponible fueron los factores limitantes para la producción de flavonoides en la especie *Sinopodophyllum hexandrum*; específicamente, la materia orgánica fue el factor limitante más importante, con una correlación positiva y significativa con los contenidos del metabolito activo (Liu et al. 2015).

2.5. CONCLUSIÓN.

Nuestros resultados sugieren que algunos de los componentes del suelo como el Nitrógeno, Fosforo y Materia orgánica (N_t, P_t y M.O) pueden ser factores importantes relacionados con el contenido de perezona en *A. cordata*. Además, queda claro que los caracteres morfológicos relacionados con talla, altura, cobertura o el número de ramas no son determinantes en la producción de perezona.

2.6. PERSPECTIVAS.

1. Realizar un estudio sobre la estructura y la diversidad genética de las plantas de los tres sitios de estudio y otros sitios en donde se encuentren poblaciones de la misma especie.
2. Establecer un cultivo controlado de plantas de *A. cordata* modificando las concentraciones de Nitrogeno, Fosforo y Materia organica y analizar la producción de perezona y otros terpenos.
3. Realizar un análisis de la producción de perezona temporalmente en plantas silvestres.
4. Analizar la influencia de algunos micronutrientes que podrían tener efecto en la producción del compuesto de interés.

REFERENCIAS.

1. Abbas Farhat., Ke Yanguo., Yu Rangcai., Yue Yuechong., Amanullah Sikandar., Jahangir Muhammad Muzammil., Fan Yanping. (2017). Volatile terpenoids: multiple functions, biosynthesis, modulation and manipulation by genetic engineering. *Planta*. 246: 803. <https://doi.org/10.1007/s00425-017-2749-x>
2. Aghaei, K. and Komatsu, S (2013) Crop and medicinal plants proteomics in response to salt stress. *Front. Plant Sci.* 4:1–9.
3. Amme S, Matros A, Schlesier B, Mock HP. 2006. Proteome analysis of cold stress response in *Arabidopsis thaliana* using DIGE-technology. *J. Exp. Bot.* 57:1537-1546.
4. Andrew, R.L., Peakall, R., Wallis, I.R., and Foley, W.J (2007) Spatial Distribution of Defense Chemicals and Markers and the Maintenance of Chemical Variation. *Ecology*, 88:716–728.
5. Araújo, W.L., Tohge, T., Ishizaki, K., Leaver, C.J., Fernie, A.R. (2011) Protein degradation – an alternative respiratory substrate for stressed plants. *Trends Plant Sci* 16:489–498.
6. Arellano, J., Vfizquez, F., Villegas, T., and Hernfindez, G (1996) Establishment of transformed root cultures of *Perezia cuernavacana* producing the sesquiterpene quinone perezone. *Plant Cell Reports* 15:455–458.
7. Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A., Hartmann, E.M (2014) Non-model organisms, a species endangered by proteogenomics. *J. Proteomics* 105:5–18.
8. Ali JG, Alborn HT, Campos-Herrera R, Kaplan F, Duncan LW, Rodriguez-Saona C, Koppenhöfer AM, Stelinski LL (2012) Subterranean, herbivore-induced plant volatile increases biological control activity of multiple beneficial nematode species in distinct habitats. *PLoS One* 7:e38146
9. Bachi A, Bonaldi T. 2008. Quantitative proteomics as a new piece of the systems biology puzzle. *J. Proteomics* 71:357.
10. Baginsky, S (2009) Plant proteomics: Concepts, applications, and novel strategies for data interpretation. *Mass Spectrom. Rev* 1:93-120.
11. Bahaji, A., Li, J., Sánchez-López, Á.M., Baroja-Fernández, E., Muñoz, F.J., Ovecka, M., Almagro, G., Montero, M., Ezquer, I., Etxeberria, E., et al (2014) Starch biosynthesis, its regulation and biotechnological approaches to improve crop yields. *Biotechnol. Adv* 32:87–106.
12. Bayer, E.M., Bottrill, A.R., Walshaw, J., Vigouroux, M., Naldrett, M.J., Thomas, C.L., Maule, A.J (2006) *Arabidopsis* cell wall proteome defined using multidimensional protein identification technology. *Proteomics* 6:301–311.
13. Bhattacharyya, D., Sinha, R., Ghanta, S., Chakraborty, A., Hazra, S., Chattopadhyay, S (2012) Proteins differentially expressed in elicited cell suspension culture of *Podophyllum hexandrum* with enhanced podophyllotoxin content. *Proteome Sci* 10:1.
14. Biemann K. 1990. Sequencing of peptides by tandem mass spectrometry and high energy collision- induced dissociation. *Methods Enzymol* 193:455-479.
15. Blanch, J.-S., Peñuelas, J., Llusà, J (2007) Sensitivity of terpene emissions to drought and fertilization in terpene-storing *Pinus halepensis* and non-storing *Quercus ilex*. *Physiol. Plant* 131: 211–225.
16. Bohler S, Bagard M, Oufir M, Planchon S, Hoffman L, Jolivet Y, Hausman JF, Dizengremel P, Renaut J. 2007. A DIGE analysis of developing poplar leaves subjected to ozone reveals major changes in carbon metabolism. *Proteomics* 7:1584-1599.
17. Borchers CH, Thapar R, Petrotchenko EV, Torres MP, Speir JP, Easterling M, Dominski Z, Marzluff WF. 2006. Combined top-down and bottom-up proteomics identifies a phosphorylation site in stem-loop-binding proteins that contributes to high-affinity RNA binding. *P .Natl. Acad. Sci. USA* 103 (9)3094-3099.
18. Bray, R.H and Kurtz, L.T (1945) Determination of total, organic, and available forms of phosphorus in soils. *Soil Sci* 59:39.
19. Bremner JM (1965) Inorganic forms of nitrogen. In: Black CA (ed) *Methods of soil analysis*. American Society of Agronomy. Madison, Wisconsin
20. Bryant, L., Flatley, B., Patole, C., Brown, G.D., and Cramer, R (2015) Proteomic analysis of *Artemisia annua* - towards elucidating the biosynthetic pathways of the antimalarial pro-drug artemisinin. *BMC Plant Biology*, 15:1–13.
21. Bryant J.P., Chapin III F.S., Klein D.R.. (1983). Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory. *Oikos*, 40:357-368.
22. Burgueño-Tapia, E., Castillo, L., González-Coloma, A., and Joseph-Nathan, P (2008) Antifeedant and phytotoxic activity of the sesquiterpene p-benzoquinone perezone and some of its derivatives. *Journal of Chemical Ecology*, 34:766–771.
23. Cabrera RL (2001) Six new species of *Acourtia* (Asteraceae) and a historical account of *Acourtia mexicana*. *Brittonia* 53:416–429.

24. Carpentier S.C., America T. (2014) Proteome Analysis of Orphan Plant Species, Fact or Fiction?. In: Jorriño Novo J., Komatsu S., Weckwerth W., Wienkoop S. (eds) Plant Proteomics. Methods in Molecular Biology (Methods and Protocols), vol 1072. Humana Press, Totowa, NJ.
25. Casati P, Zhang X, Burlingame AL, Walbot V. 2005. Analysis of leaf proteome after UV-B irradiation in maize lines differing in sensitivity. *Mol. Cell. Proteomics* 4:1673-1685.
26. Chait BT. 2006. Chemistry. Mass spectrometry: bottom-up or top-down? *Science* 314:65-66.
27. Champagne, A., Rischer, H., Oksman-Caldentey, K.M., Boutry, M (2012) In-depth proteome mining of cultured *Catharanthus roseus* cells identifies candidate proteins involved in the synthesis and transport of secondary metabolites. *Proteomics* 12:3536–3547.
28. Chappell, J. (2002) The genetics and molecular genetics of terpene and sterol origami. *Curr Opin Plant Biol*, 5, 151-157.
29. Chen J, Shiyonov P, Zhang L, Schlager JJ, Kari B. 2010. Top-down characterization of a native highly intralinked protein: concurrent cleavages of disulfide and protein backbone bonds. *Anal. Chem.* 82:6079-6089.
30. Chen, F., Ro, D.K., Petri, J., Gershenzon, J., Bohlmann, J., Pichersky, E. and Tholl, D. (2004) Characterization of a root-specific *Arabidopsis* terpene synthase responsible for the formation of the volatile monoterpene 1,8-cineole. *Plant Physiol.* 135, 1956–1966.
31. Chmielewska, K., Rodziewicz, P., Swarczewicz, B., Sawikowska, A., Krajewski, P., Marczak, Ł., Ciesiołka, D., Kuczyńska, A., Mikołajczak, K., Ogrodowicz, P., et al. (2016) Analysis of Drought-Induced Proteomic and Metabolomic Changes in Barley (*Hordeum vulgare* L.) Leaves and Roots Unravels Some Aspects of Biochemical Mechanisms Involved in Drought Tolerance. *Front. Plant Sci.* 7:1–14.
32. Closa M, Vranová E, Bortolotti C, Bigler L, Arró M, et al. (2010) The *Arabidopsis thaliana* FPP synthase isozymes have overlapping and specific functions in isoprenoid biosynthesis, and complete loss of FPP synthase activity causes early developmental arrest. *Plant J* 63: 512–525.
33. Conesa, A, Götz, S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832.
34. Coon JJ, Ueberheide B, Syka JE, Dryhurst DD, Ausio J, et al. 2005. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc. Natl. Acad. Sci. USA* 102:9463-68.
35. Corchete, P., Bru, R. (2013) Proteome alterations monitored by DIGE analysis in *Silybum marianum* cell cultures elicited with methyl jasmonate and methyl B cyclodextrin. *Journal of Proteomics.* 8599:108.
36. Craig R, Beavis RC. 2004. Tandem: matching proteins with tandem mass spectra. *Bioinformatics* 20:1466-1467.
37. Croteau, R., Kutchan, T.M., and Lewis, N.G. 2000. Natural products (Secondary metabolites). pp. 1250-1318. In: B. Buchanan., W. Gruissem., and R. Jones R. (eds.). *Biochemistry and Molecular Biology of Plants*. Vol. 24. American Society of Plant Physiologists. Maryland, USA. 1367 p
38. Cunillera N, Arro M, Delourme D, Karst F, Boronat A, Ferrer A. *Arabidopsis thaliana* contains two differentially expressed farnesyl-diphosphate synthase genes. *J Biol Chem.* 1996;271:7774–7780.
39. Cunillera N, Boronat A, Ferrer A. The *Arabidopsis thaliana* FPS1 gene generates a novel mRNA that encodes a mitochondrial farnesyl-diphosphate isoform. *J Biol Chem.* 1997;272:15381–15388.
40. Degenhardt J, Köllner TG, Gershenzon J (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70:1621–1637. doi: 10.1016/j.phytochem.2009.07.030
41. Degenhardt, J., Köllner, T.G. and Gershenzon, J. (2009a) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*, 70, 1621–1637.
42. Delory BM, Delaplace P, Fauconnier ML, Du Jardin P (2016) Root-emitted volatile organic compounds: can they mediate belowground plant–plant interactions? *Plant Soil* 402:1–26.
43. De Oliveira JM and de Graaff LH. Proteomics of industrial fungi: trends and insights for biotechnology. *Appl Microbiol Biotechnol.* 2011; 89: 225-37.
44. Di Michele, M., Chiatante, D., Plomion, C., and Scippa, G.S (2006) A proteomic analysis of Spanish broom (*Spartium junceum* L.) root growing on a slope condition. *Plant Science*, 170:926–935.
45. Dixon, R.A., Paiva, N.L., 1995. Stress-induced phenylpropanoid metabolism. *The Plant Cell* 7, 1085–1097.
46. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* 312, 212-217 (2006).
47. Du Y, Parks BA, Sohn S, Kwast KE, Kelleher NL. 2006. Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. *Anal. Chem.* 78:686- 94.
48. Dudareva, N., Pichersky, E., Gershenzon, J (2004) *Biochemistry of Plant Volatiles* 1. *Plant Physiol* 135:1893–1902.
49. Elias JE, Gygi SP. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat methods* 4:207-214.

50. Ee, S. F., Mohamed-Hussein, Z. A., Othman, R., Shaharuddin, N. A., Ismail, I., & Zainal, Z. (2014). Functional characterization of sesquiterpene synthase from *Polygonum minus*. *TheScientificWorldJournal*, 2014, 840592. doi:10.1155/2014/840592
51. Fang, X., Chen, J., Dai, L., Ma, H., Zhang, H., Yang, J., Wang, F., Yan, C (2015) Proteomic dissection of plant responses to various pathogens. *Proteomics*. 15:1525–1543.
52. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64-71 (1989).
53. Fernandez-García, N., Hernandez, M., Casado-Vela, J., Bru, R., Elortza, F., Hedden, P., Olmos, E (2011) Changes to the proteome and targeted metabolites of xylem sap in *Brassica oleracea* in response to salt stress. *Plant Cell and Environment*. 34:821 - 836.
54. Fernie, A. (2002) Sucrose to starch: a transition in molecular plant physiology. *Trends Plant Sci* 7:35–41.
55. Flores y Troncoso, F. de A (1982) *Historia de la medicina en México desde la época de los Indios hasta la presente*. 2a. ed. Instituto Mexicano del Seguro Social, Facsimile 3, México.
56. Fraser, P.D. and Bramley, P.M. (2004) The biosynthesis and nutritional uses of carotenoids. *Prog Lipid Res*, 43, 228-265.
57. Frewen B, MacCoss MJ. 2007. Using BiblioSpec for creating and searching tandem MS peptide libraries. *Curr. Protoc. Bioinformatics*. Chapter 13:Unit 13.7.
58. García de la Garma, J., M Fernandez-García, N., Bardisi, E., Pallol, B., Asensio-Rubio, J.S., Bru, R., Olmos, E (2015) New insights into plant salt acclimation: the roles of vesicle trafficking and reactive oxygen species signalling in mitochondria and the endomembrane system. *New Phytologist*. 205, pp. 216 - 239.
59. Geer LY, Markey SP, Kowalak JA, et al. 2004. Open mass spectrometry search algorithm. *J. Proteome Res*. 3:958-964. Gegenheimer P. 1990. Preparation of extracts from plants. *Meth. Enzymol*. 182:174-193.
60. Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, Kohlbacher O, Neuweger H, Schneider R, Tenenbaum D, Gavin AC. 2010. Visualization of omics data for systems biology. *Nat. methods* 7 (3) 56-S68. 7–9.
61. Gharechahi, J., Khalili, M., Hasanloo, T., Salekdeh, G.H (2013) An integrated proteomic approach to decipher the effect of methyl jasmonate elicitation on the proteome of *Silybum marianum* L. hairy roots. *Plant Physiol. Biochem* 70:115–122.
62. Ghosh, D. and Xu, J (2014) Abiotic stress responses in plant roots: a proteomics perspective. *Front. Plant Sci*. 5:1–13.
63. Gómez-Serrano G, Cristiani-Urbina E, Villegas-Garrido TL (2012) Time-dependent perezone production in different culture systems of *Acourtia cordata*. *Central European Journal of Biology* 7:507–518. doi: 10.2478/s11535-012-0035-2
64. Gómez-Serrano, G., Cristiani-Urbina, E., and Villegas-Garrido, T.L (2010) Establecimiento de protocolos para la propagación in vitro de plantas de *Acourtia cordata* (Cerv.) Turner (Compositae), colectadas en la Sierra de Guadalupe. *Polibotánica*, 30:89–110.
65. Görg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R, Weiss W 2000: The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 21:1037-1053.
66. Gupta D, Eldakak M, Rohila JS, Basu C (2014) Biochemical analysis of 'kerosene tree' *Hymenaea courbaril* L. under heat stress. *Plant Signal Behav* 9:. doi: 10.4161/15592316.2014.972851
67. Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. *Nat Chem Biol* 3:408–414
68. Guo, D., Li, H. L., & Peng, S. Q. (2015). Structure Conservation and Differential Expression of Farnesyl Diphosphate Synthase Genes in Euphorbiaceous Plants. *International journal of molecular sciences*, 16(9), 22402–22414. doi:10.3390/ijms160922402
69. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. 2000. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *P. Natl. Acad. Sci*. 97:9390-9395.
70. Gregorich, Z. R., Chang, Y. H., & Ge, Y. (2014). Proteomics in heart failure: top-down or bottom-up?. *Pflugers Archiv : European journal of physiology*, 466(6), 1199–1209. doi:10.1007/s00424-014-1471-9
71. Hajduch M, Casteel JE, Tang S, Hearne LB, Knapp S, Thelen JJ. 2007. Proteomic analysis of near isogenic sunflower varieties differing in seed oil traits. *J. Proteome Res*. 6:3232-3241.
72. Hamilton, J.G., Zangerl, A.R., DeLucia, E.H., and Berenbaum, M.R (2001) The carbon–nutrient balance hypothesis: its rise and fall. *Ecology Letters*, 4:86–95.
73. Harry JL, Wilkins MR, Herbert BR, Packer NH, Gooley AA, Williams KL. 2000 Proteomics: capacity versus utility. *Electrophoresis* 21:1071-1081.
74. Herms D.A., Mattson W.J. (1992). The dilemma of plants: to grow or defend. *The Quarterly Review of Biology*, 67:282-335

75. Hirschberg, J. (2001) Carotenoid biosynthesis in flowering plants. *Curr Opin Plant Biol*, 4, 210- 218.
76. Huang M, Sanchez-Moreiras AM, Abel C, Sohrabi R, Lee S, Gershenzon J, Tholl D (2012) The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene (E)- β -caryophyllene, is a defense against a bacterial pathogen. *New Phytol* 193:997–1008
77. Hjerno K, Alm R, Canback B, Matthiesen R, Trajkovski K, Bjork L, Roepstorff P, Emanuelsson C. 2006. Down-regulation of the strawberry Bet v 1-homologous allergen in concert with the flavonoid biosynthesis pathway in colorless strawberry mutant. *Proteomics* 6:1574-1587.
78. Hoedt E., Zhang G., Neubert T.A. (2019) Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) for Quantitative Proteomics. In: Woods A., Darie C. (eds) *Advancements of Mass Spectrometry in Biomedical Research. Advances in Experimental Medicine and Biology*, vol 1140. Springer, Cham.
79. Hol apek M, Jirásko R and Lída M. Recent developments in liquid chromatography-mass spectrometry and related techniques. *Journal of Chromatography A*. 2012; 1259: 3-15.
80. Ibdah M., Muchlinski A., Yahyaa M., Nawade B., Tholl D. (2019) Carrot Volatile Terpene Metabolism: Terpene Diversity and Biosynthetic Genes. In: Simon P., Iorizzo M., Grzebelus D., Baranski R. (eds) *The Carrot Genome. Compendium of Plant Genomes*. Springer, Cham.
81. Jaimes-Viera, M.C., Martin Del Pozzo, A.L., Layer, P.W., Benowitz, J.A., Nieto-Torres, A (2018) Timing the evolution of a monogenetic volcanic field: Sierra Chichinautzin, Central Mexico. *J. Volcanol. Geotherm. Res* 356:225–242.
82. Joseph-Nathan P, V. Mendoza, and E. García (1989) The chemistry of perezone and its consequences, in *Studies in Natural Products Chemistry*, vol. 5, Structural Elucidation (Part B), Ed. Att-ur-Rahman, Elsevier Science Publishers, Amsterdam, pp. 763–813
83. Joseph-Nathan, P., and Santillan, R. L. (1989). The chemistry of perezone and its consequences. *Studies in Natural Products Chemistry*, 5:763-813.
84. Joseph-Nathan, P., González, M.P., and Rodríguez, V.M (1972) Terpenoids of *Perezia hebeclada*. *Phytochemistry*, 11:1803–1808.
85. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K (2017) KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45:D353–D361.
86. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* 60, 2299-2301 (1988).
87. Keim, V., Manzano, D., Fernández, F. J., Closa, M., Andrade, P., Caudepón, D., Ferrer, A. (2012). Characterization of *Arabidopsis* FPS isozymes and FPS gene expression analysis provide insight into the biosynthesis of isoprenoid precursors in seeds. *PloS one*, 7(11), e49109. doi:10.1371/journal.pone.0049109
88. Kelleher NL, Lin HY, Valaskovic GA, Aaserud DJ, Fridriksson EK, McLafferty FW. 1999. Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.* 121:806-12.
89. Kettman JR, Frey JR, Lefkovits I. 2001. Proteome, transcriptome and genome: top down or bottom up analysis? *Biomol. Eng.* 18 (5):207-212.
90. Kim, J. y DellaPenna, D. (2006) Defining the primary route for lutein synthesis in plants: the role of *Arabidopsis* carotenoid beta-ring hydroxylase CYP97A3. *Proc Natl Acad Sci U S A*, 103, 3474-3479.
91. Kim, S.J., Kim, M.D., Choi, J.H., Kim, S.Y., Ryu, Y.W. and Seo, J.H. (2006) Amplification of 1-deoxy-D-xylulose 5-phosphate (DXP) synthase level increases coenzyme Q10 production in recombinant *Escherichia coli*. *Appl Microbiol Biotechnol*, 72, 982-985.
92. Koricheva, J., and Barton, K. (2012). Temporal changes in plant secondary metabolite production. In G. Iason, M. Dicke, & S. Hartley (Eds.), *The Ecology of Plant Secondary Metabolites: From Genes to Global Processes (Ecological Reviews*, pp. 34-55). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511675751.004
93. Koricheva, J., Larsson, S., Haukioja, E., Keinanem, M., 1998. Regulation of woody plant secondary metabolism by resource availability: hypothesis testing by means of meta-analysis. *Oikos* 83, 212–226.
94. Kouki, M., Manetas, Y., 2002. Resource availability affects differentially the levels of gallotannins and condensed tannins in *Ceratonia siliqua*. *Biochemical Systematics and Ecology* 30, 631–639.
95. Link AJ, Eng J, Chiertz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR. 1999. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol.* 17:676-682.
96. Liu, W., Liu, J., Yin, D., and Zhao, X (2015) Influence of ecological factors on the production of active substances in the anti-cancer plant *Sinopodophyllum hexandrum* (Royle) T.S. Ying. *PLoS ONE*, 10, 1–22.
97. Lombard, J., Moreira, D (2011) Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol. Biol. Evol* 28:87–99.

