UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS BIOLÓGICAS Y DE LA SALUD
DEPARTAMENTO DE INVESTIGACIONES CIENTÍFICAS Y TECNOLÓGICAS

POSGRADO EN BIOCIENCIAS

ANÁLISIS DEL CAMBIO EN EXPRESIÓN GÉNICA DURANTE EL PROCESO DE MADURACIÓN DE FRUTO ENTRE UNA VARIEDAD CULTIVADA Y UNA SILVESTRE DE CHILE (Capsicum annuum L.)

TESIS

que para obtener el grado de:

DOCTOR EN BIOCIENCIAS

presenta:

FERNANDO GUADALUPE RAZO MENDÍVIL

Hermosillo, Sonora, México **27 de mayo de 2021**

Hermosillo, Sonora a 21 de junio de 2021.

Asunto: Cesión de derechos

UNIVERSIDAD DE SONORA
P R E S E N T E.

Por este conducto hago constar que soy autor y titular de la obra denominada "Análisis del cambio en expresión génica durante el proceso de maduración de fruto entre una variedad cultivada y una silvestre de chile (*Capsicum annuum* L.)", en los sucesivo LA OBRA, realizada como trabajo terminal con el propósito de obtener el Grado de **Doctor en Biociencias**, en virtud de lo cual autorizo a la Universidad de Sonora (UNISON) para que efectúe la divulgación, publicación, comunicación pública, distribución, distribución pública, distribución electrónica y reproducción, así como la digitalización de la misma, con fines académicos o propios de la institución y se integren a los repositorios de la universidad, estatales, regionales, nacionales e internacionales.

La UNISON se compromete a respetar en todo momento mi autoría y a otorgarme el crédito correspondiente en todas las actividades mencionadas anteriormente.

De la misma manera, manifiesto que el contenido académico, literario, la edición y en general cualquier parte de LA OBRA son de mi entera responsabilidad, por lo que deslindo a la UNISON por cualquier violación a los derechos de autor y/o propiedad intelectual y/o cualquier responsabilidad relacionada con la OBRA que cometa el suscrito frente a terceros.

ATENTAMENTE

Dr. Fernando Guadalupe Razo Mendivil

LIC. GILBERTO LEÓN LEÓN
Abogado General
UNIVERSIDAD DE SONORA

Hermosillo, Sonora, México

Junio, 2021.

ANÁLISIS DEL CAMBIO EN EXPRESIÓN GÉNICA DURANTE EL PROCESO DE MADURACIÓN DE FRUTO ENTRE UNA VARIEDAD CULTIVADA Y UNA SILVESTRE DE CHILE (*Capsicum annuum* L.)

T E S I S

que para obtener el grado de:
DOCTOR EN BIOCIENCIAS

Presenta:
FERNANDO GUADALUPE RAZO MENDIVIL

Hermosillo, Sonora, México. 27 de Mayo de 2021

APROBACIÓN

Los miembros del Comité designado para revisar la tesis intitulada Análisis del cambio en expresión génica durante el proceso de maduración de fruto entre una variedad cultivada y una silvestre de chile (*Capsicum annuum* L.) presentada por Fernando Guadalupe Razo Mendivil, la han encontrado satisfactoria y recomiendan que sea aceptada como requisito parcial para obtener el grado de Doctor en Biociencias.

Angela Corina Hayano Kanashiro
Codirector del Comité

Octavio Martinez de la Vega
Codirector del Comité

Luis Ángel Medina Juárez
Secretario

Nohemí Gamez Meza
Sinodal interno

Miguel Angel Hernández Oñate
Sinodal externo

DEDICATORIA

A Mario. A mis padres Eva y Elpidio. A mis hermanos, Julissa, Erika y José.

AGRADECIMIENTOS

A mis padres, mi orgullo y ejemplo a seguir, gracias por su apoyo y cariño.

A mis directores de tesis, la Dra. Corina Hayano Kanashiro y el Dr. Octavio Martínez de la Vega, por ser mi paciente guía durante el desarrollo de este trabajo.

A mi comité de tesis, por sus observaciones y comentarios para mejorar este trabajo de investigación

A mis amigos, por su apoyo incondicional.

A los profesores y personal del Departamento de Investigaciones Científicas y Tecnológicas de la Universidad de Sonora por su enseñanza y hacer amena mi estancia en la institución.

A la Universidad de Sonora, por proporcionarme las facilidades para realizar este trabajo.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme el apoyo económico para realizar este doctorado.

RESUMEN

El chile (*Capsicum annuum* L.) es uno de los cultivos más importantes a nivel mundial. En Sonora, el chiltepín (*Capsicum annuum* L. *var glabriusculum*) posee un papel sociocultural y económico importante, por lo que, ha llegado a convertirse en un fruto icónico de la región. Biológicamente, el chiltepín es de peculiar interés ya que es una variedad silvestre de *Capsicum annuum* que es la especie de chile con mayor distribución y producción mundial. El objetivo de este trabajo consistió en determinar los cambios cualitativos y cuantitativos en la expresión génica durante el proceso de maduración de fruto en dos variedades de chile: una domesticada (*Capsicum annuum* L; 'Serrano Tampiqueño 74'), y su contraparte silvestre (*Capsicum annuum* L. var. *glabriusculum*; 'Chiltepín'). Para ello se utilizaron herramientas bioinformáticas como Blast2GO, Hisat2, HTSeq, BLAST, las bases de datos KEGG y GO y el paquete estadístico R project. Se generaron estudios comparativos entre los transcriptomas de chiltepín a 20 y 68 días después de antesis (DDA) y de Serrano a 20 y 60 DDA, los cuales permitieron identificar los genes con diferencias de expresión significativas, se categorizaron funcionalmente utilizando notación GO, KEGG y MapMan, y se identificaron genes relacionados con los procesos de maduración de fruto en tomate.