98. Lopez MF, Berggren K, Chernokalskaya E, Lazarev A, Robinson M, Patton WF. 2000. A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* 21:3673-3683.
99. Lodish H, Berk A, Zipursky SL, et al. *Molecular Cell Biology*. 4th edition. New York: W. H. Freeman; 2000.
100. Ma, R., Sun, L., Chen, X., Jiang, R., Sun, H., Zhao, D (2013) Proteomic changes in different growth periods of ginseng roots. *Plant Physiol. Biochem* 67, 20–32.
101. Martínez-Esteso, M.J., Casado-Vela, J., Sellés-Marchart, S., Pedreño, M.A., Bru-Martínez, R (2014) Differential plant proteome analysis by isobaric tags for relative and absolute quantitation (iTRAQ). In *Plant Proteomics*; Humana Press: New York, NY, USA, pp. 155–169
102. Martínez-Esteso, M.J., Sellés-Marchart, S., Vera-Urbina, J.C., Pedreño, M.A., Bru-Martínez, R (2011) DIGE analysis of proteome changes accompanying resveratrol production by grapevine (*Vitis vinifera* cv. Gamay) cell cultures in response to cyclodextrin and methyl jasmonate elicitors. *Journal of Proteomics*. 74:1421 - 1436.
103. Martínez-Esteso, M.J., Vilella-Anton, M.T., Selles-Marchart, S., Martínez-Márquez, A., Botta-Catala, A., Piñol-Dastis, R., Bru-Martínez, R.A (2016) A DIGE proteomic analysis of wheat flag leaf treated with TERRA-SORB® foliar, a free amino acid high content biostimulant. *J. Integr. Omics* 6:9–17
104. Martínez-Esteso, M.J.; Martínez-Márquez, A.; Sellés-Marchart, S; Morante-Carriel, J.A.; Bru-Martínez, R. (2015) The role of proteomics in progressing insights into plant secondary metabolism. *Frontiers in Plant Science*. 6:504 doi:10.3389/fpls.2015.00504
105. Martínez-Márquez, A., Morante-Carriel, J., Sellés-Marchart, S., Martínez-Esteso, M.J., Pineda-Lucas, J.L., Luque, I., Bru R (2013) Development and validation of MRM methods to quantify protein isoforms of polyphenol oxidase in loquat fruits. *Journal of Proteome Research*, 12, 5709 - 5722.
106. McDonald WH, Yates JR 3rd. 2002. Shotgun proteomics and biomarker discovery. *Dis. Markers* 18:99- 105.
107. McDonald WH, Yates JR 3rd. 2003. Shotgun proteomics: integrating technologies to answer biological questions. *Curr. Opin. Mol. Ther* 5:302-9.
108. Meleady P, Gallagher M, Clarke C, et al (2012) Impact of miR-7 over-expression on the proteome of Chinese hamster ovary cells. *Journal of Biotechnology* 160:251–262. doi: 10.1016/j.jbiotec.2012.03.002
109. Mohamad ZA, Chokchaichamnankit D, Bhinija K, et al (2011) Proteomic analysis of Chinese kale (*B. alboglabra*) leaves during growth. *Journal of Integrated OMICS* 1:102–107. doi: 10.5584/jiomics.v1i1.25
110. Monteoliva, L. and Albar, J. P. (2004), Differential proteomics: an overview of gel and non-gel based approaches., *Briefings in functional genomics & proteomics*, 3(3), 220-39.
111. Mooney BP, Krishnan HB, Thelen JJ. 2004. High-throughput peptide mass fingerprinting of soybean seed proteins: automated workflow and utility of Unigene expressed sequence tag databases for protein identification. *Phytochemistry* 65:1733-1744.
112. Moore, B.D. and Foley, W.J (2005) Tree use by koalas in a chemically complex landscape. *Nature*, 435, Muzika, R. M., Pregitzer, K. S., Hanover, J. W. (1989) Changes in terpene production following nitrogen fertilization of grand fir (*Abies grandis* (Dougl.) (Lindl) seedlings. *Oecologia* 80:485.
113. Nagegowda DA (2010) Plant volatile terpenoid metabolism: Biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Letters* 584:2965–2973. doi: 10.1016/j.febslet.2010.05.045
114. Nepusz, T., Yu, H., Paccanaro, A (2012) Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods* 9:471–472.
115. Neuhoff V, Arold N, Taube D, Ehrhardt W. 1988. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* 9:255-262.
116. Nahnsen S, Bielow C, Reinert K, et al. Tools for label-free peptide quantification. *Mol Cell Proteomics*. 2013;12: 549-56.
117. Niinemets, Ü., Seufert, G., Steinbrecher, R., Tenhunen, J.D (2002) A model coupling foliar monoterpene emissions to leaf photosynthetic characteristics in Mediterranean evergreen *Quercus* species. *New Phytol* 153:257–275.
118. O'Farrell PH. 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem* 250:4007-4021.
119. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, et al. 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* 4:1487- 502.
120. Oldham, J.T., Hincapie, M., Rejtar, T., Wall, P.K., Carlson, J.E., Lee-Parsons, C.W.T (2010) Shotgun proteomic analysis of yeast-elicited California poppy (*Eschscholzia californica*) suspension cultures producing enhanced levels of benzophenanthridine alkaloids. *J. Proteome Res* 9:4337–4345.

121. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. (2007). Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 4:709-12.
122. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1:376-86.
123. Ong SE, Kratchmarova I, Mann M. (2003). Properties of ¹³C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res* 2:173-81.
124. Ong S-E, Mann M. (2005). Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* 1:252-62.
125. Ormeño, E. and Fernandez, C. (2012) Effect of Soil Nutrient on Production and Diversity of Volatile Terpenoids from Plants. *Curr Bioact Compd* 8:71–79.
126. Pazouki, L., Kanagendran, A., Li, S., Kännaste, A., Memari, H.R., Bichele, R., Niinemets, Ü (2016) Mono- and sesquiterpene release from tomato (*Solanum lycopersicum*) leaves upon mild and severe heat stress and through recovery: from gene expression to emission responses. *Environ. Exp. Bot* 132:1–15.
127. Perkins DN, Pappin DJC, Creas DM, Cottrell JC. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551-3567.
128. Pichersky, E., Raguso, R. A., Lewinsohn, E., & Croteau, R. (1994). Floral Scent Production in Clarkia (*Onagraceae*) (I. Localization and Developmental Modulation of Monoterpene Emission and Linalool Synthase Activity). *Plant physiology*, 106(4), 1533–1540.
129. Picotti P, Aebersold R, Domon B. (2007). The implications of proteolytic background for shotgun proteomics. *Mol. Cell. Proteomics* 6(9):1589-98.
130. Pinheiro, C. and Chaves, M.M (2011) Photosynthesis and drought: can we make metabolic connections from available data? *J Exp Bot* 62:869–882.
131. Porteus B, Kocharunchitt C, Nilsson RE, et al. Utility of gel-free, label-free shotgun proteomics approaches to investigate microorganisms. *Appl Microbiol Biotechnol.* 2011; 90: 407-16.
132. Pradhan, J., Sahoo, S.K., Lalotra, S., and Sarma, R.S (2017) Positive impact of abiotic stress on medicinal and aromatic plants. *International Journal of Plant Sciences (Muzaffarnagar)* 12:309–313.
133. Rabilloud T, Adessi C, Giraudel A, Lunardi J. (1997a). Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 18:307-316.
134. Rabilloud T, Valette C, Lawrence J. (1997b). Sample application by in-gel rehydration improves the resolution of two-dimensional electrophoresis with immobilized pH gradients in the first dimension. *Electrophoresis* 18:307-316.
135. Rabilloud T. (2002). Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* 2:3-10.
136. Raghupathi, RN and Diwan, AM (1994) A protocol for protein estimation that gives a nearly constant color yield with simple proteins and nullifies the effects of four known interfering agents: microestimation of peptide groups. *Analytical biochemistry* 219:356–359.
137. Ramakrishna, A. and Ravishankar, G.A (2011) Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal Behav*, 6:1720–1731.
138. Reid GE, McLuckey SA. (2002). Top-down “ protein characterization via tandem mass spectrometry. *J. Mass Spectrom* 37:663-75.
139. Reichardt, P.B., Chapin III, F.S., Bryant, J.P., Mattes, B.R., Clausen, T.P. (1991). Carbon/nitrogen balance as a predictor to of plant defence in Alaskan balsam poplar: potential importance of metabolite turnover. *Oecologia* 88:401–406.
140. Roberts, S.C. (2007) Production and engineering of terpenoids in plant cell culture. *Nat Chem Biol*, 3, 387-395.
141. Romero-Sandoval EA, Kolano AL, Alvarado-Vázquez PA (2017) Cannabis and Cannabinoids for Chronic Pain. *Current Rheumatology Reports* 19:. doi: 10.1007/s11926-017-0693-1
142. Rose JK, Bashir S, Giovannoni JJ, Jahn MM, Saravanan RS. 2004. Tackling the plant proteome: practical approaches, hurdles and experimental tools. *Plant. J.* 39:715-733.
143. Rupasinghe, H.P., Paliyath, G., & Murr, D.P. (2000). Sesquiterpene α -Farnesene Synthase: Partial Purification, Characterization, and Activity in Relation to Superficial Scald Development in Apples.
144. Riipi, M., Ossipov, V., Lempa, K., Haukioja, E., Koricheva, J., Ossipova, S., Pihlaja, K., 2002. Seasonal changes in birch leaf chemistry: are there trade-offs between leaf growth and accumulation of phenolics? *Oecologia* 130, 380–390.

145. Rzedowski, G. C. de, J. Rzedowski y colaboradores (2005). Flora fanerogámica del Valle de México. 2a. ed., 1a reimpr., Instituto de Ecología, A.C. y Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Pátzcuaro (Michoacán), pp 967-968
146. Sakai J, Kojima S, Yanagi K, Kanaoka M. 2005. 18O-labeling quantitative proteomics using an ion trap mass spectrometer. *Proteomics* 5 (1) 16-23.
147. Sánchez-Ramos, M., Bahena, S.M., Romero-Estrada, A., Bernabé-Antonio, A., Cruz-Sosa, F., González-Christen, J., Acevedo-Fernández, J.J., Perea-Arango, I., Alvarez, L (2018) Establishment and Phytochemical Analysis of a Callus Culture from *Ageratina pichinchensis* (Asteraceae) and Its Anti-Inflammatory Activity. *Mol. J. Synth. Chem. Nat. Prod. Chem* 23:1-14.
148. Santoni V, Doumas P, Rouquié D, Mansion M, Rabilloud T. 1999. Large scale characterization of plant plasma membrane proteins. *Biochimie* 81:655-661.
149. Santoni V, Molloy M, Rabilloud T. 2000. Membrane proteins and proteomics: un amour impossible? *Electrophoresis* 21:1054-1070.
150. Senthil, K., N. Karunanithi, G.S. Kim, A. Nagappan, S. Sundareswaran, S. Natesan and R. Muthurajan (2011) Proteome analysis of in vitro and in vivo root tissue of *Withania somnifera*. *Afr. J. Biotechnol* 10:16875-16883
151. Sergeant, K and Renaut, J (2010) Plant Biotic Stress and Proteomics. *Curr. Proteomics*, 7:275–297.
152. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* 13:2498–2504
153. Simpson RJ. 2003a. Introduction to proteomics. In: Cold spring harbour laboratory press. Protein and proteomics. pp 1-37.
154. Simpson RJ. 2003b. The use of mass spectrometry in proteomics. In: Cold spring harbour laboratory press. Protein and proteomics. pp 1-37. 425-595.
155. Simon, J., Gleadow, R.M., Woodrow, I.E., 2010. Allocation of nitrogen to chemical defence and plant functional traits is constrained by soil N. *Tree Physiology* 30, 1111–1117.
156. Siuti N, Kelleher NL. 2007. Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* 4:817-21.
157. Statsoft INC (2007). STATISTICA for Windows. Tulsa, USA.
158. Sturn A, Quackenbush J, Trajanoski Z. 2002. Genesis: Cluster analysis of microarray data. *Bioinformatics* 18 (1):207-208.
159. Sud, A., Chauhan, R.S., Tandon, C (2014) Mass Spectrometric Analysis of Differentially Expressed Proteins in an Endangered Medicinal Herb, *Picrorhiza kurroa*. *BioMed Research International* 2014, 1-12.
160. Su-Fang Ee, Zeti-Azura Mohamed-Hussein, Roohaida Othman, Noor Azmi Shaharuddin, Ismanizan Ismail, and Zamri Zainal, “Functional Characterization of Sesquiterpene Synthase from *Polygonum minus*,” *The Scientific World Journal*, vol. 2014, Article ID 840592, 11 pages, 2014. <https://doi.org/10.1155/2014/840592>.
161. Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al (2017) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res* 45, D362–D368.
162. Tabb DL, McDonald HW, Yates JR III. 2002. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* 1:21-26.
163. Takáč, T., Pechan, T., and Šamaj, J (2011) Differential proteomics of plant development. *Journal of Proteomics* 74:577–588.
164. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, et al. 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 75:1895-904.
165. Tian, L., Peel, G.J., Lei, Z., Aziz, N., Dai, X., He, J., et al. (2009) Transcript and proteomic analysis of developing white lupin (*Lupinus albus* L.) roots. *BMC Plant Biology* 9:1.
166. Trupiano, D., Rocco, M., Renzone, G., Scaloni, A., Viscosi, V., Chiatante, D., Scippa, G.S (2012) The proteome of *Populus nigra* woody root: Response to bending. *Ann. Bot* 110:415–432.
167. Unlu M, Morgan ME, Minden JS. 1997. Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. *Electrophoresis* 18:2071-2077.
168. Uttenweiler JS, Claverol S, Sylvius L, Bousquet Dubouch MP, Burret Schiltz O, Monsarrat B. 2008. Toward a full characterization of the human 20S proteasome subunits and their isoforms by a combination of proteomic approaches. *Methods Mol. Biol.* 484:111-30.
169. Unsicker SB, Kunert G, Gershenzon J (2009) Protective perfumes: the role of vegetative volatiles in plant defense against herbivores. *Curr Opin Plant Biol* 12:479–485

170. Vincent D, Ergul A, Bohlman MC, Tattersall EAR, Tillett RL, Wheatley MD, Woolsey R, Quilici DR, Joets J, Schlauch K, Schooley DA, Cushman JC, Cramer GR. 2007. Proteomic analysis reveals differences between *Vitis vinifera* L. cv. Chardonnay and cv. Cabernet Sauvignon and their responses to water deficit and salinity. *J Exp Bot* 58:1873-1892.
171. Wang YK, Ma Z, Quinn DF, Fu EW. 2002. Inverse 15 N-metabolic labeling/mass spectrometry for comparative proteomics and rapid identification of protein markers/targets. *Rapid Commun. Mass Spectrom.* 16:1389-97.
172. Wang, W., Scali, M., Vignani, R., Spadafora, A., Sensi, E., Mazzuca, S., Cresti, M (2003) Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis* 24:2369–2375.
173. Washburn MP, Wolters D, Yates JR 3rd. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol* 19:242-247.
174. Waterman, P.G., Mole, S., 1994. *Analysis of Phenolic Plant Metabolites*. Blackwell Scientific Publications, London.
175. Withers, S.T. and Keasling, J.D. (2007) Biosynthesis and engineering of isoprenoid small molecules. *Appl Microbiol Biotechnol*, 73, 980-990.
176. Wolters DA, Washburn MP, Yates JR 3rd. 2001. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* 73:5683-90.
177. Wymore, T., Chen, B. Y., Nicholas, H. B., Ropelewski, A. J., and Brooks, C. L. (2011). A mechanism for evolving novel plant sesquiterpene synthase function. *Mol. Inform.* 30, 896–906. doi: 10.1002/minf.201100087
178. Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C. 2001. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* 73:2836-42.
179. Yates JR, Ruse CI, Nakorchevsky A. 2009. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu. Rev. Biomed. Eng.* 11:49-79.
180. Zar, J.H. *Biostatistical Analysis*; Prentice Hall, 2010; ISBN 978-0-13-100846-5.
181. Zhang N, Aebersold R, Schwikowski B. 2002. ProBID: a probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. *Proteomics* 2:406-1412.
182. Zhao, X., Si, J., Miao, Y., Peng, Y., Wang, L.; Cai, X (2014) Comparative proteomics of *Euphorbia kansui* Liou milky sap at two different developmental stages. *Plant Physiol. Biochem* 79:60–65.
183. Zubarev RA, Kelleher NL, McLafferty FW. 1998. Electron capture dissociation of multiply charged protein cations. A non ergodic process. *J. Am. Chem. Soc.* 120:3265-66.

ANEXOS

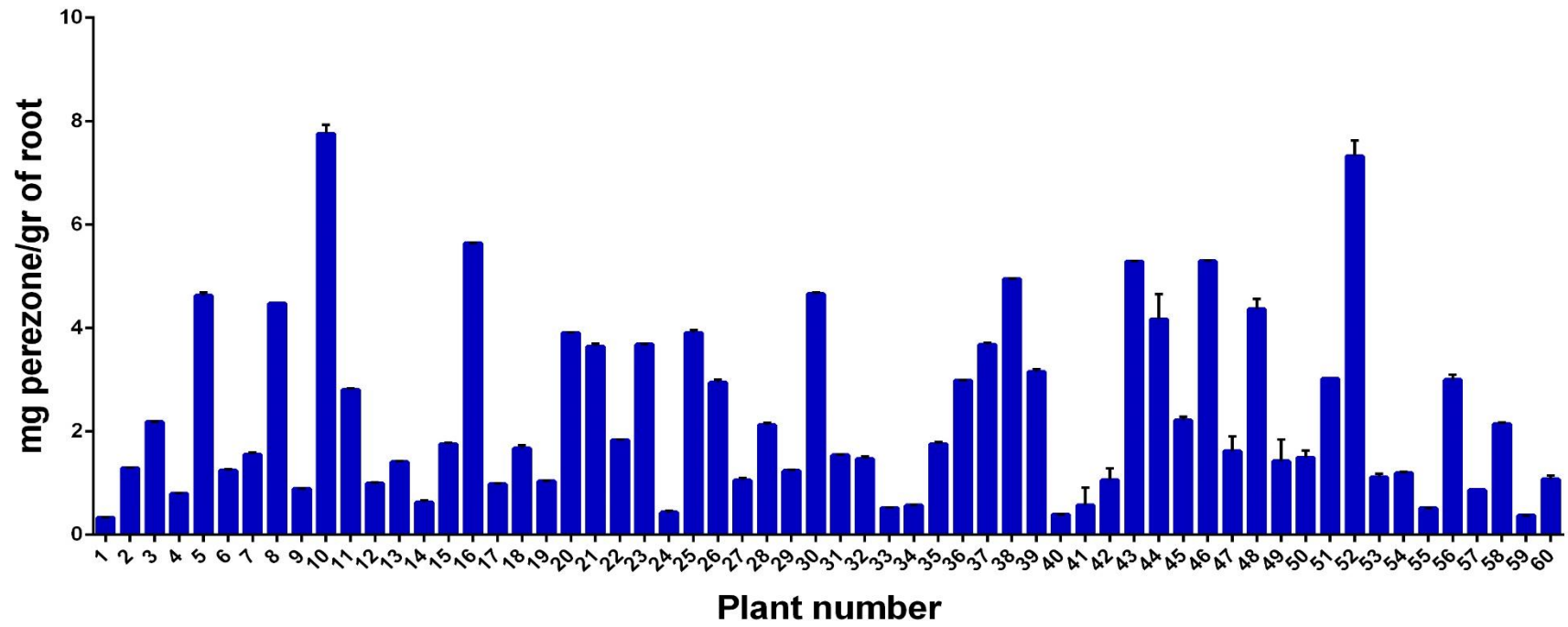


Figura Suplementaria 1. Rendimiento de perezona en las raíces de plantas de la especie *A. cordata*.

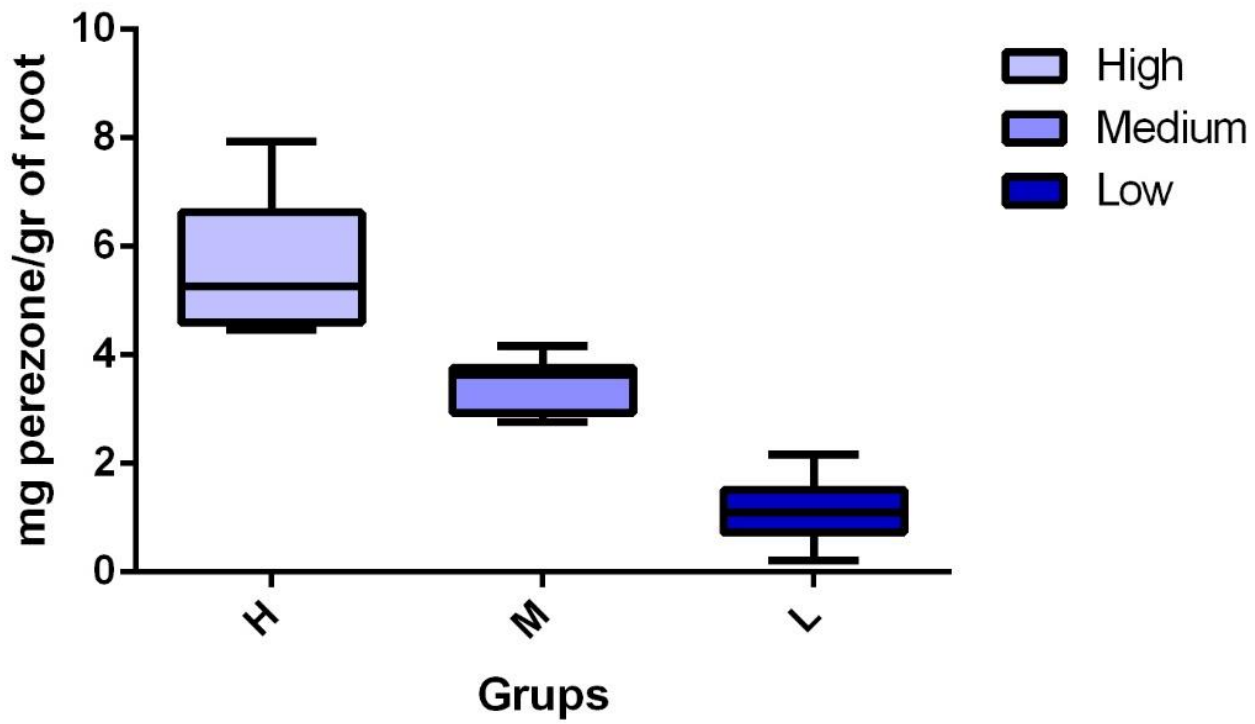


Figura Suplementaria 2. Agrupación de las plantas de *A. cordata* de acuerdo con la producción de perezona.

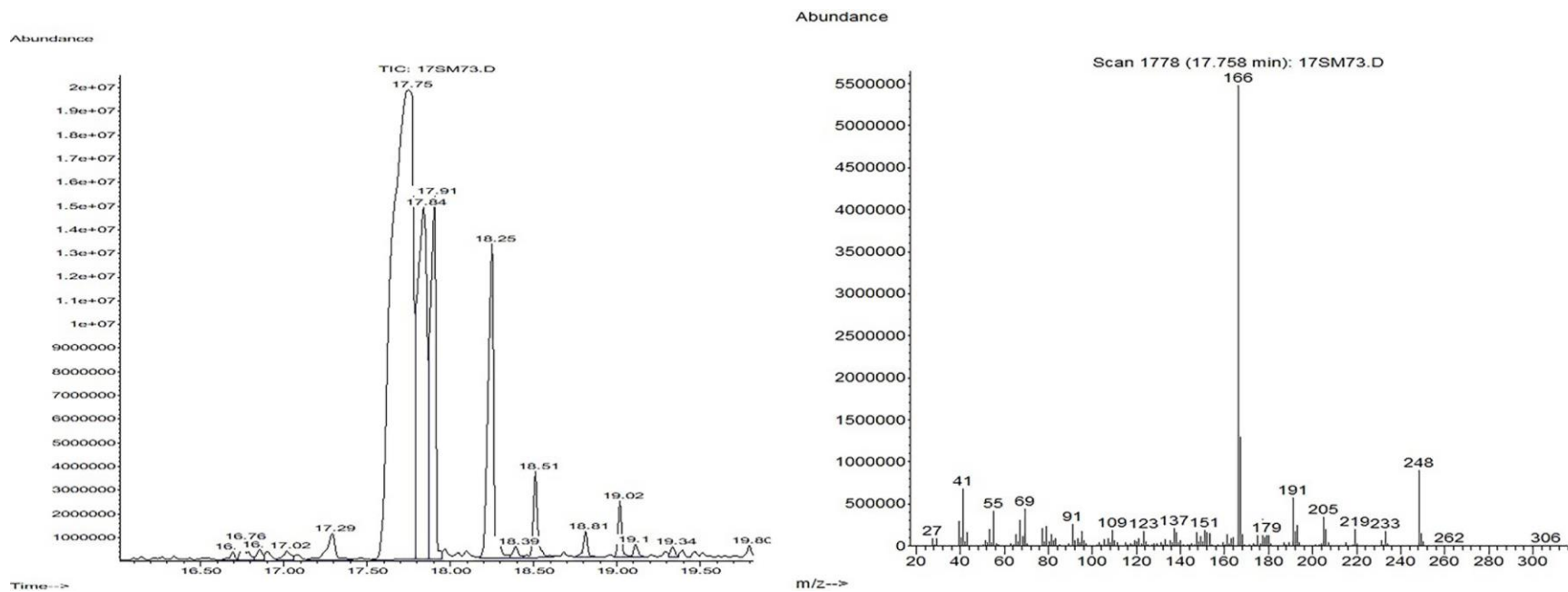


Figura Suplementaria 3. Análisis de los extractos hexánicos de las raíces *A.cordata* de por GC-MS. Identificación del compuesto en el grupo de alta producción con tiempo de retención de (RT) = 17.75 min y m/z= 166, 191,205 y 248.

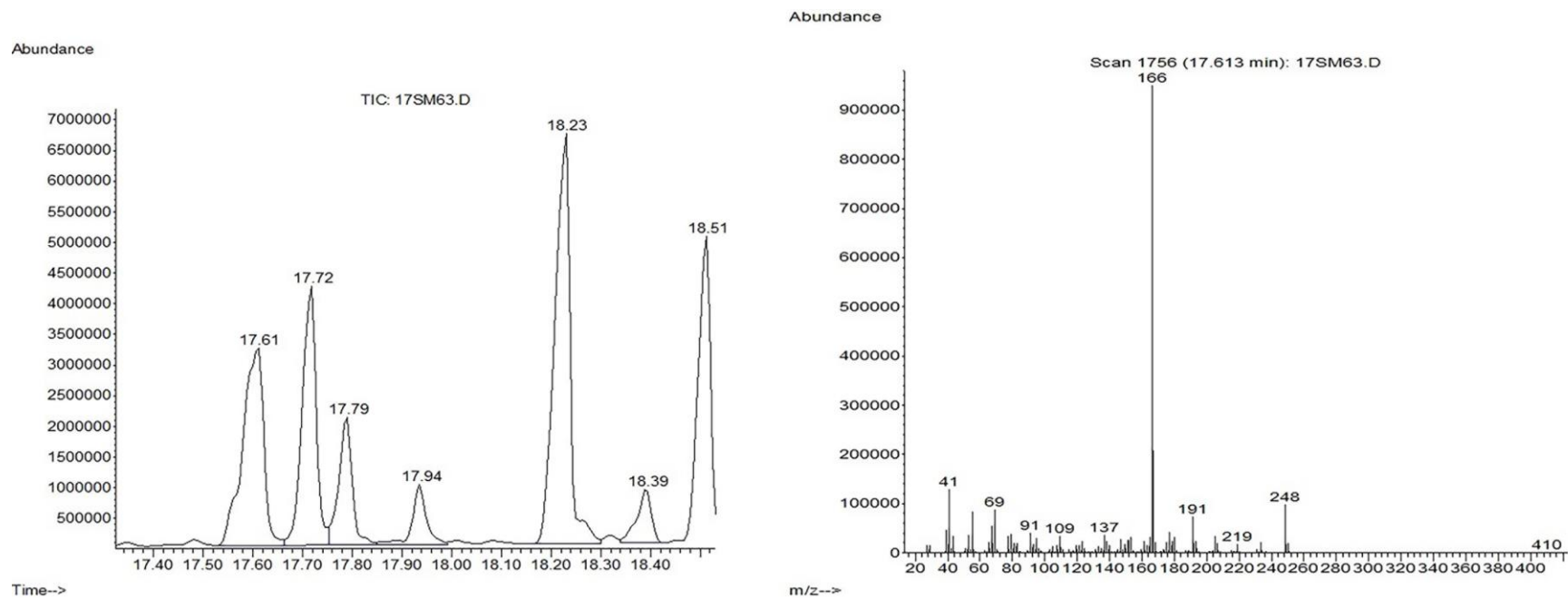


Figura suplementaria 4. Análisis de los extractos hexánicos de las raíces *A.cordata* de por GC-MS. Identificación del compuesto en el grupo de baja producción con tiempo de retención de (RT) = 17.61 min y $m/z= 166, 191, 205$ y 248.

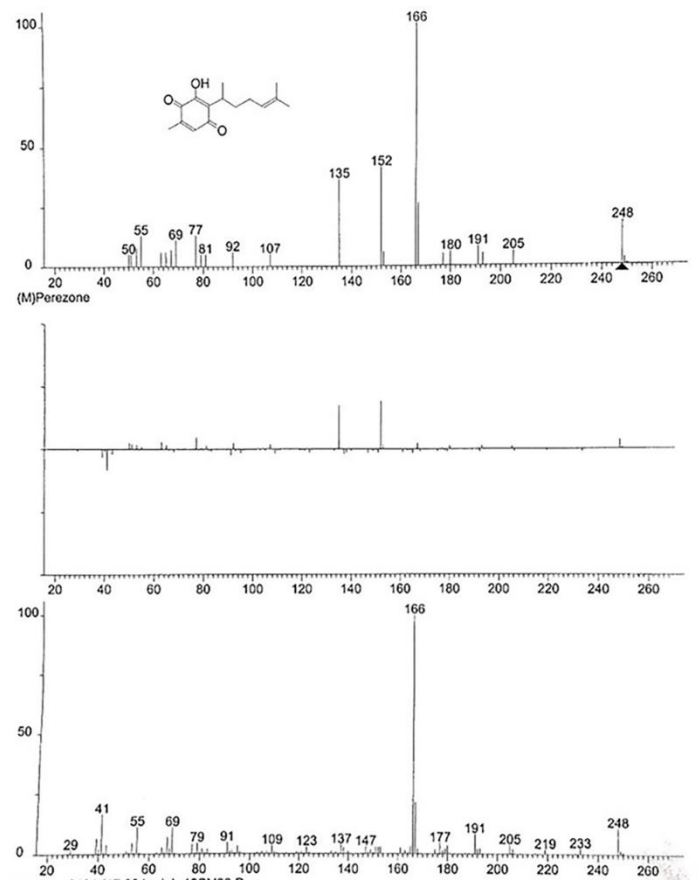
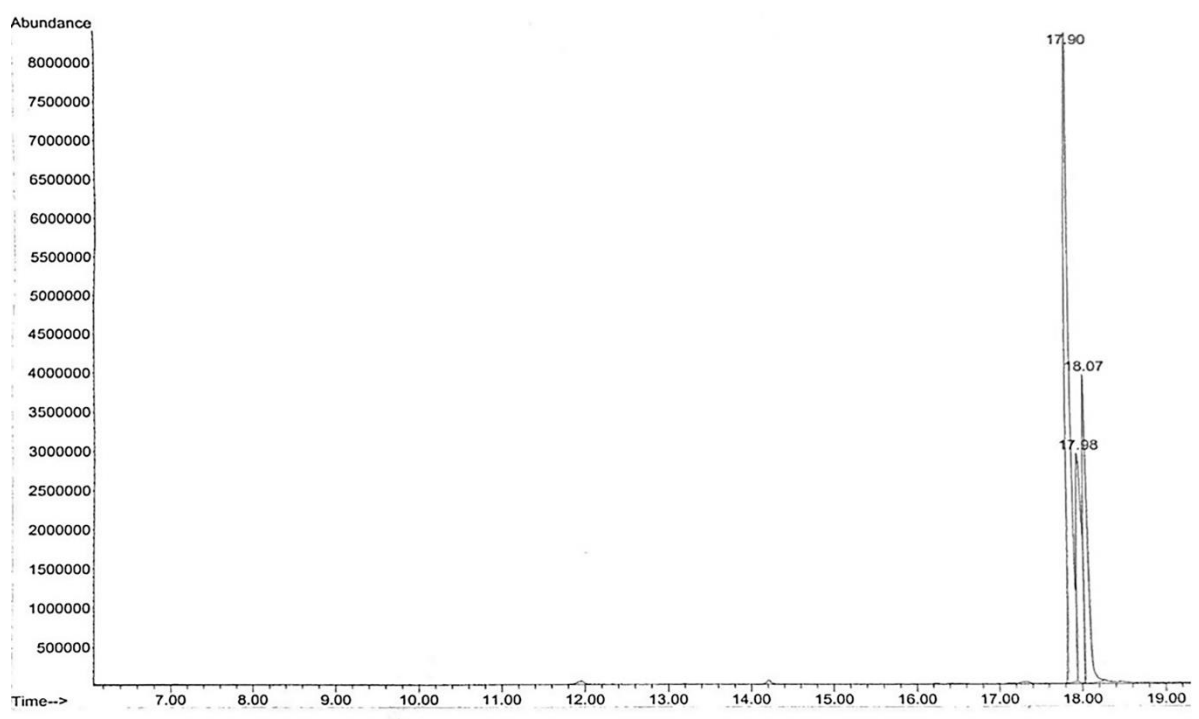


Figura suplementaria 5. Análisis por Cromatografía de Gases acoplada a Espectrometría de Masas de la perezona pura como estándar.

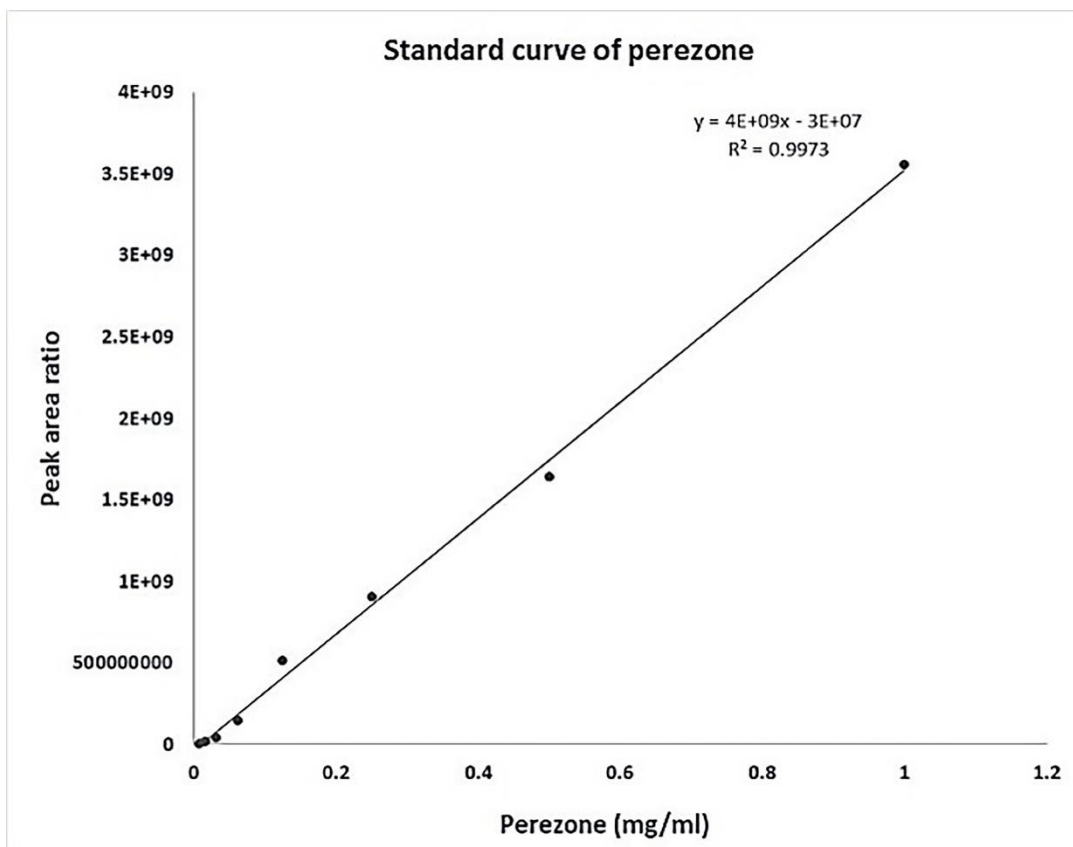


Figura suplementaria 6. Curva estándar para la perezona pura utilizada como referencia para la búsqueda y cuantificación del compuesto en las plantas silvestres de *A. coradata*.

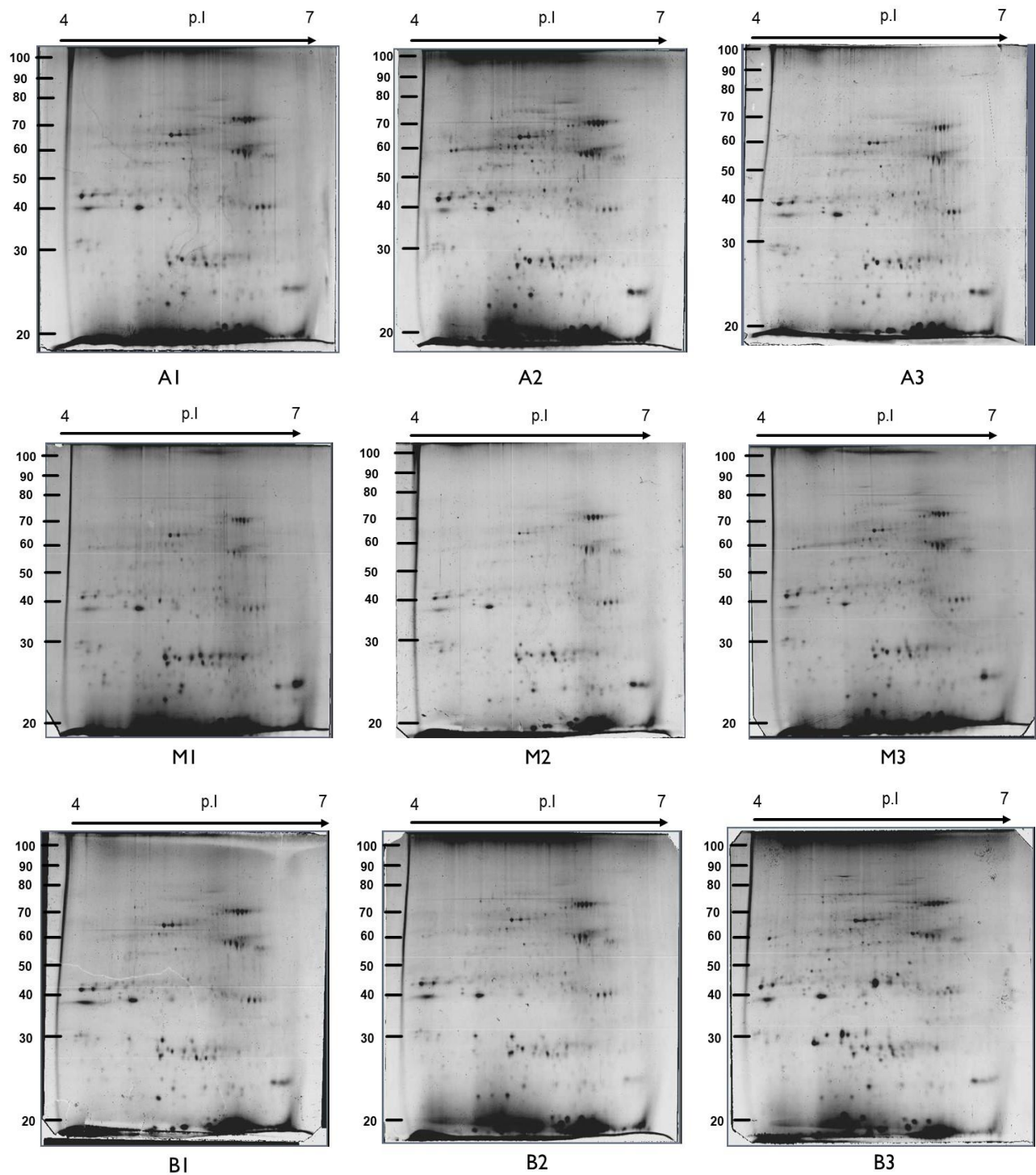


Figura suplementaria 7. Imágenes de los perfiles proteicos de los grupos altos, medios y bajos productores.

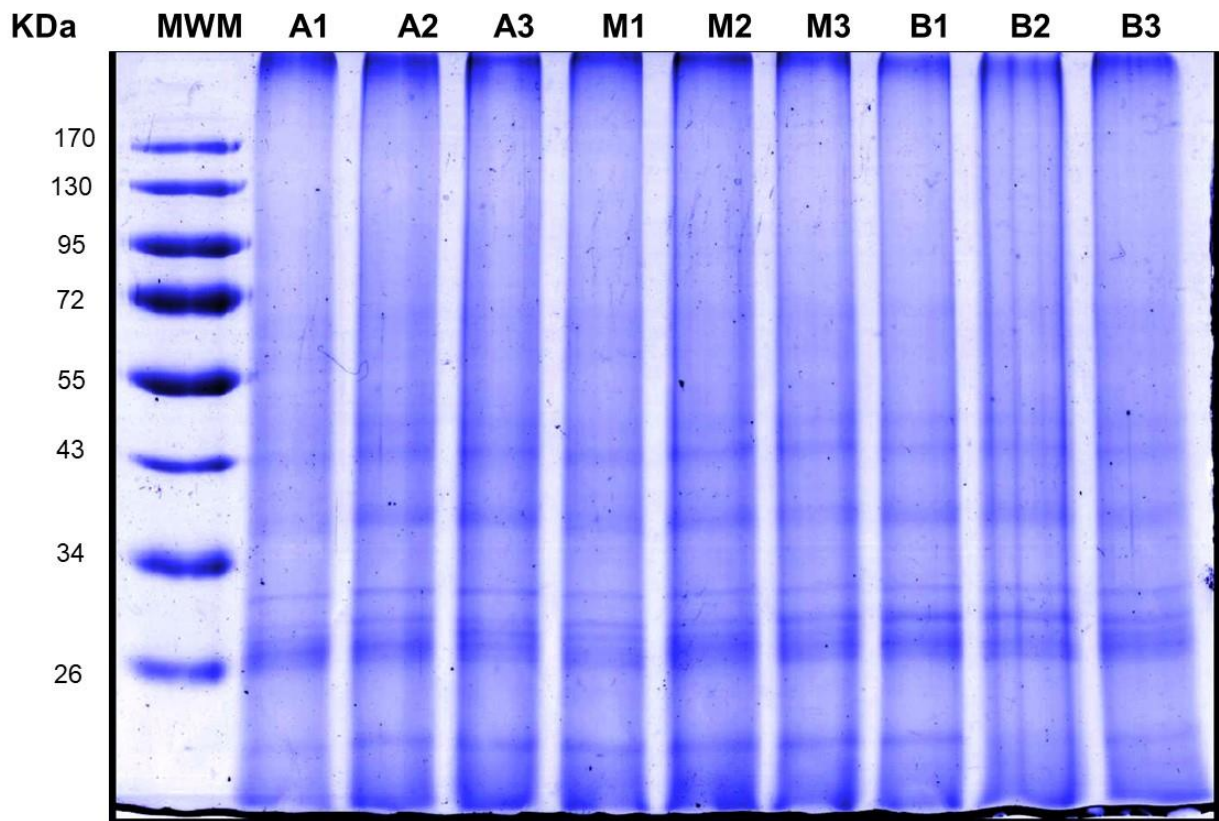


Figura suplementaria 8. Perfiles proteicos de las raíces productoras de perezona de *A. cordata*, se observa en la imagen en el primer carril el marcador de peso molecular (MWM), posteriormente del carril 2 al 4 las muestras por triplicado del grupo alto productor de perezona (A1, A2 y A3), en seguida del carril 5 al 7 se encuentran las muestras del grupo medianamente productor de perezona (M1, M2 y M3) y del carril 8 al 10 se colocaron las muestras proteicas del grupo de baja producción de perezona (B1, B2 y B3).

Tabla suplementaria S1. Identificación de proteínas expresadas diferencialmente en las raíces productoras de perezona de *A. cordata* (análisis de las imágenes 2DE).

Spot number	Anova (<i>p</i>)	Fold	Max CV (%)	Highest Mean	Lowest Mean	Area	Normalized Volume								
							High			Medium			Low		
							A1	A2	A3	M1	M2	M3	B1	B2	B3
**128	0.016	1.749	22.353	Altos	Medios	535	4278.17088	3986.0403	5020	2550.84204	1956.47943	3087.84746	4209.45309	3184.62017	3788.89859
**207	0.002	2.259	19.583	Medios	Bajos	2529	27748.4014	34855.736	26349	42063.9552	32232.9078	38251.7557	14448.6633	20351.4363	15031.5101
**271	0.007	1.254	7.864	Bajos	Medios	7426	244505.079	252379.288	256944	245529.234	262735.707	230443.042	313387.252	282596.635	330530.181
**277	0.007	1.475	14.036	Bajos	Medios	1773	20389.231	22485.6148	23017	19298.7588	23065.1622	17581.9541	32186.7958	28379.4561	27862.6887
**386	0.019	3.245	41.895	Medios	Altos	5684	112553.511	59982.7028	58681	362992.659	231277.163	155990.732	164577.396	108464.677	164124.525
**455	0.014	1.868	24.055	Bajos	Altos	782	2579.44148	2909.27862	3869	3409.43551	3494.31823	3807.34589	5399.11167	7391.39115	4687.78809

** Spots con $p < 0.02$

Tabla suplementaria S2. Cuantificación de perezona en las raíces de *A. cordata*.