ABSTRACT

The chili (*Capsicum annuum* L.) is one of the most important crops in the world, especially for its culinary use. In Sonora State, the chiltepin (*Capsicum annuum* L. var. *glabriusculum*) plays an important socio-cultural and economic role and it has become the most iconic fruit from the region. Biologically, the chiltepin is a very interesting organism because it is known as a wild variety of the *Capsicum annuum* L. species, the chili species most distributed and cultivated around the world. The aim of this study is to analyze qualitative and quantitative differences on gene expression of two varieties of chili: a domesticated one (*Capsicum annuum* L; 'Serrano Tampiqueño 74') and a wild one (*Capsicum annuum* L. var. *glabriusculum*; 'Chiltepin'), during fruit development using bioinformatic tools such as sequence alignment, Blast2GO, Hisast2, HTSeq, the KEGG and GO databases and packages of the statistical framework R. An analysis of the differential gene expression of Chiltepin fruits at 20 and 68 days post anthesis DPA and Serrano Tampiqueño 74 at 20 and 60 DPA was performed and the most differentially expressed genes were identified, subsequently categorized using GO, KEGG and Mapman functional annotation, and orthologs with tomato genes related with different aspects of fruit development and ripening were identified.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

Introducción

La domesticación es un proceso co-evolutivo producido por la relación entre un domesticador y un domesticado, usualmente, la domesticación está definida desde la perspectiva del domesticador y su rol en llevar las plantas silvestres hacia su cultivo, asumiendo el control de todos los aspectos de su ciclo vital, surgiendo así especies nuevas o poblaciones diferenciadas que se han convertido en críticas para la supervivencia humana (Pickersgill, 2007; Zeder, 2015). El proceso de domesticación comienza con el consumo de plantas silvestres que eventualmente llevará su cultivo y finalizará con la fijación de características deseables seleccionadas por humanos. Las plantas domesticadas muestran lo que es conocido como síndrome de domesticación, dependen totalmente de los humanos para su desarrollo y reproducción, haciéndolas menos capaces de sobrevivir en el medio silvestre, aumentando el tamaño de las partes recolectables de las plantas, perder la capacidad de dispersión de semillas, la dormancia de semillas y la protección mecánica/química contra ataques herbívoros (Pickersgill, 2007).

El chile (*Capsicum sp.* L.) es uno de los cultivos más importantes en el mundo (De Pace *et al.*, 2011), posee importancia comercial, agricultural, económica y cultural en países de hispano-América, Asia y África (López-Aguilar *et al.*, 2012a; Perry y Flannery, 2007a), donde sus frutos son utilizados para el consumo humano, como colorantes Chen y Zanmin (2009), como conservadores Srinivasan (2005a), y como ingredientes en compuestos repelentes Perry (2012a).

De las cinco especies de chile domesticadas, la más importante es *Capsicum annuum* L. debido a su amplio cultivo a nivel mundial De Pace *et al.* (2011); Hernández-Verdugo *et al.*

(2012a), existen una cantidad considerable de cultivares en esta especie, y sus frutos presentan diferencias morfológicas muy grandes, en tamaño, forma y niveles de pungencia que van desde frutos dulces hasta muy picantes Aguilar-Meléndez *et al.* (2009a); Zhang *et al.* (2002a).

El chiltepín (*C. annuum* L. var. *glabriusculum* (Dunal) Heiser & Pickersgill) es una variedad silvestre de *Capsicum annuum* que se encuentra distribuido desde el sur de Norteamérica hasta el norte de Sudamérica y en México se encuentra distribuida de Sonora a Chiapas por la costa del Pacífico Nabhan *et al.* (1990a); González-Jara *et al.* (2011a); Miranda-Zarazúa *et al.* (2010a); Pacheco-Olivera *et al.* (2012a), y se considera que podría ser el progenitor de los chiles domesticados de la especie *annuum* Singh (2006a), por lo cual es considerado una fuente de recursos genéticos importantes que pueden ser utilizados para el mejoramiento de los cultivares actuales Hernández-Verdugo *et al.* (1998a).

El chile posee una importancia económica en México dado que es el segundo productor a nivel mundial, a su vez el proceso de maduración de fruto de chile es de interés nutricional y biológico debido a que en este se llevan a cabo la producción de los metabolitos secundarios que brindarán las propiedades organolépticas y nutricionales del fruto. En investigaciones previas se ha mostrado la dinámica de expresión de genes durante la producción de frutos de *Capsicum*; sin embargo, la información actual es limitada a la variedad de chile serrano 'tampiqueño 74' por lo que es importante comparar el transcriptoma de otras variedades de *Capsicum*; en particular, los datos de chiltepín (*Capsicum annuum* var. *glabriusculum*), que es una accesión silvestre de chile.

I. Antecedentes

I.1. Generalidades del chile (*Capsicum annuum* L.)

El género *Capsicum* pertenece a la familia de las *Solánaceas*, nativa del norte de Sudamérica y el sur de Norteamérica (Greenlaf, 1986; Grubben y El Tahir, 2004). Esta familia incluye otros cultivos importantes como la papa (*Solanum tuberosum*), tomate (*Solanum lycopersicum*), tabaco (*Nicotiana tabacum*) , berenjena (*Solanum melongena*) y plantas ornamentales como la petunia (*Petunia sp.*) (Qin *et al.*, 2014; Martínez-López *et al.*, 2014).

El número de especies reportadas para el género *Capsicum* varía dependiendo del autor de 25 a 36 especies (Hayano-Kanashiro *et al.*, 2016); sin embargo, solo cinco especies han sido domesticadas: *C. annuum* L. (México y la parte norte de centroamérica), *C. baccatum* L. (Perú y Bolivia), *C. chinense* Jacq. (islas del Caribe), *C. frutescens* L. (principalmente el Caribe y sudamérica), y *C. pubescens* Ruiz & Pavon (encontrada en las grandes alturas de los Andes) (Wahyuni *et al.*, 2013; Russo, 2012; Zhigila *et al.*, 2014).

De las cinco especies de chile domesticadas, *Capsicum annuum* es la más importante por su distribución a nivel mundial (Alonso *et al.*, 2008; De Pace *et al.*, 2011; Hernández-Verdugo *et al.*, 2012b). Los frutos de esta especie poseen una amplia variabilidad de forma, tamaños y grados de pungencia que van desde dulce hasta muy picante (Aguilar-Meléndez *et al.*, 2009b; Zhang *et al.*, 2002b). México es considerado el centro de domesticación de la especie *C. annuum* (Pickersgill, 1984), donde se han presentado diferentes eventos de domesticación hace 6,000 a 7,000 años (Perry y Flannery, 2007b; Smith, 1967; Perry *et al.*, 2007).

Las plantas de chile han generado mecanismos de defensa específicos, como la capsaicina que le da la pungencia a sus frutos, al ser un irritante natural evita que los mamíferos los consuman, mientras que las aves pueden comerlos sin problema; las semillas pasan por sus sistemas digestivos y posteriormente son depositadas en la tierra. Además de la capsaicina, las plantas pueden sintetizar otros metabolitos secundarios que juegan un rol importante en su desarrollo, en respuesta a la falta de nutrientes, de exposición al sol, o a ambientes adversos (Sepúlveda-Jiménez *et al.*, 2003).