	R1	R2	R3	R1	R2	R3					R1				R2				R3				R1	R2	R3		
#	perezone area	perezone area	perezone area	Concentration of perezone mg/ml	Concentration of perezone mg/ml	Concentration of perezone mg/ml	Total hexanic extract (mg)	Volume (ml)	Used extract (mg)	dry plant material (gr)	perezone (mg)	perezone (mg / hexanic extract into CG-EM (mg)	hexanic extract (mg) / root (g)	perezone (mg) / root (g)	perezone (mg)	perezone (mg / hexanic extract into CG-EM (mg)	hexanic extract (mg) / root (g)	perezone (mg) / root (g)	perezone (mg)	perezone (mg / hexanic extract into CG-EM (mg)	hexanic extract (mg) / root (g)	perezone (mg) / root (g)	perezone (mg) / root (g)	perezone (mg) / root (g)	mean	standard deviation	
1	84757945	81282783	83020364	0.03406	0.03309	0.03357	70.7	0.5	2	1.8	0.017	0.00852	39.278	0.3345	0.0165	0.0083	39.278	0.3249	0.017	0.0084	39.278	0.3297	0.3345	0.3249	0.3297	0.3297	0.0048
2	4.51E+08	4.57E+08	4.54E+08	0.13667	0.13832	0.13749	82.6	0.5	2	2.2	0.0683	0.03417	37.545	1.2829	0.0692	0.0346	37.545	1.2983	0.069	0.0344	37.545	1.2906	1.2829	1.2983	1.2906	1.2906	0.0077
3	9.2E+08	9.28E+08	9.24E+08	0.26797	0.2704	0.26918	64.8	0.5	2	2	0.134	0.06699	32.4	2.1706	0.1352	0.0676	32.4	2.1902	0.135	0.0673	32.4	2.1804	2.1706	2.1902	2.1804	2.1804	0.0098
4	1.91E+08	2.01E+08	1.96E+08	0.06387	0.06672	0.06529	107	0.5	2	2.2	0.0319	0.01597	48.636	0.7766	0.0334	0.0167	48.636	0.8112	0.033	0.0163	48.636	0.7939	0.7766	0.8112	0.7939	0.7939	0.0173
5	1.48E+09	1.52E+09	1.5E+09	0.42394	0.43568	0.42981	86.1	0.5	2	2	0.212	0.10598	43.05	4.5626	0.2178	0.1089	43.05	4.689	0.215	0.1075	43.05	4.6258	4.5626	4.689	4.6258	4.6258	0.0632
6	3.45E+08	3.31E+08	3.38E+08	0.10706	0.10312	0.10509	99.3	0.5	2	2.1	0.0535	0.02677	47.286	1.2656	0.0516	0.0258	47.286	1.219	0.053	0.0263	47.286	1.2423	1.2656	1.219	1.2423	1.2423	0.0233
7	2E+08	2.11E+08	2.05E+08	0.06643	0.06934	0.06788	146.2	0.5	2	1.6	0.0332	0.01661	91.375	1.5174	0.0347	0.0173	91.375	1.5839	0.034	0.017	91.375	1.5506	1.5174	1.5839	1.5506	1.5506	0.0332
8	9.75E+08	9.75E+08	9.75E+08	0.28357	0.28357	0.28357	113.6	0.5	2	1.8	0.1418	0.07089	63.111	4.4741	0.1418	0.0709	63.111	4.4741	0.142	0.0709	63.111	4.4741	4.4741	4.4741	4.4741	4.4741	3E-05
9	2.35E+08	2.34E+08	2.34E+08	0.07608	0.07583	0.07596	79.6	0.5	2	1.7	0.038	0.01902	46.824	0.8906	0.0379	0.019	46.824	0.8877	0.038	0.019	46.824	0.8891	0.8906	0.8877	0.8891	0.8891	0.0015
10	1.78E+09	1.71E+09	1.74E+09	0.51005	0.48805	0.49905	136.8	0.5	2	2.2	0.255	0.12751	62.182	7.9289	0.244	0.122	62.182	7.587	0.25	0.1248	62.182	7.758	7.9289	7.587	7.758	7.758	0.1709
11	6.47E+08	6.62E+08	6.54E+08	0.19144	0.19587	0.19366	115.8	0.5	2	2	0.0957	0.04786	57.9	2.7711	0.0979	0.049	57.9	2.8353	0.097	0.0484	57.9	2.8032	2.7711	2.8353	2.8032	2.8032	0.0321
12	2.26E+08	2.35E+08	2.3E+08	0.07356	0.07613	0.07485	100.6	0.5	2	1.9	0.0368	0.01839	52.947	0.9737	0.0381	0.019	52.947	1.0078	0.037	0.0187	52.947	0.9907	0.9737	1.0078	0.9907	0.9907	0.017
13	3.38E+08	3.47E+08	3.42E+08	0.10507	0.10741	0.10624	100.8	0.5	2	1.9	0.0525	0.02627	53.053	1.3936	0.0537	0.0269	53.053	1.4246	0.053	0.0266	53.053	1.4091	1.3936	1.4246	1.4091	1.4091	0.0155
14	1.43E+08	1.62E+08	1.52E+08	0.05032	0.05573	0.05302	90.2	0.5	2	1.9	0.0252	0.01258	47.474	0.5972	0.0279	0.0139	47.474	0.6614	0.027	0.0133	47.474	0.6293	0.5972	0.6614	0.6293	0.6293	0.0321
15	4.41E+08	4.25E+08	4.33E+08	0.13388	0.12925	0.13156	112	0.5	2	2.1	0.0669	0.03347	53.333	1.785	0.0646	0.0323	53.333	1.7234	0.066	0.0329	53.333	1.7542	1.785	1.7234	1.7542	1.7542	0.0308
16	1.8E+09	1.8E+09	1.8E+09	0.51567	0.51478	0.51522	87.6	0.5	2	2	0.2578	0.12892	43.8	5.6466	0.2574	0.1287	43.8	5.6368	0.258	0.1288	43.8	5.6417	5.6466	5.6368	5.6417	5.6417	0.0049
17	3.72E+08	3.58E+08	3.65E+08	0.11445	0.11073	0.11259	90.6	0.5	2	2.6	0.0572	0.02861	34.846	0.9971	0.0554	0.0277	34.846	0.9646	0.056	0.0281	34.846	0.9808	0.9971	0.9646	0.9808	0.9808	0.0162
18	4.52E+08	4.17E+08	4.35E+08	0.13701	0.12727	0.13214	101.1	0.5	2	2	0.0685	0.03425	50.55	1.7315	0.0636	0.0318	50.55	1.6084	0.066	0.033	50.55	1.67	1.7315	1.6084	1.67	1.67	0.0616
19	3.55E+08	3.59E+08	3.57E+08	0.10964	0.11089	0.11027	75.2	0.5	2	2	0.0548	0.02741	37.6	1.0307	0.0554	0.0277	37.6	1.0424	0.055	0.0276	37.6	1.0365	1.0307	1.0424	1.0365	1.0365	0.0059
20	1.2E+09	1.19E+09	1.2E+09	0.34654	0.34503	0.34579	113	0.5	2	2.5	0.1733	0.08664	45.2	3.916	0.1725	0.0863	45.2	3.8988	0.173	0.0864	45.2	3.9074	3.916	3.8988	3.9074	3.9074	0.0086
21	9.21E+08	8.94E+08	9.08E+08	0.26841	0.2608	0.26461	82.6	0.5	2	1.5	0.1342	0.0671	55.067	3.6951	0.1304	0.0652	55.067	3.5904	0.132	0.0662	55.067	3.6427	3.6951	3.5904	3.6427	3.6427	0.0524
22	3.71E+08	3.75E+08	3.73E+08	0.11422	0.11539	0.1148	127.7	0.5	2	2	0.0571	0.02855	63.85	1.8232	0.0577	0.0288	63.85	1.842	0.057	0.0287	63.85	1.8326	1.8232	1.842	1.8326	1.8326	0.0094
23	9.47E+08	9.52E+08	9.49E+08	0.27551	0.27698	0.27624	95.9	0.5	2	1.8	0.1378	0.06888	53.278	3.6696	0.1385	0.0692	53.278	3.6892	0.138	0.0691	53.278	3.6794	3.6696	3.6892	3.6794	3.6794	0.0098
24	1.49E+08	1.25E+08	1.37E+08	0.05204	0.04527	0.04866	63.3	0.5	2	1.8	0.026	0.01301	35.167	0.4575	0.0226	0.0113	35.167	0.398	0.024	0.0122	35.167	0.4278	0.4575	0.398	0.4278	0.4278	0.0298
25	1.31E+09	1.35E+09	1.33E+09	0.37682	0.38737	0.3821	77.7	0.5	2	1.9	0.1884	0.0942	40.895	3.8525	0.1937	0.0968	40.895	3.9604	0.191	0.0955	40.895	3.9064	3.8525	3.9604	3.9064	3.9064	0.054
26	8.46E+08	8.83E+08	8.64E+08	0.24724	0.25767	0.25246	74.6	0.5	2	1.6	0.1236	0.06181	46.625	2.8819	0.1288	0.0644	46.625	3.0035	0.126	0.0631	46.625	2.9427	2.8819	3.0035	2.9427	2.9427	0.0608
27	2.91E+08	3.17E+08	3.04E+08	0.09189	0.09925	0.09557	83.7	0.5	2	1.9	0.0459	0.02297	44.053	1.012	0.0496	0.0248	44.053	1.0931	0.048	0.0239	44.053	1.0525	1.012	1.0931	1.0525	1.0525	0.0405
28	5.86E+08	5.62E+08	5.74E+08	0.17457	0.16772	0.17115	94.3	0.5	2	1.9	0.0873	0.04364	49.632	2.166	0.0839	0.0419	49.632	2.0811	0.086	0.0428	49.632	2.1236	2.166	2.0811	2.1236	2.1236	0.0425
29	3E+08	3.14E+08	3.07E+08	0.09444	0.09831	0.09638	87	0.5	2	1.7	0.0472	0.02361	51.176	1.2083	0.0492	0.0246	51.176	1.2578	0.048	0.0241	51.176	1.2331	1.2083	1.2578	1.2331	1.2331	0.0247

30	1.08E+09	1.1E+09	1.09E+09	0.31337	0.31757	0.31547	106.2	0.5	2	1.8	0.1567	0.07834	59	4.6222	0.1588	0.0794	59	4.6842	0.158	0.0789	59	4.6532	4.6222	4.6842	4.6532	4.6532	0.031
31	4.15E+08	4.24E+08	4.19E+08	0.1265	0.12901	0.12776	77	0.5	2	1.6	0.0633	0.03163	48.125	1.522	0.0645	0.0323	48.125	1.5522	0.064	0.0319	48.125	1.5371	1.522	1.5522	1.5371	1.5371	0.0151
32	3.7E+08	3.98E+08	3.84E+08	0.11404	0.12191	0.11798	99.4	0.5	2	2	0.057	0.02851	49.7	1.417	0.061	0.0305	49.7	1.5147	0.059	0.0295	49.7	1.4659	1.417	1.5147	1.4659	1.4659	0.0489
33	95645646	90909872	93277759	0.03711	0.03579	0.03645	102.1	0.5	2	1.8	0.0186	0.00928	56.722	0.5263	0.0179	0.0089	56.722	0.5075	0.018	0.0091	56.722	0.5169	0.5263	0.5075	0.5169	0.5169	0.0094
34	1.04E+08	1.1E+08	1.07E+08	0.03933	0.04103	0.04018	106.9	0.5	2	1.9	0.0197	0.00983	56.263	0.5533	0.0205	0.0103	56.263	0.5772	0.02	0.01	56.263	0.5652	0.5533	0.5772	0.5652	0.5652	0.012
35	5.33E+08	5.6E+08	5.47E+08	0.15974	0.16721	0.16348	85.8	0.5	2	2	0.0799	0.03993	42.9	1.7132	0.0836	0.0418	42.9	1.7933	0.082	0.0409	42.9	1.7533	1.7132	1.7933	1.7533	1.7533	0.0401
36	9.19E+08	9.27E+08	9.23E+08	0.26788	0.26994	0.26891	79.9	0.5	2	1.8	0.1339	0.06697	44.389	2.9727	0.135	0.0675	44.389	2.9956	0.134	0.0672	44.389	2.9841	2.9727	2.9956	2.9841	2.9841	0.0114
37	9.53E+08	9.71E+08	9.62E+08	0.27732	0.28246	0.27989	110.3	0.5	2	2.1	0.1387	0.06933	52.524	3.6415	0.1412	0.0706	52.524	3.709	0.14	0.07	52.524	3.6753	3.6415	3.709	3.6753	3.6753	0.0338
38	1.62E+09	1.61E+09	1.62E+09	0.46392	0.46192	0.46292	81.2	0.5	2	1.9	0.232	0.11598	42.737	4.9566	0.231	0.1155	42.737	4.9353	0.231	0.1157	42.737	4.946	4.9566	4.9353	4.946	4.946	0.0107
39	1.24E+09	1.29E+09	1.26E+09	0.35889	0.37035	0.36462	58.8	0.5	2	1.7	0.1794	0.08972	34.588	3.1034	0.1852	0.0926	34.588	3.2024	0.182	0.0912	34.588	3.1529	3.1034	3.2024	3.1529	3.1529	0.0495
40	92792099	94059041	93425570	0.03631	0.03667	0.03649	51.3	0.5	2	1.2	0.0182	0.00908	42.75	0.3881	0.0183	0.0092	42.75	0.3919	0.018	0.0091	42.75	0.39	0.3881	0.3919	0.39	0.39	0.0019
41	18443344	1.87E+08	1.03E+08	0.01548	0.06264	0.03906	93.3	0.5	2	1.6	0.0077	0.00387	58.313	0.2257	0.0313	0.0157	58.313	0.9132	0.02	0.0098	58.313	0.5694	0.2257	0.9132	0.5694	0.5694	0.3437
42	3.09E+08	4.97E+08	4.03E+08	0.09695	0.14944	0.1232	75.4	0.5	2	2.2	0.0485	0.02424	34.273	0.8307	0.0747	0.0374	34.273	1.2805	0.062	0.0308	34.273	1.0556	0.8307	1.2805	1.0556	1.0556	0.2249
43	1.09E+09	1.1E+09	1.1E+09	0.31683	0.31862	0.31772	139.7	0.5	2	2.1	0.1584	0.07921	66.524	5.2691	0.1593	0.0797	66.524	5.299	0.159	0.0794	66.524	5.2841	5.2691	5.299	5.2841	5.2841	0.0149
44	1.26E+09	1.6E+09	1.43E+09	0.36327	0.45937	0.41132	77	0.5	2	1.9	0.1816	0.09082	40.526	3.6805	0.2297	0.1148	40.526	4.6541	0.206	0.1028	40.526	4.1673	3.6805	4.6541	4.1673	4.1673	0.4868
45	9.54E+08	1.02E+09	9.86E+08	0.27757	0.29549	0.28653	67.9	0.5	2	2.2	0.1388	0.06939	30.864	2.1417	0.1477	0.0739	30.864	2.28	0.143	0.0716	30.864	2.2109	2.1417	2.28	2.2109	2.2109	0.0691
46	1.27E+09	1.27E+09	1.27E+09	0.36663	0.36488	0.36575	133.1	0.5	2	2.3	0.1833	0.09166	57.87	5.3042	0.1824	0.0912	57.87	5.2788	0.183	0.0914	57.87	5.2915	5.3042	5.2788	5.2915	5.2915	0.0127
47	4E+08	5.85E+08	4.93E+08	0.12251	0.17411	0.14831	78.5	0.5	2	1.8	0.0613	0.03063	43.611	1.3357	0.0871	0.0435	43.611	1.8983	0.074	0.0371	43.611	1.617	1.3357	1.8983	1.617	1.617	0.2813
48	1.03E+09	1.13E+09	1.08E+09	0.29852	0.32623	0.31237	89.4	0.5	2	1.6	0.1493	0.07463	55.875	4.1699	0.1631	0.0816	55.875	4.557	0.156	0.0781	55.875	4.3635	4.1699	4.557	4.3635	4.3635	0.1936
49	2.93E+08	5.68E+08	4.31E+08	0.09242	0.16945	0.13094	82.5	0.5	2	1.9	0.0462	0.0231	43.421	1.0032	0.0847	0.0424	43.421	1.8395	0.065	0.0327	43.421	1.4213	1.0032	1.8395	1.4213	1.4213	0.4181
50	3.2E+08	3.95E+08	3.58E+08	0.10001	0.12096	0.11049	91.8	0.5	2	1.7	0.05	0.025	54	1.3501	0.0605	0.0302	54	1.633	0.055	0.0276	54	1.4916	1.3501	1.633	1.4916	1.4916	0.1414
51	1.04E+09	1.04E+09	1.04E+09	0.30081	0.30104	0.30092	80.2	0.5	2	2	0.1504	0.0752	40.1	3.0156	0.1505	0.0753	40.1	3.0179	0.15	0.0752	40.1	3.0167	3.0156	3.0179	3.0167	3.0167	0.0012
52	2.11E+09	1.94E+09	2.03E+09	0.60227	0.55425	0.57826	86.1	0.5	2	1.7	0.3011	0.15057	50.647	7.6258	0.2771	0.1386	50.647	7.0178	0.289	0.1446	50.647	7.3218	7.6258	7.0178	7.3218	7.3218	0.304
53	2.23E+08	2.57E+08	2.4E+08	0.07285	0.08244	0.07765	120.1	0.5	2	2.1	0.0364	0.01821	57.19	1.0416	0.0412	0.0206	57.19	1.1787	0.039	0.0194	57.19	1.1102	1.0416	1.1787	1.1102	1.1102	0.0685
54	3.39E+08	3.54E+08	3.46E+08	0.1053	0.10937	0.10733	80	0.5	2	1.8	0.0527	0.02633	44.444	1.17	0.0547	0.0273	44.444	1.2152	0.054	0.0268	44.444	1.1926	1.17	1.2152	1.1926	1.1926	0.0226
55	1.14E+08	1.15E+08	1.15E+08	0.04233	0.04256	0.04245	96.8	0.5	2	2	0.0212	0.01058	48.4	0.5122	0.0213	0.0106	48.4	0.515	0.021	0.0106	48.4	0.5136	0.5122	0.515	0.5136	0.5136	0.0014
56	8.75E+08	9.35E+08	9.05E+08	0.25534	0.27215	0.26375	100	0.5	2	2.2	0.1277	0.06384	45.455	2.9016	0.1361	0.068	45.455	3.0926	0.132	0.0659	45.455	2.9971	2.9016	3.0926	2.9971	2.9971	0.0955
57	1.93E+08	1.93E+08	1.93E+08	0.06439	0.06439	0.06439	107.1	0.5	2	2	0.0322	0.0161	53.55	0.862	0.0322	0.0161	53.55	0.862	0.032	0.0161	53.55	0.862	0.862	0.862	0.862	0.862	9E-06
58	4.79E+08	4.91E+08	4.85E+08	0.14442	0.148	0.14621	105.5	0.5	2	1.8	0.0722	0.03611	58.611	2.1162	0.074	0.037	58.611	2.1685	0.073	0.0366	58.611	2.1424	2.1162	2.1685	2.1424	2.1424	0.0262
59	99011170	99817901	99414536	0.03805	0.03828	0.03817	77.3	0.5	2	2	0.019	0.00951	38.65	0.3677	0.0191	0.0096	38.65	0.3699	0.019	0.0095	38.65	0.3688	0.3677	0.3699	0.3688	0.3688	0.0011
60	3.83E+08	3.29E+08	3.56E+08	0.11775	0.1026	0.11018	85.6	0.5	2	2.2	0.0589	0.02944	38.909	1.1454	0.0513	0.0257	38.909	0.9981	0.055	0.0275	38.909	1.0717	1.1454	0.9981	1.0717	1.0717	0.0737

* Las plantas de 1 a 20 pertenecen al sitio de Chamilpa, las plantas de 21 a 40 pertenecen al sitio de Alarcón y las plantas de 41 a 60 pertenecen al sitio de Felipe Neri.

* Las réplicas 1, 2 y 3 se representan como: R1, R2 y R3

Tabla suplementaria S3. Análisis de enriquecimiento Gene Ontology (GO) de las proteínas sobreexpresadas de *A. cordata*.

GO ID	GO Name	GO Category	FDR	<i>p</i> -value	Nr Test	Nr Reference	Non Annot Test	Non Annot Reference
GO:2000112	regulation of cellular macromolecule biosynthetic process	BIOLOGICAL_PROCESS	1	0.0304883	10	35	87	689
GO:0032508	DNA duplex unwinding	BIOLOGICAL_PROCESS	1	0.0382872	2	1	95	723
GO:0051171	regulation of nitrogen compound metabolic process	BIOLOGICAL_PROCESS	1	0.0484789	11	44	86	680
GO:0060255	regulation of macromolecule metabolic process	BIOLOGICAL_PROCESS	1	0.0398434	12	48	85	676
GO:0009889	regulation of biosynthetic process	BIOLOGICAL_PROCESS	1	0.0403574	10	37	87	687
GO:0006355	regulation of transcription, DNA-templated	BIOLOGICAL_PROCESS	1	0.0489432	9	33	88	691
GO:0019748	secondary metabolic process	BIOLOGICAL_PROCESS	1	0.0134048	3	2	94	722
GO:0031326	regulation of cellular biosynthetic process	BIOLOGICAL_PROCESS	1	0.0403574	10	37	87	687
GO:0080090	regulation of primary metabolic process	BIOLOGICAL_PROCESS	1	0.0484789	11	44	86	680
GO:2001141	regulation of RNA biosynthetic process	BIOLOGICAL_PROCESS	1	0.0489432	9	33	88	691
GO:1903506	regulation of nucleic acid-templated transcription	BIOLOGICAL_PROCESS	1	0.0489432	9	33	88	691
GO:0010556	regulation of macromolecule biosynthetic process	BIOLOGICAL_PROCESS	1	0.0304883	10	35	87	689
GO:0030162	regulation of proteolysis	BIOLOGICAL_PROCESS	1	0.0382872	2	1	95	723
GO:0032392	DNA geometric change	BIOLOGICAL_PROCESS	1	0.0382872	2	1	95	723
GO:0009941	chloroplast envelope	CELLULAR_COMPONENT	1	0.024524	3	3	94	721
GO:0009526	plastid envelope	CELLULAR_COMPONENT	1	0.024524	3	3	94	721
GO:0031625	ubiquitin protein ligase binding	MOLECULAR_FUNCTION	1	0.0382872	2	1	95	723
GO:0016462	pyrophosphatase activity	MOLECULAR_FUNCTION	1	0.0318954	8	25	89	699
GO:0016817	hydrolase activity, acting on acid anhydrides	MOLECULAR_FUNCTION	1	0.0442239	8	27	89	697
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	MOLECULAR_FUNCTION	1	0.0442239	8	27	89	697
GO:0003678	DNA helicase activity	MOLECULAR_FUNCTION	1	0.013832	2	0	95	724
GO:0016903	oxidoreductase activity, acting on the aldehyde or oxo group of donors	MOLECULAR_FUNCTION	1	0.0144599	4	5	93	719

GO:0005515	protein binding	MOLECULAR_FUNCTION	1	0.0038357	12	33	85	691
GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	MOLECULAR_FUNCTION	1	0.0048505	4	3	93	721
GO:0046914	transition metal ion binding	MOLECULAR_FUNCTION	1	0.0268831	14	56	83	668
GO:0004386	helicase activity	MOLECULAR_FUNCTION	1	0.0053812	5	6	92	718
GO:0019899	enzyme binding	MOLECULAR_FUNCTION	1	8.75E-04	5	3	92	721
GO:0030145	manganese ion binding	MOLECULAR_FUNCTION	1	0.0382872	2	1	95	723
GO:0044389	ubiquitin-like protein ligase binding	MOLECULAR_FUNCTION	1	0.0382872	2	1	95	723

Tabla suplementaria S4. Análisis de enriquecimiento Gene Ontology (GO) de las proteínas subexpresadas de *A. cordata*.

GO ID	GO Name	GO Category	FDR	p-value	Nr Test	Nr Reference	Non Annot Test	Non Annot Reference
GO:0044092	negative regulation of molecular function	BIOLOGICAL_PROCESS	1	0.0473677	3	2	141	627
GO:0035556	intracellular signal transduction	BIOLOGICAL_PROCESS	1	0.0206031	6	7	138	622
GO:0000165	MAPK cascade	BIOLOGICAL_PROCESS	1	0.0219299	3	1	141	628
GO:0051338	regulation of transferase activity	BIOLOGICAL_PROCESS	1	0.0473677	3	2	141	627
GO:0007154	cell communication	BIOLOGICAL_PROCESS	1	0.0179034	11	20	133	609
GO:0007165	signal transduction	BIOLOGICAL_PROCESS	1	0.0094455	10	15	134	614
GO:0046916	cellular transition metal ion homeostasis	BIOLOGICAL_PROCESS	1	0.0345066	2	0	142	629
GO:0032268	regulation of cellular protein metabolic process	BIOLOGICAL_PROCESS	1	0.0473677	3	2	141	627
GO:0023052	signaling	BIOLOGICAL_PROCESS	1	0.0094455	10	15	134	614
GO:0043086	negative regulation of catalytic activity	BIOLOGICAL_PROCESS	1	0.0473677	3	2	141	627
GO:0004674	protein serine/threonine kinase activity	MOLECULAR_FUNCTION	1	0.034522	7	11	137	618
GO:0016628	oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor	MOLECULAR_FUNCTION	1	0.043681	4	4	140	625
GO:0016772	transferase activity, transferring phosphorus-containing groups	MOLECULAR_FUNCTION	1	0.0303561	31	92	113	537

Tabla suplementaria S5. Análisis de enriquecimiento KEGG de las proteínas sobreexpresadas de *A. cordata*.