I.2. Importancia del cultivo de chile

El chile (*Capsicum annuum* L.) es considerado uno de los cultivos más importantes en el mundo (De Pace *et al.*, 2011), posee importancia económica y comercial para países iberoamericanos, asiáticos y africanos (López-Aguilar *et al.*, 2012b; Perry y Flannery, 2007b) donde sus frutos son utilizados para consumo humano frescos como vegetales (Janick y Paull, 2008; Forero *et al.*, 2009) o secos como especia (DeWitt y Bosland, 2009), como colorante natural, como conservador (Srinivasan, 2005b), como ingrediente en productos farmacéuticos y como extracto pungente es el principal ingrediente en los atomizadores de defensa personal (Perry, 2012b). De todas las especias, el chile es el segundo en volumen y valor comercial con una producción total en 2018 de 3.7 millones de hectáreas y 40.9 millones de toneladas de fruto recolectado (FAOSTAT, 2020).

México posee el segundo lugar en producción mundial de chile con 3,766,757 millones de toneladas de fruto recolectado según la FAO en 2018 (FAOSTAT, 2020), varios estudios han estimado que en México el chile es el segundo vegetal de mayor consumo después del tomate, con un consumo aproximado de 7 a 9 kg por persona al año (Alvarez-Parrilla *et al.*, 2011).

I.3. Metabolitos secundarios

Los metabolitos secundarios no son considerados indispensables para la sobrevivencia, pero se ha observado que están involucrados en funciones ecológicas (González-Jara *et al.*, 2011b), y son responsables de gran parte de las características organolépticas del fruto y aportan un elevado valor nutricional. Las plantas de chile han sido valoradas a nivel mundial como una fuente importante de compuestos activos beneficiosos para la salud humana, sus frutos contienen una cantidad considerable de vitaminas, carotenoides, compuestos fenólicos y capsaicinoides, compuestos que solo se encuentran en los frutos de chile y son responsables de su sabor pungente característico (Castro-Concha *et al.*, 2012; Silva *et al.*, 2012).

I.3.1. Capsaicinoides

Los capsaicinoides es un grupo de compuestos oleorresinas, productos de la condensación de vanilinaminas y ácidos grasos, exclusivamente sintetizados y acumulados en pequeñas vesículas en la placenta de los frutos de chile (*Capsicum sp*) (Kehie *et al.*, 2015; Cuevas-Glory *et al.*, 2015), existen más de 20 capsaicinoides análogos que difieren solamente en el tamaño del ácido graso acilado a la vanilinamina, de éstos, la capsaicina y la dihidrocapsaicina son los dos capsaicinoides más abundantes y representan el 90 % del total encontrado en frutos de chile (Luo *et al.*, 2011; Nuñez-Palenius y Ochoa-Alejo, 2005). Los capsaicinoides poseen propiedades benéficas para la salud humana: son estimulantes del sistema cardiovascular (Govindarajan y Sathyanarayana, 1991), antiinflamatorios (Anogianaki *et al.*, 2007), antioxidantes (Si *et al.*, 2012), analgésicos y ayudan a prevenir la obesidad (Luo *et al.*, 2011). Estudios clínicos han mostrado que pueden ser utilizados como un tratamiento para varios tipos de artritis y neuralgia (L Deal *et al.*, 1991), se ha mostrado útil en el tratamiendo de diversos síndromes digestivos y urinarios (Chancellor y De Groat, 1999; Cruz, 2004), y en el tratamiento de cáncer por sus propiedades químico-preventivas de la tumorgénesis (Su Han *et al.*, 2001; Srinivasan, 2005b; Surh,

2002). Además, son conocidos por su actividad antimicrobiana y por su uso como especia para inhibir el crecimiento de bacterias que producen la descomposición de los alimentos (Billing y W. Sherman, 1998).

I.3.2. Carotenoides

Los carotenoides son compuestos encontrados en la naturaleza, sintetizados por todos los organismos fotosintéticos como plantas y algas y en microorganismos no fotosintéticos como hongos y algunas bacterias, representan un vasto grupo de pigmentos lipofílicos con colores que van del amarillo y naranja al rojo intenso y son responsables de la coloración de flores y frutos (Arimboor *et al.*, 2015; Gómez-García y Ochoa-Alejo, 2013; Doka *et al.*, 2013). En las plantas los carotenoides se encuentran en los plástidos, en los cloroplastos de tejidos fotosintéticos y en los cromoplastos de flores y frutos. En las hojas se encuentran como compuestos libres, mientras que en otros tejidos se encuentran de forma esterificada (Arimboor *et al.*, 2015). La principal función de estos compuestos consiste en proteger a la célula y sus organelos del daño oxidativo causado por la interacción con formas energéticamente excitadas de oxígeno molecular y atrapando radicales libres peroxilo (responsables de la oxidación de lípidos), también participan en la fotosíntesis (en el proceso de absorción de energía como fotoreceptores del aparato fotosintético), en la protección contra el daño producido por exposición a luz y como precursores del ácido abscísico (fitohormona importante para el desarrollo y respuesta de la planta ante estrés). Ecológicamente, los carotenoides con sus atractivos colores, son responsables de atraer a los polinizadores y a los agentes encargados de la dispersión de semillas (DellaPenna y Pogson, 2006; Gómez-García y Ochoa-Alejo, 2013). Cada uno de los atractivos colores observados en chiles es dado por la concentración y composición de sus carotenoides, la clorofila es responsable del color verde en los frutos, mientras que las antocianinas son pigmentos púrpuras/violetas, los colores amarillo y naranja son producidos por α-caroteno, β-caroteno, zeaxantina, luteína y

β-criptoxantina, los colores rojos son producidos por carotenoides con formas oxigenadas como la capsantina y capsorubina, siendo el primero el más común en los frutos de *Capsicum* rojos, llegando a representar hasta un 50 % del total de carotenoides (Gómez-García y Ochoa-Alejo, 2013; Guil-Guerrero y Rebolloso-Fuentes, 2009).

I.3.3. Capsinoides

Los capsinoides son compuestos estructuralmente similares a los capsaicinoides, poseen beneficios similares con un sabor menos pungente, se les atribuye la capacidad de aumentar la secreción de catecolaminas, aumentar la producción de energía, suprimir la acumulación de grasa corporal y evitar la autooxidación (Singh *et al.*, 2009).

I.3.4. Compuestos fenólicos

Los compuestos fenólicos o polifenoles son metabolitos secundarios que se sintetizan en las plantas a partir de aminoácidos aromáticos como un mecanismo de defensa; adicionalmente, contribuyen al color, sabor y astringencia de la planta y son potentes antioxidantes que pueden reducir la probabilidad de sufrir una enfermedad cardiovascular (Drago-Serrano *et al.*, 2006). Los compuestos fenólicos pueden ser producidos en las rutas metabólicas de shikimato/fenilpropanoides que produce derivados de los fenilpropanoides, la ruta metabólica acetato/malonato que produce cadenas elongadas de fenilpropanoides incluyendo el enorme grupo de los flavonoides y algunas quinonas y, en la ruta metabólica acetato/mevalonato donde se producen los terpenoides por reacciones de deshidrogenación. Las plantas pueden sintetizar miles de compuestos fenólicos diferentes, se estima que aproximadamente un 2 % del carbono fotosintetizado por las plantas es convertido en flavonoides o compuestos similares (Lattanzio, 2013; Bhattacharya *et al.*, 2010).

I.3.5. Tocoferoles

Los frutos de *Capsicum* también son ricos en tocoferoles (vitamina E), especialmente en forma deshidratada. La combinación de tocoferoles y tocotrienoles es conocida por su capacidad de inhibir la oxidación de lípidos, tanto en alimentos como en sistemas biológicos (Hayano-Kanashiro *et al.*, 2016). Durante la maduración de fruto de chile se lleva a cabo la transformación de cloroplastos en cromoplastos, la descomposición de la clorofila que se lleva a cabo en este proceso puede incitar la formación de tocoferoles, como un método para proteger los lípidos del estrés oxidativo (Chrost *et al.*, 1999). Los cromoplastos son conocidos por ser la etapa final en la diferenciación de plástidos, su aparición está asociada con diferentes cambios incluyendo el aumento de carotenoides,capsaicinodies, la pérdida de galactolípidos y un aumento en el contenido de fosfolípidos y, de forma paralela, un aumento en el contenido de alfa-tocoferoles (Koch *et al.*, 2002).

I.4. Maduración de fruto

El desarrollo del fruto de chile no ha sido completamente dilucidado; sin embargo, el tomate, un pariente cercano del chile, ha sido utilizado como modelo para el estudio de desarrollo y maduración de frutos carnosos (Dong *et al.*, 2014). El fruto de tomate se desarrolla a partir del ovario en cuatro etapas (Figura 1): (1) fertilización de la flor, (2) proliferación de las células del pericarpio, (3) crecimiento del fruto debido a la expansión celular y endoreduplicación que finaliza con un fruto verde maduro (estado VM) con el tamaño final, y la (4) maduración que finaliza con un fruto rojo maduro (estado RM) (Tanksley, 2004).

Durante el proceso de maduración, los frutos de chile sufren una transformación de color, aroma, composición nutricional, sabor y textura (Bouvier *et al.*, 1998a; Martí *et al.*, 2009). Estos cambios son el resultado de un proceso dinámico que involucra cambios moleculares y

bioquímicos complejos (Carrari y Fernie, 2006) guiados por regulación génica en respuesta a las perturbaciones del medio ambiente (Osorio *et al.*, 2012). El proceso de maduración de fruto ha evolucionado para producir frutos atractivos para los organismos dispersores de semillas, al activar rutas metabólicas que sintetizan pigmentos, azúcares, compuestos volátiles asociados al aroma y promoviendo el ablandamiento y degradación del tejido para facilitar la liberación de las semillas (Aivalakis y Katinakis, 2008; Matas *et al.*, 2011).

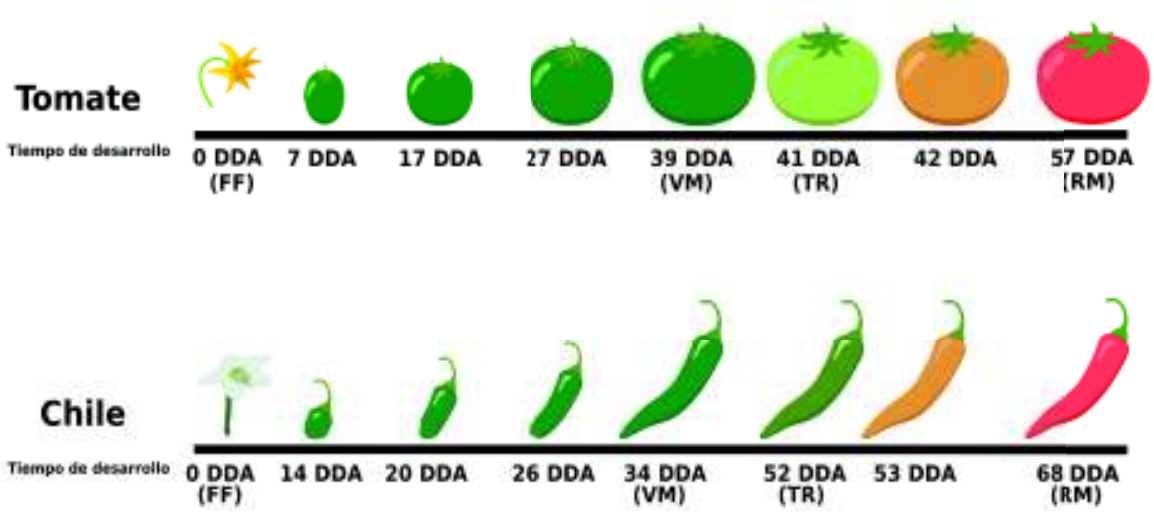


Figura 1. Etapas en el proceso de desarrollo y maduración de fruto de tomate (climatérico) y chile (no climatérico), adaptado de Klie *et al.* (2014).

I.5. Maduración de frutos climatéricos y no climatéricos

Generalmente los frutos pueden ser divididos en dos categorías dependiendo de su comportamiento al madurar y la presencia o ausencia de picos en la biosíntesis de etileno, así, en los frutos climatéricos como el tomate (*Solanum lycopersicum*), plátano (*Musa spp.*), durazno (*Prunus persica*) y la manzana (*Malus domestica*), el etileno es esencial para la maduración del fruto, si se bloquea su síntesis o sus receptores, se evita que el fruto madure (Giovannoni, 2001). Los

frutos no climatéricos como las uvas (*Vitis vinifera*), las fresas (*Fragaria ananassa*), o los chiles (*Capsicum spp.*) producen cantidades muy pequeñas de etileno y la aplicación externa de éste no produce una aceleración de la maduración de sus frutos (Perkins-Veazie, 1995). El etileno o eteno es una fitohormona gaseosa en condiciones ambientales normales, regula una amplia cantidad de aspectos del ciclo vital de las plantas, incluyendo la germinación de semillas, iniciación y producción de vellosidades en raíces, desarrollo de flores, determinación sexual, maduración de fruto, senescencia, respuesta a factores bióticos y factores abióticos (Abeles *et al.*, 1992).