Pathway	#Enzymes Test	Pathway ID	Enzyme	#Test All	# Enzymes Ref	Enzyme	#Ref All	%Test	%Ref	<i>p</i> -value
Carbapenem biosynthesis	2	map00332	ec:1.2.1.41 - dehydrogenase, ec:2.7.2.11 - 5-kinase	125	2	ec:2.7.2.11 - 5-kinase, ec:1.2.1.41 - dehydrogenase	2122	1.6	0.0942507	0.0034439
Terpenoid backbone biosynthesis	3	map00900	ec:2.5.1.1 - geranyl-diphosphate synthase, ec:2.5.1.10 - diphosphate synthase, ec:2.3.1.9 - C-acetyltransferase	125	8	ec:1.1.1.34 - reductase (NADPH), ec:2.7.1.148 - 5'-diphospho)-2-C-methyl-D-erythritol kinase, ec:4.1.1.33 - decarboxylase, ec:2.5.1.10 - diphosphate synthase, ec:2.3.1.9 - C-acetyltransferase, ec:2.7.7.60 - 4-phosphate cytidyltransferase, ec:2.5.1.1 - geranyl-diphosphate synthase, ec:5.3.3.2 - Delta-isomerase	2122	2.4	0.3770028	0.008316
Valine, leucine and isoleucine degradation	3	map00280	ec:2.3.1.16 - C-acyltransferase, ec:1.2.1.27 - dehydrogenase (CoA-acylating), ec:2.3.1.9 - C-acetyltransferase	125	11	ec:1.2.1.3 - dehydrogenase (NAD ⁺), ec:3.1.2.4 - hydrolase, ec:1.2.1.27 - dehydrogenase (CoA-acylating), ec:1.1.1.35 - dehydrogenase, ec:1.8.1.4 - dehydrogenase, ec:2.3.1.9 - C-acetyltransferase, ec:6.4.1.4 - carboxylase, ec:4.2.1.17 - hydratase, ec:1.1.1.31 - dehydrogenase, ec:4.2.1.18 - hydratase, ec:2.3.1.16 - C-acyltransferase	2122	2.4	0.5183789	0.0204986
Biosynthesis of antibiotics	9	map01130	ec:1.2.1.41 - dehydrogenase, ec:2.5.1.1 - geranyl-diphosphate synthase, ec:5.1.3.15 - 1-epimerase, ec:2.3.1.16 - C-acyltransferase, ec:2.5.1.10 - diphosphate synthase, ec:2.3.1.9 - C-acetyltransferase, ec:1.2.1.59 - dehydrogenase (NAD(P) ⁺) (phosphorylating), ec:2.7.2.11 - 5-kinase, ec:1.2.1.12 -	125	82	ec:1.1.1.34 - reductase (NADPH), ec:2.7.2.11 - 5-kinase, ec:2.3.1.1 - N-acetyltransferase, ec:3.5.1.14 - acid amidohydrolase, ec:1.3.5.1 - dehydrogenase, ec:4.1.1.32 - carboxykinase (GTP), ec:1.2.1.3 - dehydrogenase (NAD ⁺), ec:1.5.1.2 - reductase, ec:4.1.1.39 - carboxylase, ec:1.2.4.1 - dehydrogenase (acetyl-transferring), ec:4.1.1.1 - decarboxylase, ec:6.4.1.2 - carboxylase, ec:1.1.1.183 - dehydrogenase (NADP ⁺), ec:2.3.1.12 - acetyltransferase, ec:2.7.1.148 - 5'-diphospho)-2-C-methyl-D-erythritol kinase, ec:1.4.1.14 - synthase (NADH), ec:1.1.1.35 - dehydrogenase, ec:2.7.4.3 - kinase, ec:2.6.1.1 - transaminase, ec:2.3.3.8 - citrate synthase, ec:1.1.1.44 - dehydrogenase (NADP ⁺ -dependent, decarboxylating), ec:4.1.1.33 - decarboxylase, ec:2.7.7.64 - uridylyltransferase, ec:5.1.3.15 - 1-epimerase, ec:2.7.2.3 - kinase, ec:2.5.1.54 - synthase, ec:1.2.4.2 - dehydrogenase (succinyl-transferring), ec:4.2.1.46 - 4,6-dehydratase, ec:2.5.1.10 - diphosphate synthase, ec:1.8.1.4 - dehydrogenase, ec:2.7.1.1 -	2122	7.2	3.864279	0.0285492

			dehydrogenase (phosphorylating)			hexokinase type IV glucokinase, ec:1.2.1.12 - dehydrogenase (phosphorylating), ec:2.7.7.9 - uridylyltransferase, ec:1.2.1.41 - dehydrogenase, ec:2.2.1.6 - synthase, ec:4.2.1.11 - hydratase, ec:2.3.1.9 - C-acetyltransferase, ec:3.5.3.1 - arginine amidinase, ec:1.1.1.1 - dehydrogenase, ec:4.2.1.20 - synthase, ec:4.3.1.19 - ammonia-lyase, ec:2.7.4.6 - kinase, ec:2.7.7.60 - 4-phosphate cytidylyltransferase, ec:1.1.1.86 - reductoisomerase (NADP+), ec:2.7.7.4 - adenylyltransferase, ec:1.1.1.37 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:2.7.4.10 - kinase, ec:1.1.1.27 - dehydrogenase, ec:1.1.1.95 - dehydrogenase, ec:5.3.1.6 - isomerase, ec:3.1.3.3 - phosphatase, ec:2.5.1.47 - synthase, ec:6.3.3.1 - cyclo-ligase, ec:5.4.2.8 - mannose phosphomutase, ec:1.1.1.41 - dehydrogenase (NAD+), ec:1.2.1.59 - dehydrogenase (NAD(P)+) (phosphorylating), ec:4.1.3.27 - synthase, ec:2.5.1.1 - geranyl-diphosphate synthase, ec:3.1.1.31 - phosphogluconolactonase, ec:1.11.1.6 - equilase, ec:2.1.2.1 - hydroxymethyltransferase, ec:4.2.1.3 - hydratase, ec:2.2.1.1 - glycolaldehydetransferase, ec:5.3.3.2 - Delta-isomerase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:6.2.1.4 - ligase (GDP-forming), ec:2.5.1.19 - 1-carboxyvinyltransferase, ec:1.1.1.42 - dehydrogenase (NADP+), ec:4.2.1.9 - dehydratase, ec:2.3.1.16 - C-acyltransferase, ec:5.3.1.1 - isomerase, ec:4.1.2.13 - aldolase, ec:1.1.1.49 - dehydrogenase (NADP+), ec:5.4.99.5 - mutase, ec:1.1.3.15 - oxidase, ec:2.3.3.1 - (Si)-synthase, ec:2.7.1.40 - kinase, ec:2.2.1.2 - dihydroxyacetone transferase, ec:5.3.1.8 - isomerase, ec:6.2.1.5 - ligase (ADP-forming), ec:2.3.1.30 - O-acetyltransferase				
Benzoate degradation	2	map00362	ec:2.3.1.16 - C-acyltransferase, ec:2.3.1.9 - C-acetyltransferase	125	5	ec:1.1.1.35 - dehydrogenase, ec:2.3.1.9 - C-acetyltransferase, ec:4.1.3.17 - aldolase, ec:4.2.1.17 - hydratase, ec:2.3.1.16 - C-acyltransferase	2122	1.6	0.2356268	0.0287829
Nitrotoluene degradation	1	map00633	ec:2.3.1.5 - N-acetyltransferase	125	1	ec:2.3.1.5 - N-acetyltransferase	2122	0.8	0.0471254	0.0589067
Ethylbenzene degradation	1	map00642	ec:2.3.1.16 - C-acyltransferase	125	1	ec:2.3.1.16 - C-acyltransferase	2122	0.8	0.0471254	0.0589067
Glycolysis / Gluconeogenesis	3	map00010	ec:5.1.3.15 - 1-epimerase, ec:1.2.1.59 - dehydrogenase	125	18	ec:4.1.1.32 - carboxykinase (GTP), ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.2.4.1 - dehydrogenase (acetyl-transferring), ec:4.1.1.1 - decarboxylase, ec:2.3.1.12 - acetyltransferase,	2122	2.4	0.8482564	0.0667868

			(NAD(P)+) (phosphorylating), ec:1.2.1.12 - dehydrogenase (phosphorylating)			ec:5.1.3.15 - 1-epimerase, ec:2.7.2.3 - kinase, ec:1.8.1.4 - dehydrogenase, ec:2.7.1.1 - hexokinase type IV glucokinase, ec:1.2.1.12 - dehydrogenase (phosphorylating), ec:4.2.1.11 - hydratase, ec:1.1.1.1 - dehydrogenase, ec:1.1.1.27 - dehydrogenase, ec:1.2.1.59 - dehydrogenase (NAD(P)+) (phosphorylating), ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:5.3.1.1 - isomerase, ec:4.1.2.13 - aldolase, ec:2.7.1.40 - kinase				
Drug metabolism - other enzymes	2	map00983	ec:2.3.1.5 - N- acetyltransferase, ec:3.1.1.1 - ali-esterase	125	8	ec:2.7.1.21 - kinase, ec:2.7.1.48 - kinase, ec:3.1.1.1 - ali-esterase, ec:2.7.4.6 - kinase, ec:3.2.1.31 - beta-glucuronide glucuronohydrolase glucuronidase, ec:2.5.1.18 - transferase, ec:2.3.1.5 - N-acetyltransferase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase	2122	1.6	0.3770028	0.0673391
Fatty acid degradation	2	map00071	ec:2.3.1.16 - C- acyltransferase, ec:2.3.1.9 - C-acetyltransferase	125	9	ec:1.2.1.3 - dehydrogenase (NAD+), ec:4.2.1.74 - hydratase, ec:1.1.1.35 - dehydrogenase, ec:2.3.1.9 - C-acetyltransferase, ec:1.1.1.1 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:6.2.1.3 - ligase, ec:2.3.1.16 - C-acyltransferase, ec:1.14.14.1 - monooxygenase	2122	1.6	0.4241282	0.0815414
Arginine and proline metabolism	2	map00330	ec:1.2.1.41 - dehydrogenase, ec:2.7.2.11 - 5-kinase	125	9	ec:2.7.2.11 - 5-kinase, ec:1.5.3.16 - oxidase, ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.5.1.2 - reductase, ec:2.6.1.1 - transaminase, ec:1.5.3.17 - polyamine oxidase, ec:3.5.3.11 - agmatine ureohydrolase, ec:1.2.1.41 - dehydrogenase, ec:3.5.3.1 - arginine amidinase	2122	1.6	0.4241282	0.0815414
Propanoate metabolism	2	map00640	ec:1.2.1.27 - dehydrogenase (CoA- acylating), ec:2.3.1.9 - C- acetyltransferase	125	9	ec:3.1.2.4 - hydrolase, ec:1.2.1.27 - dehydrogenase (CoA- acylating), ec:6.4.1.2 - carboxylase, ec:1.8.1.4 - dehydrogenase, ec:2.3.1.9 - C-acetyltransferase, ec:4.2.1.17 - hydratase, ec:1.1.1.27 - dehydrogenase, ec:6.2.1.4 - ligase (GDP-forming), ec:6.2.1.5 - ligase (ADP-forming)	2122	1.6	0.4241282	0.0815414
Caffeine metabolism	1	map00232	ec:2.3.1.5 - N- acetyltransferase	125	2	ec:2.3.1.5 - N-acetyltransferase, ec:1.14.14.1 - monooxygenase	2122	0.8	0.0942507	0.1109257
Synthesis and degradation of ketone bodies	1	map00072	ec:2.3.1.9 - C- acetyltransferase	125	2	ec:2.3.1.9 - C-acetyltransferase, ec:1.1.1.30 - dehydrogenase	2122	0.8	0.0942507	0.1109257
Cutin, suberine and wax biosynthesis	1	map00073	ec:1.2.1.84 - fatty acyl- CoA reductase	125	3	ec:2.3.1.20 - O-acyltransferase, ec:2.3.1.75 - O-fatty- acyltransferase, ec:1.2.1.84 - fatty acyl-CoA reductase	2122	0.8	0.1413761	0.1566563
Carbon fixation in	2	map00710	ec:1.2.1.59 - dehydrogenase	125	15	ec:4.1.1.39 - carboxylase, ec:2.6.1.1 - transaminase, ec:2.7.2.3 - kinase, ec:1.2.1.12 - dehydrogenase (phosphorylating), ec:1.1.1.40 -	2122	1.6	0.7068803	0.1658805

photosynthetic organisms			(NAD(P)+) (phosphorylating), ec:1.2.1.12 - dehydrogenase (phosphorylating)			dehydrogenase (oxaloacetate-decarboxylating) (NADP+), ec:3.1.3.37 - SBPase, ec:1.1.1.37 - dehydrogenase, ec:5.3.1.6 - isomerase, ec:1.2.1.13 - dehydrogenase (NADP+) (phosphorylating), ec:1.2.1.59 - dehydrogenase (NAD(P)+) (phosphorylating), ec:2.2.1.1 - glycolaldehydetransferase, ec:4.1.1.31 - carboxylase, ec:5.3.1.1 - isomerase, ec:4.1.2.13 - aldolase, ec:1.1.1.39 - dehydrogenase (decarboxylating)				
Purine metabolism	2	map00230	ec:3.6.1.3 - adenylypyrophosphatase, ec:3.6.1.15 - phosphatase	125	16	ec:3.6.1.15 - phosphatase, ec:2.7.7.7 - DNA polymerase, ec:2.7.4.3 - kinase, ec:3.6.1.9 - diphosphatase, ec:2.7.7.8 - nucleotidyltransferase, ec:3.6.1.3 - adenylypyrophosphatase, ec:2.7.4.6 - kinase, ec:2.7.7.4 - adenylyltransferase, ec:2.7.4.10 - kinase, ec:6.3.3.1 - cyclo-ligase, ec:3.6.1.5 - ATP-diphosphatase, ec:3.1.3.5 - uridine 5'-nucleotidase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:2.4.2.7 - phosphoribosyltransferase, ec:2.7.7.6 - RNA polymerase, ec:2.7.1.40 - kinase	2122	1.6	0.7540057	0.1785108
Fatty acid elongation	1	map00062	ec:2.3.1.16 - C-acyltransferase	125	4	ec:4.2.1.74 - hydratase, ec:1.1.1.35 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:2.3.1.16 - C-acyltransferase	2122	0.8	0.1885014	0.1966521
Thiamine metabolism	1	map00730	ec:3.6.1.15 - phosphatase	125	4	ec:3.6.1.15 - phosphatase, ec:2.5.1.3 - phosphate synthase, ec:2.7.4.3 - kinase, ec:3.1.3.2 - phosphatase	2122	0.8	0.1885014	0.1966521
Geraniol degradation	1	map00281	ec:2.3.1.16 - C-acyltransferase	125	4	ec:1.1.1.183 - dehydrogenase (NADP+), ec:1.1.1.35 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:2.3.1.16 - C-acyltransferase	2122	0.8	0.1885014	0.1966521
Biosynthesis of unsaturated fatty acids	1	map01040	ec:2.3.1.16 - C-acyltransferase	125	5	ec:1.1.1.100 - reductase, ec:1.14.19.2 - 9-desaturase, ec:4.2.1.17 - hydratase, ec:2.3.1.16 - C-acyltransferase, ec:1.14.19.6 - (9+3)-desaturase	2122	0.8	0.2356268	0.2314237
Ubiquinone and other terpenoid-quinone biosynthesis	1	map00130	ec:2.1.1.201 - methylase	125	5	ec:2.2.1.9 - synthase, ec:1.6.5.2 - dehydrogenase (quinone), ec:4.2.99.20 - synthase, ec:2.1.1.201 - methylase, ec:6.2.1.12 - ligase	2122	0.8	0.2356268	0.2314237
Lysine degradation	1	map00310	ec:2.3.1.9 - C-acetyltransferase	125	6	ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.1.1.35 - dehydrogenase, ec:1.2.4.2 - dehydrogenase (succinyl-transferring), ec:2.3.1.9 - C-acetyltransferase, ec:4.2.1.17 - hydratase, ec:2.1.1.43 - N-methyltransferase	2122	0.8	0.2827521	0.2614421
Butanoate metabolism	1	map00650	ec:2.3.1.9 - C-acetyltransferase	125	7	ec:1.3.5.1 - dehydrogenase, ec:1.1.1.35 - dehydrogenase, ec:2.2.1.6 - synthase, ec:2.3.1.9 - C-acetyltransferase, ec:1.1.1.30 -	2122	0.8	0.3298775	0.2871415

						dehydrogenase, ec:4.2.1.17 - hydratase, ec:1.2.1.16 - dehydrogenase [NAD(P)+]				
Galactose metabolism	1	map00052	ec:3.2.1.26 - invertase	125	8	ec:3.2.1.26 - invertase, ec:2.7.7.64 - uridylyltransferase, ec:3.2.1.20 - maltase, ec:2.7.1.1 - hexokinase type IV glucokinase, ec:2.7.7.9 - uridylyltransferase, ec:3.2.1.23 - lactase (ambiguous), ec:3.2.1.22 - melibiase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent)	2122	0.8	0.3770028	0.308922
Pentose and glucuronate interconversions	1	map00040	ec:3.2.1.15 - pectin depolymerase	125	8	ec:3.2.1.15 - pectin depolymerase, ec:2.7.7.64 - uridylyltransferase, ec:5.3.1.5 - isomerase, ec:2.7.7.9 - uridylyltransferase, ec:3.2.1.31 - beta-glucuronide glucuronohydrolase glucuronidase, ec:3.2.1.67 - 1,4-alpha-galacturonidase, ec:3.1.1.11 - pectin demethoxylase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase	2122	0.8	0.3770028	0.308922
Tryptophan metabolism	1	map00380	ec:2.3.1.9 - C-acetyltransferase	125	9	ec:3.5.1.9 - kynurenine formamidase, ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.1.1.35 - dehydrogenase, ec:1.2.4.2 - dehydrogenase (succinyl-transferring), ec:2.3.1.9 - C-acetyltransferase, ec:4.2.1.17 - hydratase, ec:1.11.1.6 - equilase, ec:4.1.1.28 - decarboxylase, ec:1.14.14.1 - monooxygenase	2122	0.8	0.4241282	0.3271519
alpha-Linolenic acid metabolism	1	map00592	ec:2.3.1.16 - C-acyltransferase	125	9	ec:4.2.1.92 - dehydratase, ec:1.3.1.42 - reductase, ec:5.3.99.6 - cyclase, ec:1.1.1.1 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:1.13.11.12 - 13S-lipoxygenase, ec:3.1.1.4 - A2, ec:2.3.1.16 - C-acyltransferase, ec:3.1.1.32 - A1	2122	0.8	0.4241282	0.3271519
Glycerolipid metabolism	1	map00561	ec:3.1.1.23 - lipase	125	9	ec:2.3.1.20 - O-acyltransferase, ec:1.2.1.3 - dehydrogenase (NAD+), ec:3.1.1.23 - lipase, ec:2.3.1.22 - O-acyltransferase, ec:2.3.1.51 - O-acyltransferase, ec:2.7.7.9 - uridylyltransferase, ec:3.1.1.3 - lipase, ec:2.3.1.15 - 1-O-acyltransferase, ec:3.2.1.22 - melibiase	2122	0.8	0.4241282	0.3271519
Carbon fixation pathways in prokaryotes	1	map00720	ec:2.3.1.9 - C-acetyltransferase	125	11	ec:1.3.5.1 - dehydrogenase, ec:6.4.1.2 - carboxylase, ec:1.1.1.35 - dehydrogenase, ec:2.3.3.8 - citrate synthase, ec:2.3.1.9 - C-acetyltransferase, ec:1.1.1.37 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:4.2.1.3 - hydratase, ec:1.1.1.42 - dehydrogenase (NADP+), ec:4.1.1.31 - carboxylase, ec:6.2.1.5 - ligase (ADP-forming)	2122	0.8	0.5183789	0.3542888
Alanine, aspartate and glutamate metabolism	1	map00250	ec:6.3.5.4 - synthase (glutamine-hydrolysing)	125	11	ec:3.5.1.111 - amidase, ec:1.4.1.14 - synthase (NADH), ec:2.6.1.1 - transaminase, ec:2.1.3.2 - carbamoyltransferase, ec:1.4.1.2 - dehydrogenase, ec:1.4.1.3 - dehydrogenase [NAD(P)+], ec:6.3.5.4 - synthase (glutamine-hydrolysing), ec:6.3.1.2 - synthetase, ec:3.5.1.3	2122	0.8	0.5183789	0.3542888

						- alpha-keto acid-omega-amidase, ec:1.2.1.16 - dehydrogenase [NAD(P)+], ec:1.4.1.4 - dehydrogenase (NADP+)				
Starch and sucrose metabolism	1	map00500	ec:3.2.1.26 - invertase	125	23	ec:3.2.1.4 - endo-1,4-beta-D-glucanase, ec:2.4.1.13 - synthase, ec:3.2.1.26 - invertase, ec:3.6.1.9 - diphosphatase, ec:2.4.1.34 - synthase, ec:2.4.1.14 - synthase, ec:3.2.1.20 - maltase, ec:2.7.7.27 - adenylyltransferase, ec:2.7.1.1 - hexokinase type IV glucokinase, ec:2.7.7.9 - uridylyltransferase, ec:2.4.1.1 - phosphorylase, ec:2.4.1.25 - disproportionating enzyme, ec:3.2.1.21 - gentiobiase, ec:3.2.1.1 - glycogenase, ec:3.1.3.12 - trehalose 6-phosphatase, ec:3.2.1.2 - saccharogen amylase, ec:2.4.1.21 - synthase (glycosyl-transferring), ec:3.2.1.39 - endo-1,3-beta-D-glucosidase, ec:2.4.1.11 - synthase, ec:3.2.1.48 - alpha-glucosidase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:2.4.1.12 - synthase (UDP-forming), ec:2.7.1.4 - fructokinase (phosphorylating)	2122	0.8	1.0838831	0.3575671
Glyoxylate and dicarboxylate metabolism	1	map00630	ec:2.3.1.9 - C-acetyltransferase	125	14	ec:3.5.1.9 - kynurenine formamidase, ec:1.1.1.79 - reductase (NADP+), ec:4.1.1.39 - carboxylase, ec:4.1.1.8 - decarboxylase, ec:2.3.1.9 - C-acetyltransferase, ec:1.1.1.37 - dehydrogenase, ec:1.1.99.14 - dehydrogenase, ec:1.11.1.6 - equilase, ec:2.1.2.1 - hydroxymethyltransferase, ec:4.2.1.3 - hydratase, ec:1.1.1.81 - reductase, ec:6.3.1.2 - synthetase, ec:1.1.3.15 - oxidase, ec:2.3.3.1 - (Si)-synthase	2122	0.8	0.6597549	0.3759959
Pyruvate metabolism	1	map00620	ec:2.3.1.9 - C-acetyltransferase	125	18	ec:4.1.1.32 - carboxykinase (GTP), ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.1.1.79 - reductase (NADP+), ec:1.2.4.1 - dehydrogenase (acetyl-transferring), ec:6.4.1.2 - carboxylase, ec:2.3.1.12 - acetyltransferase, ec:1.8.1.4 - dehydrogenase, ec:2.3.1.9 - C-acetyltransferase, ec:1.1.2.4 - dehydrogenase (cytochrome), ec:1.1.1.40 - dehydrogenase (oxaloacetate-decarboxylating) (NADP+), ec:1.1.1.37 - dehydrogenase, ec:1.1.1.27 - dehydrogenase, ec:1.1.1.38 - dehydrogenase (oxaloacetate-decarboxylating), ec:1.1.1.28 - dehydrogenase, ec:4.4.1.5 - lyase, ec:4.1.1.31 - carboxylase, ec:1.1.1.39 - dehydrogenase (decarboxylating), ec:2.7.1.40 - kinase	2122	0.8	0.8482564	0.3792577

Tabla suplementaria S6. Análisis de enriquecimiento KEGG de las proteínas subexpresadas de *A. cordata*.