El rol del etileno en la maduración de frutos no climatéricos aún es desconocido, en general, su maduración no es afectada por tratamientos de etileno (Brady y Speirs, 1991; Pretel *et al.*, 1995; Tian *et al.*, 2000; Atta-Aly *et al.*, 2000). Los frutos del género *Capsicum* poseen un comportamiento único, si fueron recolectados durante o después del estado de transición (TR en Figura 1) el proceso de maduración continúa correctamente, esto no sucede si se recolecta durante el punto de maduración verde (VM) (Krajayklang *et al.*, 2000), lo que sugiere que los reguladores de maduración se activan a partir del estadío de transición.

I.6. El Chiltepín (Capsicum annuum L. var. *glabriusculum*)

El chiltepín (*C. annuum* L. var. *glabriusculum* (Dunal) Heiser & Pickersgill) es una variedad silvestre de la especie *C. annuum*, es un arbusto perenne con un vasto número de ramificaciones delgadas, que puede trepar en otros arbustos y alcanzar los 2 m de altura (Figura 2,a), puede llegar a vivir hasta 50 años, alcanzando la madurez reproductiva entre los 6 y 10 meses de edad, produce flores de mayo a agosto y frutos de junio a octubre, los frutos son pequeños (de 3 a 6 mm de diámetro) de un rojo vibrante, los frutos de chiltepín que se encuentran en los estados de Sonora, Sinaloa y Baja California poseen una forma redonda (Figura 2,d), mientras que en el resto del país poseen una forma oblonga (Figura 2,c) (Gentry, 1942; Nabhan *et al.*, 1990b).

que se encuentra distribuida desde el suroeste de Estados Unidos de América, México, cen-

troamérica, hasta Colombia (González-Jara *et al.*, 2011b), en México se encuentra distribuída desde Sonora hasta Chiapas a lo largo de toda la costa del pacífico (Nabhan *et al.*, 1990b; González-Jara *et al.*, 2011b; Miranda-Zarazúa *et al.*, 2010b; Pacheco-Olivera *et al.*, 2012b) y de Tamaulipas a la península de Yucatán por el Golfo de México (Gar, 2004). En el estado de Sonora se encuentra usualmente en los municipios cercanos a la sierra de 240 a 900 metros sobre el nivel del mar, las áreas más ricas en poblaciones de chiltepín silvestre son las del Río Sonora y la Sierra de Álamos (Figura 2,b)(Gentry, 1942; Felger y Moser, 1995), donde crece bajo la protección de plantas "nodrizasçomo Tepeguaje (*Lysiloma watsonii*), Mezquite, (*Prosopis velutina*), Cúmaro (*Celtis reticulata*) y el garambullo (*Celtis pallida*), que le ofrecen microclimas que le protegen de las inclemencias del clima y favorecen su crecimiento (Bañuelos *et al.*, 2008).

Se considera que el chiltepín podría ser la progenitora de los cultivares de *C. annuum* domesticados actuales (Singh, 2006b), y es considerada una importante fuente de recursos genéticos que puede ser utilizado para el mejoramiento de los chiles cultivados, para resolver problemas como susceptibilidad a enfermedades, plagas y estrés hídrico. (Hernández-Verdugo *et al.*, 1998b).

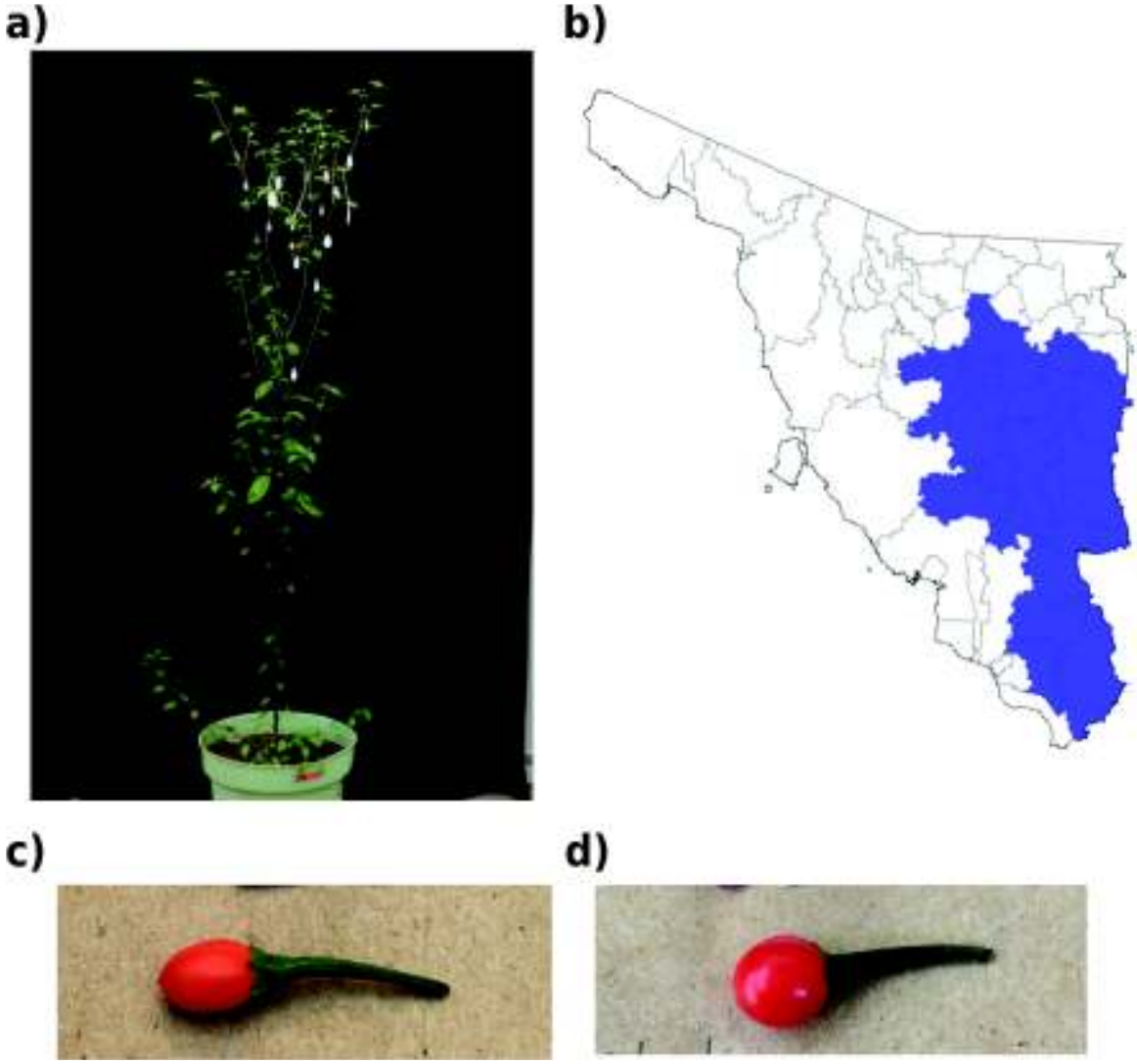


Figura 2. Generalidades del chiltepín: a) Planta de chiltepín cultivada en invernadero, b) Distribución geográfica del chiltepín en el estado de Sonora, elaborado con información del Consejo para la Promoción Económica del Estado de Sonora; c) Fruto de chiltepín de forma oblonga, d) fruto de chiltepín de Sonora con forma redonda

I.7. Regulación de la expresión génica en plantas

En plantas existen una amplia cantidad de factores que influyen en la regulación de genes, ya que a diferencias de los mamíferos, las plantas no pueden huir de condiciones de estrés o depredadores, han creado mecanismos para reaccionar ante cada uno de estos. Factores como la relación entre las horas de luz y oscuridad, la concentración de sales en el medio de crecimiento, la presencia de patógenos, las bajas temperaturas, estrés hídrico, entre otros, pueden producir un cambio en los patrones de expresión génica del organismo a través de uno de los métodos mencionados anteriormente o a través de un mecanismo propio del organismo en cuestión (Hoopes, 2008).

I.7.1. Mecanismo básico de regulación de expresión génica

Existen una amplia variedad de mecanismos que pueden regular la expresión génica en cada una de las etapas de producción de proteínas; sin embargo, la que es utilizada para controlar la mayoría de los genes es el proceso de transcripción, las etapas posteriores son utilizadas principalmente para refinar los patrones de expresión génica (Roundtree *et al.*, 2017; Beas *et al.*, 2009). Básicamente la regulación de expresión de genes se lleva a cabo por un proceso iniciado por proteínas "transductoras de señales", es decir, proteínas cuya función consiste en censar el ambiente y producir una modificación de la expresión de genes que producirán una respuesta a este estímulo ambiental. Este tipo de complejos son conocidos como sistema de regulación de dos componentes ya que utiliza una proteína sensora y una proteína reguladora de respuesta, esta última posee una región afín a secuencias de ADN. En los organismos eucariotas los genes posee regiones llamadas promotores que son reconocidas por proteínas especiales llamadas Factores generales de transcripción (entre ellos se encuentran las proteínas reguladoras de respuesta), estos a su vez reclutan a la enzima ARN polimerasa que se encargará de realizar la síntesis del ARN mensajero, la cantidad de regiones promotoras, su afinidad con los factores de transcripción y la

presencia de potenciadores (regiones distantes a los promotores que interactúan con los factores generales de transcripción y ayudan a reclutar la maquinaria de transcripción), o represores (regiones distantes al gen que inhiben su transcripción) regulan la velocidad de reclutamiento de la maquinaria de transcripción, es decir, la velocidad con que se pueden sintetizar los ARN mensajeros y, en algunos casos evitan que ciertos genes sean transcritos (Beas *et al.*, 2009). De esta forma se regula básicamente la expresión de cada uno de los genes en células eucariotas; sin embargo, existen otros factores que influyen en la respuesta de los mecanismos mencionados anteriormente, por ejemplo, factores epigenéticos como las modificaciones de histonas son comunes en todas las células eucariotas y afectan la capacidad de transcripción de un gen ya que para llevar a cabo la transcripción es necesario reclutar la ARN polimerasa al promotor del gen a transcribir, esto se lleva a cabo sin problema en el ADN desnudo; sin embargo, el ADN rara vez se encuentra en esta forma, usualmente se encuentra super-empaquetado gracias a nucleosomas que son complejos de proteínas llamadas histonas que tienen afinidad por el ADN, por lo cual este se enrolla alrededor de ellos produciendo una estructura de cuentas, que posteriormente por interacción entre las histonas produce una estructura helicoidal llamada fibra de 30 nm (Beas *et al.*, 2009). Para llevar a cabo la transcripción es necesario deshacer el enrollamiento del ADN en los nucleosomas, en las modificaciones de histonas se encuentran alteraciones epigenéticas que agregan grupos funcionales a las histonas, de este modo, la acetilación agrega grupos acetilos a las lisinas de la terminal N de las histonas, esto reduce la atracción entre el ADN y las histonas, lo cual permite a la maquinaria de transcripción acceder al ADN desnudo, posteriormente la desacetilación de las histonas les permite volver a su estructura original (Struhl, 1998). La metilación es otra modificación de histonas que agrega un grupo metilo a aminoácidos de las histonas y puede reprimir o activar la transcripción dependiendo de los aminoácidos metilados y la cantidad de grupos metilo agregados Greer y Shi (2012).

I.7.2. Otros mecanismos de regulación de expresión génica

Aunque la mayor parte de la regulación de genes se lleva a cabo durante el proceso de transcripción, existen otros mecanismos de regulación que pueden llevarse a cabo en varias etapas del proceso de biosíntesis de una proteína:

- Procesamiento del transcrito primario de ARN. A través de eventos de splicing y splicing alternativos es posible regular la isoforma del gen que será traducido, e incluso se ha mostrado que a través de los mecanismos NDM (Nonsense Decay Mechanism) pueden evitar que algunos ARN mensajeros sean traducidos, induciendo su degradación (Beas *et al.*, 2009; Chang *et al.*, 2007).

- Transporte del ARN mensajero al citoplasma. El ARN mensajero maduro debe salir del núcleo para ser traducido por los ribosomas en el citoplasma, proteínas transportadoras reconocen el ARN maduro y lo conducen por los poros nucleares al citoplasma, en este proceso es importante que el ARN mensajero sea estable, caso contrario será degradado antes de llegar a ser traducido. Antes de salir del núcleo el ARNm puede ser atacado por complejos de ARNs unidos a proteínas endonucleasas que procederán a degradarlo (Beas *et al.*, 2009).