Pathway	#Enzymes Test	Pathway ID	Enzyme	#Test All	# Enzymes Ref	Enzyme	#Ref All	%Test	%Ref	<i>p</i> -value
Pyruvate metabolism	13	map00620	ec:4.1.1.3 - decarboxylase, ec:1.1.1.37 - dehydrogenase, ec:4.4.1.5 - lyase, ec:1.1.1.39 - dehydrogenase (decarboxylating), ec:1.1.1.38 - dehydrogenase (oxaloacetate-decarboxylating), ec:2.3.1.12 - acetyltransferase, ec:4.1.1.31 - carboxylase, ec:4.1.1.32 - carboxykinase (GTP), ec:1.1.1.40 - dehydrogenase (oxaloacetate-decarboxylating) (NADP+), ec:1.1.1.28 - dehydrogenase, ec:1.1.2.4 - dehydrogenase (cytochrome), ec:2.7.1.40 - kinase, ec:1.1.1.79 - reductase (NADP+)	491	18	ec:4.1.1.32 - carboxykinase (GTP), ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.1.1.79 - reductase (NADP+), ec:1.2.4.1 - dehydrogenase (acetyl-transferring), ec:6.4.1.2 - carboxylase, ec:2.3.1.12 - acetyltransferase, ec:1.8.1.4 - dehydrogenase, ec:2.3.1.9 - C-acetyltransferase, ec:1.1.2.4 - dehydrogenase (cytochrome), ec:1.1.1.40 - dehydrogenase (oxaloacetate-decarboxylating) (NADP+), ec:1.1.1.37 - dehydrogenase, ec:1.1.1.27 - dehydrogenase, ec:1.1.1.38 - dehydrogenase (oxaloacetate-decarboxylating), ec:1.1.1.28 - dehydrogenase, ec:4.4.1.5 - lyase, ec:4.1.1.31 - carboxylase, ec:1.1.1.39 - dehydrogenase (decarboxylating), ec:2.7.1.40 - kinase	2122	2.6476578	0.8482564	1.1401E-05
Purine metabolism	10	map00230	ec:3.6.1.15 - phosphatase, ec:2.7.4.3 - kinase, ec:3.1.3.5 - uridine 5'-nucleotidase, ec:3.6.1.3 - adenylypyrophosphatase, ec:2.7.7.4 - adenylyltransferase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:2.7.7.6 - RNA polymerase, ec:2.7.7.7 - DNA polymerase, ec:2.7.7.8 - nucleotidyltransferase, ec:2.7.1.40 - kinase	491	16	ec:3.6.1.15 - phosphatase, ec:2.7.7.7 - DNA polymerase, ec:2.7.4.3 - kinase, ec:3.6.1.9 - diphosphatase, ec:2.7.7.8 - nucleotidyltransferase, ec:3.6.1.3 - adenylypyrophosphatase, ec:2.7.4.6 - kinase, ec:2.7.7.4 - adenylyltransferase, ec:2.7.4.10 - kinase, ec:6.3.3.1 - cyclo-ligase, ec:3.6.1.5 - ATP-diphosphatase, ec:3.1.3.5 - uridine 5'-nucleotidase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:2.4.2.7 - phosphoribosyltransferase, ec:2.7.7.6 - RNA polymerase, ec:2.7.1.40 - kinase	2122	2.0366599	0.7540057	0.000694537
Glycolysis / Gluconeogenesis	10	map00010	ec:4.1.1.1 - decarboxylase, ec:5.3.1.1 - isomerase,	491	18	ec:4.1.1.32 - carboxykinase (GTP), ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.2.4.1 - dehydrogenase (acetyl-transferring), ec:4.1.1.1 - decarboxylase,	2122	2.0366599	0.8482564	0.002259216

			ec:4.1.2.13 - aldolase, ec:1.1.1.1 - dehydrogenase, ec:2.3.1.12 - acetyltransferase, ec:4.1.1.32 - carboxykinase (GTP), ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:2.7.1.40 - kinase, ec:2.7.1.1 - hexokinase type IV glucokinase, ec:1.2.1.9 - dehydrogenase (NADP+)			ec:2.3.1.12 - acetyltransferase, ec:5.1.3.15 - 1-epimerase, ec:2.7.2.3 - kinase, ec:1.8.1.4 - dehydrogenase, ec:2.7.1.1 - hexokinase type IV glucokinase, ec:1.2.1.12 - dehydrogenase (phosphorylating), ec:4.2.1.11 - hydratase, ec:1.1.1.1 - dehydrogenase, ec:1.1.1.27 - dehydrogenase, ec:1.2.1.59 - dehydrogenase (NAD(P)+) (phosphorylating), ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:5.3.1.1 - isomerase, ec:4.1.2.13 - aldolase, ec:2.7.1.40 - kinase				
Thiamine metabolism	4	map00730	ec:2.5.1.3 - phosphate synthase, ec:3.6.1.15 - phosphatase, ec:2.7.4.3 - kinase, ec:3.1.3.2 - phosphatase	491	4	ec:3.6.1.15 - phosphatase, ec:2.5.1.3 - phosphate synthase, ec:2.7.4.3 - kinase, ec:3.1.3.2 - phosphatase	2122	0.814664	0.1885014	0.002839575
Glyoxylate and dicarboxylate metabolism	8	map00630	ec:2.1.2.1 - hydroxymethyltransferase, ec:1.1.1.37 - dehydrogenase, ec:6.3.1.2 - synthetase, ec:1.1.99.14 - dehydrogenase, ec:1.1.1.79 - reductase (NADP+), ec:1.1.3.15 - oxidase, ec:1.1.1.81 - reductase, ec:3.5.1.9 - kynurenine formamidase	491	14	ec:3.5.1.9 - kynurenine formamidase, ec:1.1.1.79 - reductase (NADP+), ec:4.1.1.39 - carboxylase, ec:4.1.1.8 - decarboxylase, ec:2.3.1.9 - C-acetyltransferase, ec:1.1.1.37 - dehydrogenase, ec:1.1.99.14 - dehydrogenase, ec:1.1.1.6 - equilase, ec:2.1.2.1 - hydroxymethyltransferase, ec:4.2.1.3 - hydratase, ec:1.1.1.81 - reductase, ec:6.3.1.2 - synthetase, ec:1.1.3.15 - oxidase, ec:2.3.3.1 - (Si)-synthase	2122	1.6293279	0.6597549	0.004968985
Oxidative phosphorylation	4	map00190	ec:1.9.3.1 - oxidase, ec:1.6.99.3 - dehydrogenase, ec:1.6.5.3 - reductase (H+-translocating), ec:3.6.3.6 - ATPase	491	5	ec:3.6.1.1 - diphosphatase, ec:1.3.5.1 - dehydrogenase, ec:1.9.3.1 - oxidase, ec:1.10.2.2 - reductase, ec:1.6.99.3 - dehydrogenase	2122	0.814664	0.2356268	0.010933303
Steroid hormone biosynthesis	4	map00140	ec:1.3.99.6 - 4-dehydrogenase, ec:1.3.1.3 - 5beta-reductase, ec:1.1.1.145 - dehydrogenase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase	491	5	ec:1.1.1.145 - dehydrogenase, ec:1.3.1.3 - 5beta-reductase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase, ec:1.3.99.6 - 4-dehydrogenase, ec:1.14.14.1 - monooxygenase	2122	0.814664	0.2356268	0.010933303
Galactose metabolism	5	map00052	ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:3.2.1.20 - maltase, ec:3.2.1.26 - invertase, ec:3.2.1.23 - lactase (ambiguous), ec:2.7.1.1 - hexokinase type IV glucokinase	491	8	ec:3.2.1.26 - invertase, ec:2.7.7.64 - uridylyltransferase, ec:3.2.1.20 - maltase, ec:2.7.1.1 - hexokinase type IV glucokinase, ec:2.7.7.9 - uridylyltransferase, ec:3.2.1.23 - lactase (ambiguous), ec:3.2.1.22 - melibiase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent)	2122	1.0183299	0.3770028	0.016713222

Drug metabolism - other enzymes	5	map00983	ec:2.3.1.5 - N-acetyltransferase, ec:2.5.1.18 - transferase, ec:2.7.1.48 - kinase, ec:3.1.1.1 - ali-esterase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase	491	8	ec:2.7.1.21 - kinase, ec:2.7.1.48 - kinase, ec:3.1.1.1 - ali-esterase, ec:2.7.4.6 - kinase, ec:3.2.1.31 - beta-glucuronide glucuronohydrolase glucuronidase, ec:2.5.1.18 - transferase, ec:2.3.1.5 - N-acetyltransferase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase	2122	1.0183299	0.3770028	0.016713222
Pyrimidine metabolism	5	map00240	ec:3.1.3.5 - uridine 5'-nucleotidase, ec:2.7.7.6 - RNA polymerase, ec:2.7.7.7 - DNA polymerase, ec:2.7.7.8 - nucleotidyltransferase, ec:2.7.1.48 - kinase	491	8	ec:5.4.3.8 - 2,1-aminomutase, ec:1.2.1.70 - reductase, ec:4.2.1.75 - synthase, ec:1.3.3.3 - oxidase, ec:3.2.1.31 - beta-glucuronide glucuronohydrolase glucuronidase, ec:1.14.13.81 - IX monomethyl ester (oxidative) cyclase, ec:1.16.3.1 - ceruloplasmin, ec:2.4.1.17 - 1-naphthol glucuronyltransferase	2122	1.0183299	0.3770028	0.016713222
Arginine biosynthesis	5	map00220	ec:2.3.1.1 - N-acetyltransferase, ec:3.5.3.1 - arginine amidinase, ec:6.3.1.2 - synthetase, ec:1.4.1.3 - dehydrogenase [NAD(P)+], ec:1.4.1.2 - dehydrogenase	491	8	ec:2.3.1.1 - N-acetyltransferase, ec:3.5.1.14 - acid amidohydrolase, ec:2.6.1.1 - transaminase, ec:3.5.3.1 - arginine amidinase, ec:1.4.1.2 - dehydrogenase, ec:1.4.1.3 - dehydrogenase [NAD(P)+], ec:6.3.1.2 - synthetase, ec:1.4.1.4 - dehydrogenase (NADP+)	2122	1.0183299	0.3770028	0.016713222
Pentose phosphate pathway	5	map00030	ec:4.1.2.13 - aldolase, ec:1.1.1.44 - dehydrogenase (NADP+-dependent, decarboxylating), ec:1.1.1.43 - 2-dehydrogenase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:1.2.1.9 - dehydrogenase (NADP+)	491	8	ec:1.1.1.44 - dehydrogenase (NADP+-dependent, decarboxylating), ec:5.3.1.6 - isomerase, ec:3.1.1.31 - phosphogluconolactonase, ec:2.2.1.1 - glycolaldehydetransferase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:4.1.2.13 - aldolase, ec:1.1.1.49 - dehydrogenase (NADP+), ec:2.2.1.2 - dihydroxyacetone transferase	2122	1.0183299	0.3770028	0.016713222
Amino sugar and nucleotide sugar metabolism	7	map00520	ec:5.3.1.8 - isomerase, ec:4.1.1.35 - decarboxylase, ec:5.4.99.30 - mutase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:1.1.1.22 - 6-dehydrogenase, ec:2.7.7.27 - adenylyltransferase, ec:2.7.1.1 - hexokinase type IV glucokinase	491	14	ec:3.2.1.55 - end alpha-L-arabinofuranosidase, ec:2.7.7.64 - uridylyltransferase, ec:5.4.99.30 - mutase, ec:2.7.7.27 - adenylyltransferase, ec:2.7.1.1 - hexokinase type IV glucokinase, ec:2.7.7.9 - uridylyltransferase, ec:4.2.1.76 - 4,6-dehydratase, ec:4.1.1.35 - decarboxylase, ec:5.4.2.8 - mannose phosphomutase, ec:3.2.1.52 - hexosaminidase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:1.6.2.2 - reductase, ec:5.3.1.8 - isomerase, ec:2.7.1.4 - fructokinase (phosphorylating)	2122	1.4256619	0.6597549	0.019066353
Biosynthesis of antibiotics	25	map01130	ec:1.1.1.183 - dehydrogenase (NADP+), ec:2.1.2.1 - hydroxymethyltransferase, ec:4.1.1.1 - decarboxylase, ec:5.3.1.1 - isomerase,	491	82	ec:1.1.1.34 - reductase (NADPH), ec:2.7.2.11 - 5-kinase, ec:2.3.1.1 - N-acetyltransferase, ec:3.5.1.14 - acid amidohydrolase, ec:1.3.5.1 - dehydrogenase, ec:4.1.1.32 - carboxykinase (GTP), ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.5.1.2 - reductase, ec:4.1.1.39 - carboxylase, ec:1.2.4.1 - dehydrogenase (acetyl-transferring), ec:4.1.1.1 - decarboxylase,	2122	5.0916497	3.864279	0.028915996

			<p>ec:5.3.1.8 - isomerase, ec:4.1.2.13 - aldolase, ec:2.7.4.3 - kinase, ec:2.3.1.1 - N-acetyltransferase, ec:1.1.1.37 - dehydrogenase, ec:1.1.1.1 - dehydrogenase, ec:2.3.1.12 - acetyltransferase, ec:4.1.1.32 - carboxykinase (GTP), ec:1.1.1.44 - dehydrogenase (NADP+-dependent, decarboxylating), ec:2.7.7.4 - adenylyltransferase, ec:1.1.1.41 - dehydrogenase (NAD+), ec:4.2.1.9 - hydratase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:3.5.3.1 - arginine amidinase, ec:1.2.4.2 - dehydrogenase (succinyl-transferring), ec:4.2.1.17 - hydratase, ec:2.7.1.40 - kinase, ec:1.1.3.15 - oxidase, ec:1.3.1.13 - dehydrogenase (NADP+), ec:1.3.1.12 - dehydrogenase, ec:2.7.1.1 - hexokinase type IV glucokinase</p>		<p>ec:6.4.1.2 - carboxylase, ec:1.1.1.183 - dehydrogenase (NADP+), ec:2.3.1.12 - acetyltransferase, ec:2.7.1.148 - 5'-diphospho)-2-C-methyl-D-erythritol kinase, ec:1.4.1.14 - synthase (NADH), ec:1.1.1.35 - dehydrogenase, ec:2.7.4.3 - kinase, ec:2.6.1.1 - transaminase, ec:2.3.3.8 - citrate synthase, ec:1.1.1.44 - dehydrogenase (NADP+-dependent, decarboxylating), ec:4.1.1.33 - decarboxylase, ec:2.7.7.64 - uridylyltransferase, ec:5.1.3.15 - 1-epimerase, ec:2.7.2.3 - kinase, ec:2.5.1.54 - synthase, ec:1.2.4.2 - dehydrogenase (succinyl-transferring), ec:4.2.1.46 - 4,6-dehydratase, ec:2.5.1.10 - diphosphate synthase, ec:1.8.1.4 - dehydrogenase, ec:2.7.1.1 - hexokinase type IV glucokinase, ec:1.2.1.12 - dehydrogenase (phosphorylating), ec:2.7.7.9 - uridylyltransferase, ec:1.2.1.41 - dehydrogenase, ec:2.2.1.6 - synthase, ec:4.2.1.11 - hydratase, ec:2.3.1.9 - C-acetyltransferase, ec:3.5.3.1 - arginine amidinase, ec:1.1.1.1 - dehydrogenase, ec:4.2.1.20 - synthase, ec:4.3.1.19 - ammonia-lyase, ec:2.7.4.6 - kinase, ec:2.7.7.60 - 4-phosphate cytidylyltransferase, ec:1.1.1.86 - reductoisomerase (NADP+), ec:2.7.7.4 - adenylyltransferase, ec:1.1.1.37 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:2.7.4.10 - kinase, ec:1.1.1.27 - dehydrogenase, ec:1.1.1.95 - dehydrogenase, ec:5.3.1.6 - isomerase, ec:3.1.3.3 - phosphatase, ec:2.5.1.47 - synthase, ec:6.3.3.1 - cyclo-ligase, ec:5.4.2.8 - mannose phosphomutase, ec:1.1.1.41 - dehydrogenase (NAD+), ec:1.2.1.59 - dehydrogenase (NAD(P)+) (phosphorylating), ec:4.1.3.27 - synthase, ec:2.5.1.1 - geranyl-diphosphate synthase, ec:3.1.1.31 - phosphogluconolactonase, ec:1.11.1.6 - equilase, ec:2.1.2.1 - hydroxymethyltransferase, ec:4.2.1.3 - hydratase, ec:2.2.1.1 - glycolaldehydetransferase, ec:5.3.3.2 - Delta-isomerase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:6.2.1.4 - ligase (GDP-forming), ec:2.5.1.19 - 1-carboxyvinyltransferase, ec:1.1.1.42 - dehydrogenase (NADP+), ec:4.2.1.9 - dehydratase, ec:2.3.1.16 - C-acyltransferase, ec:5.3.1.1 - isomerase, ec:4.1.2.13 - aldolase, ec:1.1.1.49 - dehydrogenase (NADP+), ec:5.4.99.5 - mutase, ec:1.1.3.15 - oxidase, ec:2.3.3.1 - (Si)-synthase, ec:2.7.1.40 - kinase, ec:2.2.1.2 - dihydroxyacetone transferase, ec:5.3.1.8 - isomerase, ec:6.2.1.5 - ligase (ADP-forming), ec:2.3.1.30 - O-acetyltransferase</p>					
Starch and sucrose metabolism	9	map00500	<p>ec:3.2.1.48 - alpha-glucosidase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:3.2.1.20 - maltase, ec:3.2.1.21 - gentiobiase, ec:3.2.1.26 - invertase, ec:2.7.7.27 - adenylyltransferase, ec:3.2.1.39</p>	491	23	<p>ec:3.2.1.4 - endo-1,4-beta-D-glucanase, ec:2.4.1.13 - synthase, ec:3.2.1.26 - invertase, ec:3.6.1.9 - diphosphatase, ec:2.4.1.34 - synthase, ec:2.4.1.14 - synthase, ec:3.2.1.20 - maltase, ec:2.7.7.27 - adenylyltransferase, ec:2.7.1.1 - hexokinase type IV glucokinase, ec:2.7.7.9 - uridylyltransferase, ec:2.4.1.1 - phosphorylase, ec:2.4.1.25 - disproportionating enzyme, ec:3.2.1.21 - gentiobiase, ec:3.2.1.1 - glycogenase, ec:3.1.3.12 - trehalose 6-phosphatase, ec:3.2.1.2 - saccharogen amylase, ec:2.4.1.21 - synthase (glycosyl-transferring), ec:3.2.1.39 - endo-1,3-beta-D-glucosidase, ec:2.4.1.11 -</p>	2122	1.8329939	1.0838831	0.038628336

			- endo-1,3-beta-D-glucosidase, ec:2.4.1.13 - synthase, ec:2.7.1.1 - hexokinase type IV glucokinase			synthase, ec:3.2.1.48 - alpha-glucosidase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:2.4.1.12 - synthase (UDP-forming), ec:2.7.1.4 - fructokinase (phosphorylating)				
Nitrogen metabolism	4	map00910	ec:4.2.1.1 - anhydrase, ec:6.3.1.2 - synthetase, ec:1.4.1.3 - dehydrogenase [NAD(P)+], ec:1.4.1.2 - dehydrogenase	491	7	ec:1.4.1.14 - synthase (NADH), ec:4.2.1.1 - anhydrase, ec:1.4.1.2 - dehydrogenase, ec:4.2.1.104 - cyanate lyase, ec:1.4.1.3 - dehydrogenase [NAD(P)+], ec:6.3.1.2 - synthetase, ec:1.4.1.4 - dehydrogenase (NADP+)	2122	0.814664	0.3298775	0.045365072
Cyanoamino acid metabolism	2	map00460	ec:2.1.2.1 - hydroxymethyltransferase, ec:3.2.1.21 - gentiobiase	491	2	ec:3.2.1.21 - gentiobiase, ec:2.1.2.1 - hydroxymethyltransferase	2122	0.407332	0.0942507	0.053455393
Pentose and glucuronate interconversions	4	map00040	ec:3.1.1.11 - pectin demethoxylase, ec:1.1.1.22 - 6-dehydrogenase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase, ec:3.2.1.15 - pectin depolymerase	491	8	ec:3.2.1.15 - pectin depolymerase, ec:2.7.7.64 - uridylyltransferase, ec:5.3.1.5 - isomerase, ec:2.7.7.9 - uridylyltransferase, ec:3.2.1.31 - beta-glucuronide glucuronohydrolase glucuronidase, ec:3.2.1.67 - 1,4-alpha-galacturonidase, ec:3.1.1.11 - pectin demethoxylase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase	2122	0.814664	0.3770028	0.069838617
Carbon fixation in photosynthetic organisms	6	map00710	ec:5.3.1.1 - isomerase, ec:4.1.2.13 - aldolase, ec:1.1.1.37 - dehydrogenase, ec:1.1.1.39 - dehydrogenase (decarboxylating), ec:4.1.1.31 - carboxylase, ec:1.1.1.40 - dehydrogenase (oxaloacetate-decarboxylating) (NADP+)	491	15	ec:4.1.1.39 - carboxylase, ec:2.6.1.1 - transaminase, ec:2.7.2.3 - kinase, ec:1.2.1.12 - dehydrogenase (phosphorylating), ec:1.1.1.40 - dehydrogenase (oxaloacetate-decarboxylating) (NADP+), ec:3.1.3.37 - SBPase, ec:1.1.1.37 - dehydrogenase, ec:5.3.1.6 - isomerase, ec:1.2.1.13 - dehydrogenase (NADP+) (phosphorylating), ec:1.2.1.59 - dehydrogenase (NAD(P)+) (phosphorylating), ec:2.2.1.1 - glycolaldehydetransferase, ec:4.1.1.31 - carboxylase, ec:5.3.1.1 - isomerase, ec:4.1.2.13 - aldolase, ec:1.1.1.39 - dehydrogenase (decarboxylating)	2122	1.2219959	0.7068803	0.071683077
Metabolism of xenobiotics by cytochrome P450	3	map00980	ec:2.5.1.18 - transferase, ec:1.1.1.1 - dehydrogenase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase	491	5	ec:1.1.1.1 - dehydrogenase, ec:2.5.1.18 - transferase, ec:3.3.2.9 - epoxide hydrolase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase, ec:1.14.14.1 - monooxygenase	2122	0.610998	0.2356268	0.073038051
Drug metabolism - cytochrome P450	3	map00982	ec:2.5.1.18 - transferase, ec:1.1.1.1 - dehydrogenase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase	491	5	ec:1.1.1.1 - dehydrogenase, ec:1.14.13.8 - monooxygenase, ec:2.5.1.18 - transferase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase, ec:1.14.14.1 - monooxygenase	2122	0.610998	0.2356268	0.073038051
Phenylpropanoid biosynthesis	4	map00940	ec:1.1.1.195 - dehydrogenase, ec:2.1.1.68 - O-methyltransferase, ec:3.2.1.21 - gentiobiase, ec:1.11.1.7 - lactoperoxidase	491	9	ec:1.1.1.195 - dehydrogenase, ec:3.2.1.21 - gentiobiase, ec:2.3.1.133 - O-hydroxycinnamoyltransferase, ec:4.3.1.24 - ammonia-lyase, ec:2.1.1.68 - O-methyltransferase, ec:6.2.1.12 - ligase, ec:1.11.1.7 - lactoperoxidase, ec:2.1.1.104 - O-methyltransferase, ec:4.3.1.25 - ammonia-lyase	2122	0.814664	0.4241282	0.09674994

Phenylalanine, tyrosine and tryptophan biosynthesis	3	map00400	ec:1.3.1.78 - dehydrogenase (NADP+), ec:1.3.1.13 - dehydrogenase (NADP+), ec:1.3.1.12 - dehydrogenase	491	6	ec:2.6.1.1 - transaminase, ec:2.5.1.54 - synthase, ec:4.2.1.20 - synthase, ec:4.1.3.27 - synthase, ec:2.5.1.19 - 1-carboxyvinyltransferase, ec:5.4.99.5 - mutase	2122	0.610998	0.2827521	0.112403387
Ascorbate and aldarate metabolism	3	map00053	ec:1.6.5.4 - reductase (NADH), ec:1.1.1.22 - 6-dehydrogenase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase	491	6	ec:1.2.1.3 - dehydrogenase (NAD+), ec:2.7.7.64 - uridylyltransferase, ec:1.6.5.4 - reductase (NADH), ec:4.1.2.20 - aldolase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase, ec:1.10.3.3 - oxidase	2122	0.610998	0.2827521	0.112403387
Selenocompound metabolism	2	map00450	ec:2.7.7.4 - adenylyltransferase, ec:2.1.1.14 - S-methyltransferase	491	3	ec:1.8.1.9 - reductase, ec:2.1.1.14 - S-methyltransferase, ec:2.7.7.4 - adenylyltransferase	2122	0.407332	0.1413761	0.123376055
Aminobenzoate degradation	2	map00627	ec:3.1.3.41 - nitrophenyl phosphatase, ec:4.2.1.17 - hydratase	491	3	ec:4.2.1.17 - hydratase, ec:3.1.3.41 - nitrophenyl phosphatase, ec:1.14.14.1 - monooxygenase	2122	0.407332	0.1413761	0.123376055
Flavone and flavonol biosynthesis	2	map00944	ec:2.1.1.42 - 3'-O-methyltransferase, ec:2.1.1.76 - 3-O-methyltransferase	491	3	ec:3.2.1.31 - beta-glucuronide glucuronohydrolase glucuronidase, ec:2.1.1.76 - 3-O-methyltransferase, ec:2.1.1.42 - 3'-O-methyltransferase	2122	0.407332	0.1413761	0.123376055
Diterpenoid biosynthesis	2	map00904	ec:1.14.11.13 - 2beta-dioxygenase, ec:4.2.3.19 - synthase	491	3	ec:5.5.1.13 - diphosphate synthase, ec:4.2.3.19 - synthase, ec:1.14.11.13 - 2beta-dioxygenase	2122	0.407332	0.1413761	0.123376055
Cutin, suberine and wax biosynthesis	2	map00073	ec:2.3.1.75 - O-fatty-acyltransferase, ec:2.3.1.20 - O-acyltransferase	491	3	ec:2.3.1.20 - O-acyltransferase, ec:2.3.1.75 - O-fatty-acyltransferase, ec:1.2.1.84 - fatty acyl-CoA reductase	2122	0.407332	0.1413761	0.123376055
Streptomycin biosynthesis	2	map00521	ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent), ec:2.7.1.1 - hexokinase type IV glucokinase	491	3	ec:4.2.1.46 - 4,6-dehydratase, ec:2.7.1.1 - hexokinase type IV glucokinase, ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent)	2122	0.407332	0.1413761	0.123376055
Fructose and mannose metabolism	4	map00051	ec:5.3.1.1 - isomerase, ec:5.3.1.8 - isomerase, ec:4.1.2.13 - aldolase, ec:2.7.1.1 - hexokinase type IV glucokinase	491	10	ec:4.1.2.53 - aldolase, ec:5.3.1.5 - isomerase, ec:2.7.1.1 - hexokinase type IV glucokinase, ec:3.2.1.80 - beta-fructosidase, ec:5.3.1.6 - isomerase, ec:5.4.2.8 - mannose phosphomutase, ec:5.3.1.1 - isomerase, ec:4.1.2.13 - aldolase, ec:5.3.1.8 - isomerase, ec:2.7.1.4 - fructokinase (phosphorylating)	2122	0.814664	0.4712535	0.124085347
Methane metabolism	4	map00680	ec:2.1.2.1 - hydroxymethyltransferase, ec:4.1.2.13 - aldolase, ec:1.1.1.37 - dehydrogenase, ec:4.1.1.31 - carboxylase	491	10	ec:4.2.1.11 - hydratase, ec:1.1.1.37 - dehydrogenase, ec:1.1.1.95 - dehydrogenase, ec:1.12.98.1 - F420 hydrogenase, ec:3.1.3.3 - phosphatase, ec:2.1.2.1 - hydroxymethyltransferase, ec:1.1.1.284 - dehydrogenase, ec:3.1.2.12 - hydrolase, ec:4.1.1.31 - carboxylase, ec:4.1.2.13 - aldolase	2122	0.814664	0.4712535	0.124085347
Citrate cycle (TCA cycle)	5	map00020	ec:1.1.1.37 - dehydrogenase, ec:2.3.1.12 - acetyltransferase, ec:4.1.1.32 - carboxykinase	491	14	ec:1.3.5.1 - dehydrogenase, ec:4.1.1.32 - carboxykinase (GTP), ec:1.2.4.1 - dehydrogenase (acetyl-transferring), ec:2.3.1.12 - acetyltransferase, ec:2.3.3.8 - citrate synthase, ec:1.2.4.2 - dehydrogenase (succinyl-	2122	1.0183299	0.6597549	0.124368663

			(GTP), ec:1.1.1.41 - dehydrogenase (NAD+), ec:1.2.4.2 - dehydrogenase (succinyl-transferring)			transferring), ec:1.8.1.4 - dehydrogenase, ec:1.1.1.37 - dehydrogenase, ec:1.1.1.41 - dehydrogenase (NAD+), ec:4.2.1.3 - hydratase, ec:6.2.1.4 - ligase (GDP-forming), ec:1.1.1.42 - dehydrogenase (NADP+), ec:2.3.3.1 - (Si)-synthase, ec:6.2.1.5 - ligase (ADP-forming)				
Glutathione metabolism	4	map00480	ec:2.5.1.18 - transferase, ec:1.1.1.44 - dehydrogenase (NADP+-dependent, decarboxylating), ec:1.1.1.43 - 2-dehydrogenase, ec:3.4.19.13 - hydrolase	491	11	ec:3.4.19.13 - hydrolase, ec:1.11.1.12 - glutathione peroxidase, ec:1.1.1.44 - dehydrogenase (NADP+-dependent, decarboxylating), ec:1.11.1.9 - peroxidase, ec:6.3.2.2 - ligase, ec:4.3.2.9 - gamma-glutamyl-amino acid cyclotransferase, ec:2.5.1.18 - transferase, ec:1.11.1.15 - thioredoxin peroxidase, ec:1.1.1.42 - dehydrogenase (NADP+), ec:1.1.1.49 - dehydrogenase (NADP+), ec:1.8.1.7 - reductase	2122	0.814664	0.5183789	0.150028786
Porphyrin and chlorophyll metabolism	4	map00860	ec:1.14.13.81 - IX monomethyl ester (oxidative) cyclase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase, ec:1.3.3.5 - oxidase, ec:1.16.3.1 - ceruloplasmin	491	11	ec:2.7.7.7 - DNA polymerase, ec:2.7.1.21 - kinase, ec:2.7.1.48 - kinase, ec:3.6.1.9 - diphosphatase, ec:2.7.7.8 - nucleotidyltransferase, ec:6.3.4.2 - synthase (glutamine hydrolysing), ec:2.1.3.2 - carbamoyltransferase, ec:2.7.4.6 - kinase, ec:3.6.1.5 - ATP-diphosphatase, ec:3.1.3.5 - uridine 5'-nucleotidase, ec:2.7.7.6 - RNA polymerase	2122	0.814664	0.5183789	0.150028786
Glycerophospholipid metabolism	4	map00564	ec:3.1.1.5 - lecithinase B, ec:2.3.1.51 - O-acyltransferase, ec:2.3.1.43 - O-acyltransferase, ec:3.1.4.4 - D	491	11	ec:3.1.4.4 - D, ec:2.3.1.51 - O-acyltransferase, ec:1.1.1.94 - dehydrogenase [NAD(P)+], ec:2.3.1.43 - O-acyltransferase, ec:1.1.1.8 - dehydrogenase (NAD+), ec:2.3.1.15 - 1-O-acyltransferase, ec:2.1.1.103 - N-methyltransferase, ec:3.1.1.5 - lecithinase B, ec:3.1.4.3 - C, ec:3.1.1.4 - A2, ec:3.1.1.32 - A1	2122	0.814664	0.5183789	0.150028786
Retinol metabolism	2	map00830	ec:1.1.1.1 - dehydrogenase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase	491	4	ec:1.2.1.36 - dehydrogenase, ec:1.1.1.1 - dehydrogenase, ec:2.4.1.17 - 1-naphthol glucuronyltransferase, ec:1.14.14.1 - monooxygenase	2122	0.407332	0.1885014	0.189809315
Geraniol degradation	2	map00281	ec:1.1.1.183 - dehydrogenase (NADP+), ec:4.2.1.17 - hydratase	491	4	ec:1.1.1.183 - dehydrogenase (NADP+), ec:1.1.1.35 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:2.3.1.16 - C-acyltransferase	2122	0.407332	0.1885014	0.189809315
Tryptophan metabolism	3	map00380	ec:1.2.4.2 - dehydrogenase (succinyl-transferring), ec:4.2.1.17 - hydratase, ec:3.5.1.9 - kynurenine formamidase	491	9	ec:3.5.1.9 - kynurenine formamidase, ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.1.1.35 - dehydrogenase, ec:1.2.4.2 - dehydrogenase (succinyl-transferring), ec:2.3.1.9 - C-acetyltransferase, ec:4.2.1.17 - hydratase, ec:1.11.1.6 - equilase, ec:4.1.1.28 - decarboxylase, ec:1.14.14.1 - monooxygenase	2122	0.610998	0.4241282	0.214911752
alpha-Linolenic acid metabolism	3	map00592	ec:1.1.1.1 - dehydrogenase, ec:1.13.11.12 - 13S-lipoxygenase, ec:4.2.1.17 - hydratase	491	9	ec:4.2.1.92 - dehydratase, ec:1.3.1.42 - reductase, ec:5.3.99.6 - cyclase, ec:1.1.1.1 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:1.13.11.12 - 13S-lipoxygenase, ec:3.1.1.4 - A2, ec:2.3.1.16 - C-acyltransferase, ec:3.1.1.32 - A1	2122	0.610998	0.4241282	0.214911752
Glycine, serine and threonine metabolism	3	map00260	ec:2.1.2.1 - hydroxymethyltransferase, ec:1.1.1.1 - dehydrogenase, ec:1.1.1.81 - reductase	491	9	ec:1.8.1.4 - dehydrogenase, ec:1.1.1.1 - dehydrogenase, ec:4.2.1.20 - synthase, ec:4.3.1.19 - ammonia-lyase, ec:1.1.1.95 - dehydrogenase, ec:1.4.3.21 - oxidase, ec:3.1.3.3 - phosphatase, ec:2.1.2.1 - hydroxymethyltransferase, ec:1.1.1.81 - reductase	2122	0.610998	0.4241282	0.214911752

Cysteine and methionine metabolism	2	map00270	ec:1.1.1.37 - dehydrogenase, ec:2.1.1.14 - S-methyltransferase	491	13	ec:3.3.1.1 - S-adenosylhomocysteine synthase, ec:2.1.1.14 - S-methyltransferase, ec:2.6.1.1 - transaminase, ec:3.2.2.9 - nucleosidase, ec:6.3.2.2 - ligase, ec:1.1.1.37 - dehydrogenase, ec:1.1.1.27 - dehydrogenase, ec:2.5.1.47 - synthase, ec:2.5.1.6 - adenosyltransferase, ec:2.1.1.37 - (cytosine-5-)-methyltransferase, ec:5.3.1.23 - isomerase, ec:2.8.1.2 - sulfurtransferase, ec:2.3.1.30 - O-acetyltransferase	2122	0.407332	0.6126296	0.231203443
Neomycin, kanamycin and gentamicin biosynthesis	1	map00524	ec:2.7.1.1 - hexokinase type IV glucokinase	491	1	ec:2.7.1.1 - hexokinase type IV glucokinase	2122	0.203666	0.0471254	0.231385485
Monobactam biosynthesis	1	map00261	ec:2.7.7.4 - adenyltransferase	491	1	ec:2.7.7.4 - adenyltransferase	2122	0.203666	0.0471254	0.231385485
Steroid degradation	1	map00984	ec:1.1.1.145 - dehydrogenase	491	1	ec:1.1.1.145 - dehydrogenase	2122	0.203666	0.0471254	0.231385485
Fluorobenzoate degradation	1	map00364	ec:3.1.1.45 - maleylacetate enol-lactonase	491	1	ec:3.1.1.45 - maleylacetate enol-lactonase	2122	0.203666	0.0471254	0.231385485
Nitrotoluene degradation	1	map00633	ec:2.3.1.5 - N-acetyltransferase	491	1	ec:2.3.1.5 - N-acetyltransferase	2122	0.203666	0.0471254	0.231385485
mTOR signaling pathway	1	map04150	ec:2.7.11.24 - protein kinase	491	1	ec:2.7.11.24 - protein kinase	2122	0.203666	0.0471254	0.231385485
Novobiocin biosynthesis	1	map00401	ec:1.3.1.12 - dehydrogenase	491	1	ec:2.6.1.1 - transaminase	2122	0.203666	0.0471254	0.231385485
Taurine and hypotaurine metabolism	1	map00430	ec:1.4.1.2 - dehydrogenase	491	1	ec:1.4.1.2 - dehydrogenase	2122	0.203666	0.0471254	0.231385485
Sesquiterpenoid and triterpenoid biosynthesis	1	map00909	ec:4.2.3.23 - synthase	491	1	ec:4.2.3.23 - synthase	2122	0.203666	0.0471254	0.231385485
Naphthalene degradation	1	map00626	ec:1.1.1.1 - dehydrogenase	491	1	ec:1.1.1.1 - dehydrogenase	2122	0.203666	0.0471254	0.231385485
One carbon pool by folate	1	map00670	ec:2.1.2.1 - hydroxymethyltransferase	491	1	ec:2.1.2.1 - hydroxymethyltransferase	2122	0.203666	0.0471254	0.231385485
D-Glutamine and D-glutamate metabolism	1	map00471	ec:1.4.1.3 - dehydrogenase [NAD(P)+]	491	1	ec:1.4.1.3 - dehydrogenase [NAD(P)+]	2122	0.203666	0.0471254	0.231385485
Chlorocyclohexane and chlorobenzene degradation	1	map00361	ec:3.1.1.45 - maleylacetate enol-lactonase	491	1	ec:3.1.1.45 - maleylacetate enol-lactonase	2122	0.203666	0.0471254	0.231385485
Sphingolipid metabolism	2	map00600	ec:2.4.1.80 - glucosyltransferase, ec:3.2.1.23 - lactase (ambiguous)	491	5	ec:3.5.1.23 - acylsphingosine deacylase, ec:4.1.2.27 - aldolase, ec:3.2.1.23 - lactase (ambiguous), ec:3.2.1.22 - melibiase, ec:2.4.1.80 - glucosyltransferase	2122	0.407332	0.2356268	0.243310808

Alanine, aspartate and glutamate metabolism	3	map00250	ec:6.3.1.2 - synthetase, ec:1.4.1.3 - dehydrogenase [NAD(P)+], ec:1.4.1.2 - dehydrogenase	491	11	ec:3.5.1.111 - amidase, ec:1.4.1.14 - synthase (NADH), ec:2.6.1.1 - transaminase, ec:2.1.3.2 - carbamoyltransferase, ec:1.4.1.2 - dehydrogenase, ec:1.4.1.3 - dehydrogenase [NAD(P)+], ec:6.3.5.4 - synthase (glutamine-hydrolysing), ec:6.3.1.2 - synthetase, ec:3.5.1.3 - alpha-keto acid-omega-amidase, ec:1.2.1.16 - dehydrogenase [NAD(P)+], ec:1.4.1.4 - dehydrogenase (NADP+)	2122	0.610998	0.5183789	0.249638062
Carbon fixation pathways in prokaryotes	3	map00720	ec:1.1.1.37 - dehydrogenase, ec:4.1.1.31 - carboxylase, ec:4.2.1.17 - hydratase	491	11	ec:1.3.5.1 - dehydrogenase, ec:6.4.1.2 - carboxylase, ec:1.1.1.35 - dehydrogenase, ec:2.3.3.8 - citrate synthase, ec:2.3.1.9 - C-acetyltransferase, ec:1.1.1.37 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:4.2.1.3 - hydratase, ec:1.1.1.42 - dehydrogenase (NADP+), ec:4.1.1.31 - carboxylase, ec:6.2.1.5 - ligase (ADP-forming)	2122	0.610998	0.5183789	0.249638062
Propanoate metabolism	1	map00640	ec:4.2.1.17 - hydratase	491	9	ec:3.1.2.4 - hydrolase, ec:1.2.1.27 - dehydrogenase (CoA-acylating), ec:6.4.1.2 - carboxylase, ec:1.8.1.4 - dehydrogenase, ec:2.3.1.9 - C-acetyltransferase, ec:4.2.1.17 - hydratase, ec:1.1.1.27 - dehydrogenase, ec:6.2.1.4 - ligase (GDP-forming), ec:6.2.1.5 - ligase (ADP-forming)	2122	0.203666	0.4241282	0.253605016
Valine, leucine and isoleucine degradation	2	map00280	ec:4.2.1.18 - hydratase, ec:4.2.1.17 - hydratase	491	11	ec:1.2.1.3 - dehydrogenase (NAD+), ec:3.1.2.4 - hydrolase, ec:1.2.1.27 - dehydrogenase (CoA-acylating), ec:1.1.1.35 - dehydrogenase, ec:1.8.1.4 - dehydrogenase, ec:2.3.1.9 - C-acetyltransferase, ec:6.4.1.4 - carboxylase, ec:4.2.1.17 - hydratase, ec:1.1.1.31 - dehydrogenase, ec:4.2.1.18 - hydratase, ec:2.3.1.16 - C-acyltransferase	2122	0.407332	0.5183789	0.27618444
Lysine degradation	2	map00310	ec:1.2.4.2 - dehydrogenase (succinyl-transferring), ec:4.2.1.17 - hydratase	491	6	ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.1.1.35 - dehydrogenase, ec:1.2.4.2 - dehydrogenase (succinyl-transferring), ec:2.3.1.9 - C-acetyltransferase, ec:4.2.1.17 - hydratase, ec:2.1.1.43 - N-methyltransferase	2122	0.407332	0.2827521	0.280663671
Phenylalanine metabolism	1	map00360	ec:4.2.1.17 - hydratase	491	8	ec:2.6.1.1 - transaminase, ec:4.2.1.17 - hydratase, ec:1.4.3.21 - oxidase, ec:4.3.1.24 - ammonia-lyase, ec:6.2.1.12 - ligase, ec:2.1.1.104 - O-methyltransferase, ec:4.1.1.28 - decarboxylase, ec:4.3.1.25 - ammonia-lyase	2122	0.203666	0.3770028	0.293443352
Tyrosine metabolism	1	map00350	ec:1.1.1.1 - dehydrogenase	491	8	ec:4.1.1.68 - decarboxylase, ec:2.6.1.1 - transaminase, ec:1.1.1.1 - dehydrogenase, ec:1.10.3.1 - oxidase, ec:1.4.3.21 - oxidase, ec:4.1.1.28 - decarboxylase, ec:1.2.1.16 - dehydrogenase [NAD(P)+], ec:3.7.1.5 - hydrolase	2122	0.203666	0.3770028	0.293443352
Riboflavin metabolism	1	map00740	ec:3.1.3.2 - phosphatase	491	8	ec:2.5.1.78 - synthase, ec:1.1.1.193 - reductase, ec:3.5.4.26 - deaminase, ec:3.6.1.9 - diphosphatase, ec:2.7.7.2 - synthetase, ec:3.1.3.2 - phosphatase, ec:1.5.1.38 - reductase (NADPH), ec:1.5.1.39 - reductase [NAD(P)H]	2122	0.203666	0.3770028	0.293443352
Inositol phosphate metabolism	2	map00562	ec:5.3.1.1 - isomerase, ec:2.7.1.68 - 5-kinase	491	7	ec:3.1.4.11 - phospholipase C, ec:2.7.1.159 - 5/6-kinase, ec:2.7.1.68 - 5-kinase, ec:2.7.1.134 - 1-kinase, ec:3.1.4.3 - C, ec:5.3.1.1 - isomerase, ec:2.7.1.150 - 5-kinase	2122	0.407332	0.3298775	0.302124627
Fatty acid degradation	2	map00071	ec:1.1.1.1 - dehydrogenase, ec:4.2.1.17 - hydratase	491	9	ec:1.2.1.3 - dehydrogenase (NAD+), ec:4.2.1.74 - hydratase, ec:1.1.1.35 - dehydrogenase, ec:2.3.1.9 - C-acetyltransferase, ec:1.1.1.1 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:6.2.1.3 - ligase, ec:2.3.1.16 - C-acyltransferase, ec:1.14.14.1 - monooxygenase	2122	0.407332	0.4241282	0.306075019