- Traducción. Durante el proceso de traducción de ARN a polipéptido es posible regular la velocidad de traducción a través de elementos reguladores, miARNs un tipo de ARN pequeño en unión con proteínas como la ARGONAUTA1 se adhieren a ARN mensajeros y evitan que sean traducidos; existen también proteínas de control negativo de la traducción que pueden disminuir la producción de proteínas a partir de un ARN mensajero o incluso evitar completamente que se traduzca (Beas *et al.*, 2009).

- Degradación del ARN mensajero. Los ARN mensajeros tienen un tiempo de vida definido por varios factores, como la presencia de regiones ricas en Adeninas y Uracilos en la

región UTR 3', la presencia de sitios de anclaje para endonucleasas, el tamaño de la cola de poliadeninas, etc., estos factores a su vez pueden cambiar dependiendo de las condiciones de la célula regulando el tiempo de vida del ARN mensajero y con ello la probabilidad de que este sea traducido en proteínas (Beas *et al.*, 2009).

- Activación e inactivación de proteínas. Las proteínas sintetizadas en la traducción pueden ser objeto de una serie de modificaciones post-traduccionales que van desde eliminar segmentos de aminoácidos, agregarles grupos funcionales (fosfato, acilo, metilo, glúcido, etc), e incluso destruirlas completamente, estas modificaciones son reguladas y producen cambios que pueden modificar la actividad de las proteínas (Beas *et al.*, 2009; Mata *et al.*, 2005).

I.8. Transcriptómica

La transcriptómica es el estudio del transcriptoma, el término 'transcriptoma' se utiliza para referirse al conjunto de todos los transcritos (y sus abundancias) encontrados en un tejido específico, célula u organismo en una etapa específica de desarrollo o bajo cierto tratamiento, a diferencia del genoma que se mantiene constante durante la vida del organismo, el transcriptoma es altamente variable y cambia dependiendo de las etapas de desarrollo y como respuesta a los cambios en el ambiente Wang *et al.* (2009); Qian *et al.* (2014); Assis *et al.* (2014).

El análisis del transcriptoma es importante porque puede ayudar a comprender la relación entre el genotipo y fenotipo Evans (2015), proveer información sobre cómo la expresión génica cambia entre organismos Melé *et al.* (2015) o bajo condiciones o tratamientos específicos Geniza y Jaiswal (2017), sobre los mecanismos que controlan el destino celular Li *et al.* (2014), el proceso de desarrollo Rangan *et al.* (2017), e incluso el progreso de enfermedades Lowe *et al.* (2017); Sandberg (2014).

La transcriptómica moderna utiliza tecnologías de secuenciación de nueva generación (NGS), para realizar secuenciación de ARN (RNA-Seq) que producen lecturas de secuenciación de manera masiva y paralela, de este modo es posible analizar todos los tipos de transcritos, incluyendo ARN mensajero (ARNm), microARNs (miARN) y diferentes tipos de ARN no codificantes (ncARN) como los ARN no codificantes largos (lncARN) (Milward *et al.*, 2016; Wang *et al.*, 2009). Permite el estudio de estructuras de isoformas, extraer información alélica a través de SNPs Haas *et al.* (2013), y realizar análisis de expresión diferencial entre grupos Zhang *et al.* (2016). A diferencia de las tecnologías previas, es posible realizar estudios de RNA-Seq sin conocimiento previo de las secuencias o del organismo de interés Costa-Silva *et al.* (2017), además, la cantidad de ARN necesario para realizar la secuenciación es menor (nanogramos) comparado con los microarreglos (microgramos) Lowe *et al.* (2017) y combinado con la amplificación linear de ADNc puede ser utilizado para realizar experimentos al nivel de una sola célula Hashimshony *et al.* (2012); Kolodziejczyk *et al.* (2015).

I.9. Análisis de la expresión génica en plantas de chile

Durante las últimas décadas se han realizado diferentes estudios del transcriptoma de frutos de chile; por ejemplo, se creó una base de datos ESTs (secuencias cortas obtenidas de ADN complementario), utilizadas como identificadores de genes. Esta base de datos fue diseñada como un punto de referencia para la detección de genes en plantas de chile; se utilizaron 11 tipos diferentes de tejido, de los cuales se obtuvieron 21 librerías de ADNc, de éstas, fueron secuenciadas 122,588 ESTs y se refinaron 116,412 ESTs (Kim *et al.*, 2008).

Se ha reportado el ensamble del transcriptoma de dos tipos de chile, a partir de secuencias generadas por secuenciación Sanger (>125,000 ESTs) y la plataforma Illumina NGS (200 millones de lecturas), con el objetivo de identificar SNPs y SSRs como marcadores moleculares para su reproducción o polimorfismos de nucleótido único para genotipado (Ashrafi *et al.*, 2012).

Góngora-Castillo *et al.* (2012) documentan la creación de una base de datos del transcriptoma de *Capsicum* generada de forma híbrida entre una colección de ESTs extraídos de cinco tejidos de *Capsicum annuum* (raíz, tallo, hoja, flores y fruto) secuenciados con el método Sanger y múltiples transcriptomas obtenidos de las hojas por pirosecuenciación (secuenciación 454). Esta colección de secuencias contiene 60,000 singletons y 32,314 contigs de alta calidad, de los cuales el 75.5 % es similar a registros en bases de datos de proteínas y ácidos nucléicos, mientras que el 23 % no había sido reportado previamente para *Capsicum annuum*.

Lee y Choi (2013) realizaron un análisis global del transcriptoma de las plantas de chile sometidas a diferentes tipos de estrés (biótico y abiótico), con el objetivo de identificar los componentes (regulones) que se activan durante dichas condiciones. Se encontró que el ácido salicílico (SA) puede afectar la activación de genes de respuesta a condiciones de estrés abiótico, mientras que los genes responsables del metil jasmonato (MeJA) y etileno (ET) se activan principalmente bajo estímulos bióticos y los genes responsables del ácido abscísico (ABA) se activan durante ambos tipos de estrés. Osorio *et al.* (2012) llevaron a cabo un estudio integrativo analizando el transcriptoma y el metaboloma de tomate (climatérico) y chile (no climatérico) durante el proceso de maduración de fruto. Se utilizaron muestra en 11 etapas de maduración: 14, 20, 26, 34, 51, 52, 53, 55, 57, 62 y 68 DDA (Días después de antesis), para identificar los genes expresados en el pericarpio del fruto de chile se utilizaron microarreglos TOM1, con una creación de perfil de abundancias calculada por hibridación de dos colores. El microarreglo TOM1 contiene 12,899 sondas de ESTs para obtener un total de 8,500 genes de tomate.