Arginine and proline metabolism	2	map00330	ec:3.5.3.11 - agmatine ureohydrolase, ec:3.5.3.1 - arginine amidinase	491	9	ec:2.7.2.11 - 5-kinase, ec:1.5.3.16 - oxidase, ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.5.1.2 - reductase, ec:2.6.1.1 - transaminase, ec:1.5.3.17 - polyamine oxidase, ec:3.5.3.11 - agmatine ureohydrolase, ec:1.2.1.41 - dehydrogenase, ec:3.5.3.1 - arginine amidinase	2122	0.407332	0.4241282	0.306075019
Glycerolipid metabolism	2	map00561	ec:2.3.1.20 - O-acyltransferase, ec:2.3.1.51 - O-acyltransferase	491	9	ec:2.3.1.20 - O-acyltransferase, ec:1.2.1.3 - dehydrogenase (NAD+), ec:3.1.1.23 - lipase, ec:2.3.1.22 - O-acyltransferase, ec:2.3.1.51 - O-acyltransferase, ec:2.7.7.9 - uridylyltransferase, ec:3.1.1.3 - lipase, ec:2.3.1.15 - 1-O-acyltransferase, ec:3.2.1.22 - melibiase	2122	0.407332	0.4241282	0.306075019
Butanoate metabolism	1	map00650	ec:4.2.1.17 - hydratase	491	7	ec:1.3.5.1 - dehydrogenase, ec:1.1.1.35 - dehydrogenase, ec:2.2.1.6 - synthase, ec:2.3.1.9 - C-acetyltransferase, ec:1.1.1.30 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:1.2.1.16 - dehydrogenase [NAD(P)+]	2122	0.203666	0.3298775	0.334186832
Caffeine metabolism	1	map00232	ec:2.3.1.5 - N-acetyltransferase	491	2	ec:2.3.1.5 - N-acetyltransferase, ec:1.14.14.1 - monooxygenase	2122	0.203666	0.0942507	0.355860185
Toluene degradation	1	map00623	ec:3.1.1.45 - maleylacetate enol-lactonase	491	2	ec:1.1.1.35 - dehydrogenase, ec:3.1.1.45 - maleylacetate enol-lactonase	2122	0.203666	0.0942507	0.355860185
Aminoacyl-tRNA biosynthesis	1	map00970	ec:6.1.1.6 - ligase	491	2	ec:6.1.1.22 - ligase, ec:6.1.1.6 - ligase	2122	0.203666	0.0942507	0.355860185
Folate biosynthesis	1	map00790	ec:2.8.1.12 - synthase	491	2	ec:2.6.1.85 - synthase, ec:2.8.1.12 - synthase	2122	0.203666	0.0942507	0.355860185
T cell receptor signaling pathway	1	map04660	ec:2.7.10.2 - protein-tyrosine kinase	491	2	ec:3.1.3.16 - phosphatase, ec:2.7.10.2 - protein-tyrosine kinase	2122	0.203666	0.0942507	0.355860185
Caprolactam degradation	1	map00930	ec:4.2.1.17 - hydratase	491	2	ec:1.1.1.35 - dehydrogenase, ec:4.2.1.17 - hydratase	2122	0.203666	0.0942507	0.355860185
Limonene and pinene degradation	1	map00903	ec:4.2.1.17 - hydratase	491	2	ec:1.2.1.3 - dehydrogenase (NAD+), ec:4.2.1.17 - hydratase	2122	0.203666	0.0942507	0.355860185
beta-Alanine metabolism	1	map00410	ec:4.2.1.17 - hydratase	491	6	ec:1.5.3.16 - oxidase, ec:1.2.1.3 - dehydrogenase (NAD+), ec:3.1.2.4 - hydrolase, ec:1.5.3.17 - polyamine oxidase, ec:4.2.1.17 - hydratase, ec:1.4.3.21 - oxidase	2122	0.203666	0.2827521	0.372767178
Biosynthesis of unsaturated fatty acids	1	map01040	ec:4.2.1.17 - hydratase	491	5	ec:1.1.1.100 - reductase, ec:1.14.19.2 - 9-desaturase, ec:4.2.1.17 - hydratase, ec:2.3.1.16 - C-acyltransferase, ec:1.14.19.6 - (9+3)-desaturase	2122	0.203666	0.2356268	0.404193872
Sulfur metabolism	1	map00920	ec:2.7.7.4 - adenylyltransferase	491	5	ec:2.7.7.4 - adenylyltransferase, ec:2.5.1.47 - synthase, ec:2.8.1.1 - sulfurtransferase, ec:3.6.3.25 - ATPase, ec:2.3.1.30 - O-acetyltransferase	2122	0.203666	0.2356268	0.404193872
Pantothenate and CoA biosynthesis	1	map00770	ec:4.2.1.9 - dehydratase	491	5	ec:3.6.1.9 - diphosphatase, ec:2.2.1.6 - synthase, ec:2.7.1.33 - kinase, ec:1.1.1.86 - reductoisomerase (NADP+), ec:4.2.1.9 - dehydratase	2122	0.203666	0.2356268	0.404193872
Nicotinate and nicotinamide metabolism	1	map00760	ec:3.1.3.5 - uridine 5'-nucleotidase	491	5	ec:3.6.1.9 - diphosphatase, ec:2.7.1.23 - kinase, ec:2.1.1.7 - N-methyltransferase, ec:3.1.3.5 - uridine 5'-nucleotidase, ec:1.2.1.16 - dehydrogenase [NAD(P)+]	2122	0.203666	0.2356268	0.404193872
Benzoate degradation	1	map00362	ec:4.2.1.17 - hydratase	491	5	ec:1.1.1.35 - dehydrogenase, ec:2.3.1.9 - C-acetyltransferase, ec:4.1.3.17 - aldolase, ec:4.2.1.17 - hydratase, ec:2.3.1.16 - C-acyltransferase	2122	0.203666	0.2356268	0.404193872

Other glycan degradation	1	map00511	ec:3.2.1.23 - lactase (ambiguous)	491	5	ec:3.2.1.25 - mannanase, ec:3.2.1.51 - alpha-fucosidase, ec:3.2.1.23 - lactase (ambiguous), ec:3.2.1.52 - hexosaminidase, ec:3.2.1.24 - alpha-D-mannosidase	2122	0.203666	0.2356268	0.404193872
Fatty acid elongation	1	map00062	ec:4.2.1.17 - hydratase	491	3	ec:2.4.99.4 - alpha-2,3-sialyltransferase, ec:3.2.1.23 - lactase (ambiguous), ec:3.2.1.52 - hexosaminidase	2122	0.203666	0.1413761	0.410414223
Linoleic acid metabolism	1	map00591	ec:1.13.11.12 - 13S-lipoxygenase	491	3	ec:1.13.11.12 - 13S-lipoxygenase, ec:3.1.1.4 - A2, ec:1.14.14.1 - monooxygenase	2122	0.203666	0.1413761	0.410414223
Ether lipid metabolism	1	map00565	ec:3.1.4.4 - D	491	3	ec:3.1.4.4 - D, ec:3.1.4.3 - C, ec:3.1.1.4 - A2	2122	0.203666	0.1413761	0.410414223
Chloroalkane and chloroalkene degradation	1	map00625	ec:1.1.1.1 - dehydrogenase	491	3	ec:1.2.1.3 - dehydrogenase (NAD+), ec:1.1.1.1 - dehydrogenase, ec:3.3.2.10 - epoxide hydrolase	2122	0.203666	0.1413761	0.410414223
Primary bile acid biosynthesis	1	map00120	ec:1.3.1.3 - 5beta-reductase	491	3	ec:1.1.1.35 - dehydrogenase, ec:2.3.1.176 - C-acyltransferase, ec:1.3.1.3 - 5beta-reductase	2122	0.203666	0.1413761	0.410414223
Glycosaminoglycan degradation	1	map00531	ec:3.2.1.23 - lactase (ambiguous)	491	3	ec:3.2.1.31 - beta-glucuronide glucuronohydrolase glucuronidase, ec:3.2.1.23 - lactase (ambiguous), ec:3.2.1.52 - hexosaminidase	2122	0.203666	0.1413761	0.410414223
Valine, leucine and isoleucine biosynthesis	1	map00290	ec:4.2.1.9 - dehydratase	491	4	ec:2.2.1.6 - synthase, ec:4.3.1.19 - ammonia-lyase, ec:1.1.1.86 - reductoisomerase (NADP+), ec:4.2.1.9 - dehydratase	2122	0.203666	0.1885014	0.420679421
Glycosphingolipid biosynthesis - ganglio series	1	map00604	ec:3.2.1.23 - lactase (ambiguous)	491	4	ec:4.2.1.74 - hydratase, ec:1.1.1.35 - dehydrogenase, ec:4.2.1.17 - hydratase, ec:2.3.1.16 - C-acyltransferase	2122	0.203666	0.1885014	0.420679421
Phosphatidylinositol signaling system	1	map04070	ec:2.7.1.68 - 5-kinase	491	4	ec:3.1.4.11 - phospholipase C, ec:2.7.1.159 - 5/6-kinase, ec:2.7.1.68 - 5-kinase, ec:2.7.1.150 - 5-kinase	2122	0.203666	0.1885014	0.420679421
Isoquinoline alkaloid biosynthesis	1	map00950	ec:1.21.3.3 - oxidase	491	4	ec:2.6.1.1 - transaminase, ec:1.10.3.1 - oxidase, ec:1.4.3.21 - oxidase, ec:4.1.1.28 - decarboxylase	2122	0.203666	0.1885014	0.420679421

Tabla suplementaria S7. Clusters de las interacciones entre las proteínas sobreexpresadas de la comparación entre grupo altamente productor y bajo productor.

Cluster	Size	Density	Internal weight	External weight	<i>p</i> -value	Members
1	14	0.2825	25.71	4.095	4.86E-06	AT1G11660 FPS1 HSC70-1 P5CS2 AT5G47720 MFP2 AIM1 GAPC2 PKT4 PKT3 AT4G13010 ACAT2 AOR KAT5
2	7	0.7994	16.79	0.652	0.0007069	PHYA FUS5 COP8 COP13 FUS12 COP9 FUS6
3	8	0.3783	10.59	2.137	0.0009663	ESP3 At1g80070 MAC3A MAC3B CDC5 AT5G46840 MOS4 AT3G19810
4	5	0.6873	6.873	1.433	0.0060929	UPF3 DCP2 LBA1 DCP1 AT2G39260
5	3	0.3303	0.991	0.666	0.0404278	AT2G40770 AT1G30680 SRS2
6	2	0.976	0.976	0	0.0969654	VAM3 SGR4
7	2	0.401	0.401	0	0.0969654	AT3G13350 AT1G62660
8	2	0.396	0.396	0	0.0969654	ABCG15 ABCG40
9	2	0.363	0.363	0	0.0969654	AT5G23910 AT5G18590
10	2	0.417	0.417	0.188	0.110336	AT5G47540 PIRL4
11	4	0.3495	2.097	3.772	0.442275	ESP3 At1g80070 AT5G41140 AT3G19810
12	4	0.3175	1.905	7.018	0.557383	CSP3 emb1138 HSC70-1 P5CS2

Tabla suplementaria S8. Clusters de las interacciones entre las proteínas subexpresadas de la comparación entre grupo altamente productor y bajo productor.

Cluster	Size	Density	Internal weight	External weight	<i>p</i> -value	Members
1	33	0.2797	147.7	43.97	0.00000000	LOS1 AT2G45030 RPS11 RPS5A AT1G67430 RPS5B AT1G07210 AT3G09630 AT2G09990 AT3G13580 AT1G80750 RPL23AB AT4G36130 AT3G49910 AT4G15000 RPL3B P40 AT4G34670 AT1G70600 AT2G19750 AT1G04480 AT3G02560 EMB3010 AT5G04800 AT4G11120 BBC1 RPL5B AT5G48760 rps15 AtCg00790 AT5G59240 AT5G39850 AT5G65220
2	6	0.4544	6.816	2.707	0.00411951	AT5G51570 VHA-A3 AT3G42050 AT3G28715 VHA-A TUF
3	5	0.3777	3.777	0.261	0.00485039	NUP155 NUP1 NTF2B EMB3142 RAN3
4	4	0.4705	2.823	1.297	0.0147005	EMB1467 NDHF MATK NAD7
5	3	0.3843	1.153	0	0.0318013	AT3G53980 AGP30 AT1G17860
6	14	0.3	27.3	39.78	0.0347671	MDH GDH2 PGMP ALDH11A3 PKP-ALPHA NADP-ME4 HXK2 NADP-ME1 PPC3 CSY2 PMDH1 TPI GLN1-1 IDH-V
7	3	0.377	1.131	0.165	0.0382613	LD AT2G47820 FLK
8	3	0.575	1.725	0.56	0.0382613	AT5G46940 UGD3 RGP5
9	3	0.304	0.912	0.331	0.0404278	ABCF4 MSR2 AT2G46290
10	14	0.2732	24.86	35.93	0.0805466	AT1G64190 MDH GDH2 PGMP PKP-ALPHA NADP-ME4 HXK2 NADP-ME1 FBA4 CA2 PPC3 CSY2 PMDH1 TPI
11	3	0.352	1.056	0.543	0.0920193	AT5G20870 PME3 AT1G72160
12	7	0.312	6.552	14.24	0.124816	MOM NRPD1B AT4G40030 AT1G09200 SUVH5 JMJ18 AT2G28720
13	4	0.3475	2.085	2.076	0.156161	AT5G44400 AT1G03220 AT5G48540 scpl50
14	4	0.3475	2.085	2.076	0.156161	AT5G44400 AT1G03220 AT5G48540 scpl17
15	4	0.3475	2.085	2.076	0.156161	AT5G44400 AT1G03220 AT5G48540 SCPL27
16	13	0.3059	23.86	39.6	0.164939	MDH GDH2 PGMP PKP-ALPHA NADP-ME4 NADP-ME1 PPC3 CSY2 PMDH1 TPI GLN1-1 AT3G55410 IDH-V
17	8	0.3016	8.444	47.32	0.213567	AT3G47570 RPK2 CYP71B34 MPL12.8 CYP87A2 BAS1 CYP81D8 UXS3
18	4	0.3555	2.133	2.275	0.235243	PME3 AT5G44400 AT1G03220 AT5G48540

19	12	0.3108	20.51	37.91	0.25336	MDH GDH2 PGMP PKP-ALPHA NADP-ME4 PHO1 NADP-ME1 CA2 PPC3 CSY2 PMDH1 TPI
20	6	0.3077	4.616	13.04	0.287587	MOM NRPD1B AT4G40030 SUVH5 DRB2 JMJ18
21	4	0.305	1.83	3.41	0.332503	MAPKKK21 AT5G01850 PP2AB2 B'GAMMA
22	7	0.311	6.532	46.95	0.350268	AT3G47570 RPK2 AT3G20150 AT5G59100 MPL12.8 CYP87A2 BAS1
23	4	0.3143	1.886	7.328	0.557383	PGMP PHO1 ADG1 SGR2
24	11	0.3205	17.63	72.67	0.628702	FAB1D AT1G12310 MPK16 CPK1 AT3G47570 RPK2 MPL12.8 AT2G21130 TUB3 Hsp81.4 AT4G05230
25	5	0.359	3.59	13.76	0.661948	MOM NRPD1B AT4G40030 SUVH5 XPB1
26	5	0.3616	3.616	14.08	0.661948	MOM NRPD1B AT4G40030 SUVH5 NRPB8A
27	3	0.3977	1.193	3.786	0.66874	SDR1 AT2G43420 BAS
28	6	0.3153	4.729	9.609	0.712748	AT4G05230 AT4G40030 AT1G09200 SUVH5 AT2G28720 HTA7
29	4	0.3537	2.122	10.72	0.764757	CLPP5 AT3G01920 CLP2 AT4G11120
30	4	0.3147	1.888	10.1	0.764757	AT2G35490 CLPP5 CLP2 AT4G11120
31	4	0.3058	1.835	15.72	0.764757	SHM6 AT2G09990 AT5G06580 HPR3
32	3	0.4597	1.379	4.027	0.808633	AT4G19880 MDAR6 AT1G54220
33	7	0.3276	6.879	18.54	0.846657	AT1G65280 Hsp81.4 HSP21 AT1G53540 AR192 GFA2 AT3G57340
34	5	0.3157	3.157	13.91	0.851865	PGMP SUS3 SUS4 AT4G33070 ADG1
35	5	0.3127	3.127	29.27	0.851865	UKL4 NAGS1 SHM6 OVA5 ATMS1
36	9	0.3027	10.9	38.38	0.855343	MDH GDH2 PKP-ALPHA HXK2 FBA4 CSY2 AT5G22620 PMDH1 TPI
37	4	0.3488	2.093	16.3	0.903035	SHM6 AT2G09990 OVA5 HPR3
38	4	0.3488	2.093	17.16	0.943824	SHM6 ATKRS-1 AT2G09990 HPR3
39	4	0.3063	1.838	27.49	0.943824	AT2G34480 AT1G29320 AT3G09630 RPL23AB
40	6	0.3094	4.641	50.32	0.977336	FAB1D UKL4 At1g49670 PDIL1-1 RD21A PMDH1



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
DOCTORADO EN CIENCIAS NATURALES

Cuernavaca, Morelos a 13-11-19

COMITÉ DE REVISION DE TESIS

Dr. José de Jesús Arellano García (Tutor)
Dr. Sergio Manuel Encarnación Guevara (Co Tutor)
Dra. Laura Patricia Álvarez Berber
Dra. Rosa Victoria Pando Robles
Dr. Alexandre Toshirrico Cardoso Taketa
Dra. María del Carmen Gutiérrez Villafuerte
Dr. Efraín Tovar Sánchez

Tesis: "Estudio del proteoma de las raíces de *Acourtia cordata* y la influencia de factores edáficos en la producción de perezona"

Alumno que la presenta a revisión: **MA. DEL CARMEN GARCIA MENDEZ**

Programa: DOCTORADO EN CIENCIAS NATURALES

VOTO

El documento ha sido revisado y reúne los requisitos para editarse como TESIS por lo que es **APROBADO**

ATENTAMENTE

DRA. ROSA VICTORIA PANDO ROBLES



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
DOCTORADO EN CIENCIAS NATURALES

Cuernavaca, Morelos a 13-11-19

COMITÉ DE REVISION DE TESIS

Dr. José de Jesús Arellano García (Tutor)
Dr. Sergio Manuel Encarnación Guevara (Co Tutor)
Dra. Laura Patricia Álvarez Berber
Dra. Rosa Victoria Pando Robles
Dr. Alexandre Toshirrico Cardoso Taketa
Dra. María del Carmen Gutiérrez Villafuerte
Dr. Efraín Tovar Sánchez

Tesis: "Estudio del proteoma de las raíces de *Acourtia cordata* y la influencia de factores edáficos en la producción de perezona"

Alumno que la presenta a revisión: **MA. DEL CARMEN GARCIA MENDEZ**

Programa: DOCTORADO EN CIENCIAS NATURALES

VOTO

El documento ha sido revisado y reúne los requisitos para editarse como TESIS por lo que es **APROBADO**

ATENTAMENTE

DR. JOSE DE JESUS ARELLANO GARCIA



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
DOCTORADO EN CIENCIAS NATURALES

Cuernavaca, Morelos a 13-11-2019

COMITÉ DE REVISION DE TESIS

Dr. José de Jesús Arellano García (Tutor)
Dr. Sergio Manuel Encarnación Guevara (Co Tutor)
Dra. Laura Patricia Álvarez Berber
Dra. Rosa Victoria Pando Robles
Dr. Alexandre Toshirrico Cardoso Taketa
Dra. María del Carmen Gutiérrez Villafuerte
Dr. Efraín Tovar Sánchez

Tesis: "Estudio del proteoma de las raíces de *Acourtia cordata* y la influencia de factores edáficos en la producción de perezona"

Alumno que la presenta a revisión: **MA. DEL CARMEN GARCIA MENDEZ**

Programa: DOCTORADO EN CIENCIAS NATURALES

VOTO

El documento ha sido revisado y reúne los requisitos para editarse como TESIS por lo que es **APROBADO**

ATENTAMENTE

DRA. LAURA PATRICIA ALVAREZ BERBER



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
DOCTORADO EN CIENCIAS NATURALES

Cuernavaca, Morelos a 13 NOV 2019

COMITÉ DE REVISION DE TESIS

Dr. José de Jesús Arellano García (Tutor)
Dr. Sergio Manuel Encarnación Guevara (Co Tutor)
Dra. Laura Patricia Álvarez Berber
Dra. Rosa Victoria Pando Robles
Dr. Alexandre Toshirrico Cardoso Taketa
Dra. María del Carmen Gutiérrez Villafuerte
Dr. Efraín Tovar Sánchez

Tesis: "Estudio del proteoma de las raíces de *Acourtia cordata* y la influencia de factores edáficos en la producción de perezona"

Alumno que la presenta a revisión: **MA. DEL CARMEN GARCIA MENDEZ**

Programa: DOCTORADO EN CIENCIAS NATURALES

VOTO

El documento ha sido revisado y reúne los requisitos para editarse como TESIS por lo que es **APROBADO**

ATENTAMENTE

DR. ALEXANDRE TOSHIRRICO CARDOSO TAKETA



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
DOCTORADO EN CIENCIAS NATURALES

Cuernavaca, Morelos a 13-11-19

COMITÉ DE REVISION DE TESIS

Dr. José de Jesús Arellano García (Tutor)
Dr. Sergio Manuel Encarnación Guevara (Co Tutor)
Dra. Laura Patricia Álvarez Berber
Dra. Rosa Victoria Pando Robles
Dr. Alexandre Toshirrico Cardoso Taketa
Dra. María del Carmen Gutiérrez Villafuerte
Dr. Efraín Tovar Sánchez

Tesis: "Estudio del proteoma de las raíces de *Acourtia cordata* y la influencia de factores edáficos en la producción de perezona"

Alumno que la presenta a revisión: **MA. DEL CARMEN GARCIA MENDEZ**

Programa: DOCTORADO EN CIENCIAS NATURALES

VOTO

El documento ha sido revisado y reúne los requisitos para editarse como TESIS por lo que es **APROBADO**

ATENTAMENTE

DRA. MARIA DEL CARMEN GUTIERREZ VILLAFUERTE



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
DOCTORADO EN CIENCIAS NATURALES

Cuernavaca, Morelos a 13-11-19

COMITÉ DE REVISION DE TESIS

Dr. José de Jesús Arellano García (Tutor)
Dr. Sergio Manuel Encarnación Guevara (Co Tutor)
Dra. Laura Patricia Álvarez Berber
Dra. Rosa Victoria Pando Robles
Dr. Alexandre Toshirrico Cardoso Taketa
Dra. María del Carmen Gutiérrez Villafuerte
Dr. Efraín Tovar Sánchez

Tesis: "Estudio del proteoma de las raíces de *Acourtia cordata* y la influencia de factores edáficos en la producción de perezona"

Alumno que la presenta a revisión: **MA. DEL CARMEN GARCIA MENDEZ**

Programa: DOCTORADO EN CIENCIAS NATURALES

VOTO

El documento ha sido revisado y reúne los requisitos para editarse como TESIS por lo que es **APROBADO**

ATENTAMENTE

DR. SERGIO MANUEL ENCARNACION GUEVARA



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
DOCTORADO EN CIENCIAS NATURALES

Cuernavaca, Morelos a 13-11-14

COMITÉ DE REVISION DE TESIS

Dr. José de Jesús Arellano García (Tutor)
Dr. Sergio Manuel Encarnación Guevara (Co Tutor)
Dra. Laura Patricia Álvarez Berber
Dra. Rosa Victoria Pando Robles
Dr. Alexandre Toshirrico Cardoso Taketa
Dra. María del Carmen Gutiérrez Villafuerte
Dr. Efraín Tovar Sánchez

Tesis: "Estudio del proteoma de las raíces de *Acourtia cordata* y la influencia de factores edáficos en la producción de perezona"

Alumno que la presenta a revisión: **MA. DEL CARMEN GARCIA MENDEZ**

Programa: DOCTORADO EN CIENCIAS NATURALES

VOTO

El documento ha sido revisado y reúne los requisitos para editarse como TESIS por lo que es **APROBADO**

ATENTAMENTE

DR. EFRAIN TOVAR SANCHEZ