Los resultados muestran que el nivel de expresión génica en cualquier etapa de maduración de fruto es al menos 2 veces mayor que la referencia (14 DDA). Existe una clara transición entre las etapas de desarrollo y maduración de fruto detectada a 42 DDA en tomate y 26-34 DDA en chile; por último, se observó que las etapas finales en la maduración tanto de tomate como de chile son muy similares (chile 55-68 DDA, tomate 42-47 DDA). Además, se detectó que ambos

frutos tienen señalamiento mediado por etileno similar; sin embargo, la regulación es diferente y puede verse reflejada en la alteración de la sensibilidad de etileno en chiles, donde los genes involucrados en la biosíntesis no se encuentran inducidos en los frutos de chile como en los de tomate; sin embargo, los efectos en pared celular, síntesis de carotenoides y otros que son inducidos por etileno, se comportan de forma similar entre tomate y chile.

Liu *et al.* (2013) realizaron un análisis de transcriptoma de *Capsicum frutescens* 'xiaomila', el único chile encontrado de forma silvestre en China y que se ha adaptado a condiciones poco favorables, como temperaturas altas, terrenos con una humedad elevada, bajos niveles de nutrientes y luminosidad baja, además de poseer resistencia a enfermedades. Se utilizaron muestras de pericarpio de frutos a 20 y 30 DDA (Días después de Antesis), de las cuales se extrajo ARN, con el que se crearon bibliotecas de cDNAs que fueron secuenciadas utilizando la plataforma Illumina HiSeq para obtener un total de 27,533,332 lecturas que se ensamblaron con la herramienta Trinity, con 54,045 genes no redundantes de los cuales el 45.82 % son mayores a 500pb, 22.65 % mayores a 1000pb Y 5.07 % mayores a 2,000pb. El 92.65 % de los genes ensamblados tuvieron un alineamiento con secuencias homólogas en bases de datos públicas, en el genoma de papa o tomate, o en la base de datos de ESTs de chile (*C. annuum*). Se cree que los genes restantes (7.35 %) pueden ser transcritos no reportados previamente y probablemente sean genes únicos propios de la especie *C. frutescens*.

Se identificaron 197 transcritos que codifican enzimas putativas relacionadas con las rutas metabólicas responsables de la biosíntesis de capsaicinoides, utilizando un análisis KEGG. De estos, 162 son genes candidatos que no han sido reportados previamente, 21 de ellos se encuentran involucrados en la ruta metabólica de la valina, leucina e isoleucina, 30 se encuentran relacionados con la ruta metabólica de ácidos grasos, 109 genes candidatos forman parte de la ruta metabólica de la biosíntesis de los fenilpropanoides, 56 genes participan en el metabolismo de la fenilalanina y 32 genes pertenecen a la ruta metabólica de la biosíntesis de fenilalanina, ti-

rosina y triptófano. Además, se utilizaron los genes ensamblados para predecir un total de 9,150 SNPs y 4,072 marcadores moleculares SSR entre *C. frutescens* y *C. annuum*, lo cual podría proporcionar un medio para estudiar el polimorfismo de los chiles.

Actualmente, la única investigación que muestra un panorama general del transcriptoma de *Capsicum* durante el proceso de maduración de fruto es la reportada por Martínez-López *et al.* (2014), en ella, se llevó a cabo recolectando muestras de frutos de chile serrano (*Capsicum annuum* L.; 'Tampiqueño 74') a los 10, 20 40 y 60 días después de antesis (DDA). Se produjeron 15,550,468 lecturas utilizando la plataforma Illumina MiSeq, estas lecturas fueron ensambladas *de novo* en 34,066 genes.

Se descubrió que en los intervalos de 10 a 20, y 40 a 60 DDA se presentaron los cambios en niveles de expresión más drásticos y en el último intervalo, entre 40 y 60 DAA, se presentaban el 49 % de los cambios en niveles de expresión significativos, caracterizados por un descenso en la expresión de genes, marcando el fin de la maduración y el comienzo de la senescencia. También se infirió que los genes relacionados con la biosíntesis de metabolitos importantes como los capsaicinoides y el ácido ascórbico tienen una expresión más abundante a los 20 DDA, mientras que otros como los genes relacionados en la biosíntesis de carotenoides presentaron una mayor abundancia a los 60 DDA.

Dada la importancia económica del chile en México ya que es el segundo productor a nivel mundial, y considerando que nutricional y biológicamente el proceso de maduración de fruto es de especial interés, ya que en éste se llevan a cabo la producción de los metabolitos secundarios, se considera importante comparar el transcriptoma de otras variedades de *Capsicum*, en particular, los datos del chiltepín (*Capsicum annuum var. glabriusculum*), que es una variedad silvestre del chile cultivado actual, por lo cual es considerado una fuente de recursos genéticos importantes que pueden ser utilizados para el mejoramiento de los cultivares actuales.

II. Hipótesis

La domesticación del chile ha generado cambios cuantitativos y cualitativos en la expresión génica durante el desarrollo del fruto de chile.

III. Objetivos

III.1. Objetivo General

Determinar los cambios cuantitativos y cualitativos en la expresión génica durante el desarrollo y maduración del fruto en dos variedades de chile, una domesticada (serrano 'tampiqueño 74') y otra silvestre (chiltepín colectado en la sierra de Sonora).

III.2. Objetivos Específicos

- Identificar los genes homólogos expresados en las variedades cultivada y silvestre.
- Estimar cuantitativamente las diferencias de expresión génica entre las dos variedades.
- Determinar las rutas metabólicas y procesos de morfogénesis que se encuentran modificados entre las dos variedades.
- Determinar las diferencias en los patrones de expresión de genes asociados al proceso de maduración de fruto en las dos variedades.
- Identificar ARNs no codificantes expresados durante el proceso de maduración de fruto en Chiltepín.

IV. MATERIALES Y MÉTODOS

Para cumplir con los objetivos planteados, el trabajo experimental se dividió en dos fases principales, la fase 1 (Figura 3) corresponde a la obtención de material biológico, extracción de ARN y secuenciación, mientras que la fase 2 (Figura 5) consiste en el análisis bioinformático que se realizó una vez obtenidos los datos de secuenciación resultado de la fase 1. Es importante mencionar que la fase 1 correspondiente a este trabajo de investigación fue realizada previamente por integrantes de nuestro equipo de trabajo, por lo cual, los resultados plasmados en este trabajo de tesis corresponden al análisis bioinformático de lecturas de RNA-Seq de chiltepín y Serrano en dos etapas de maduración.

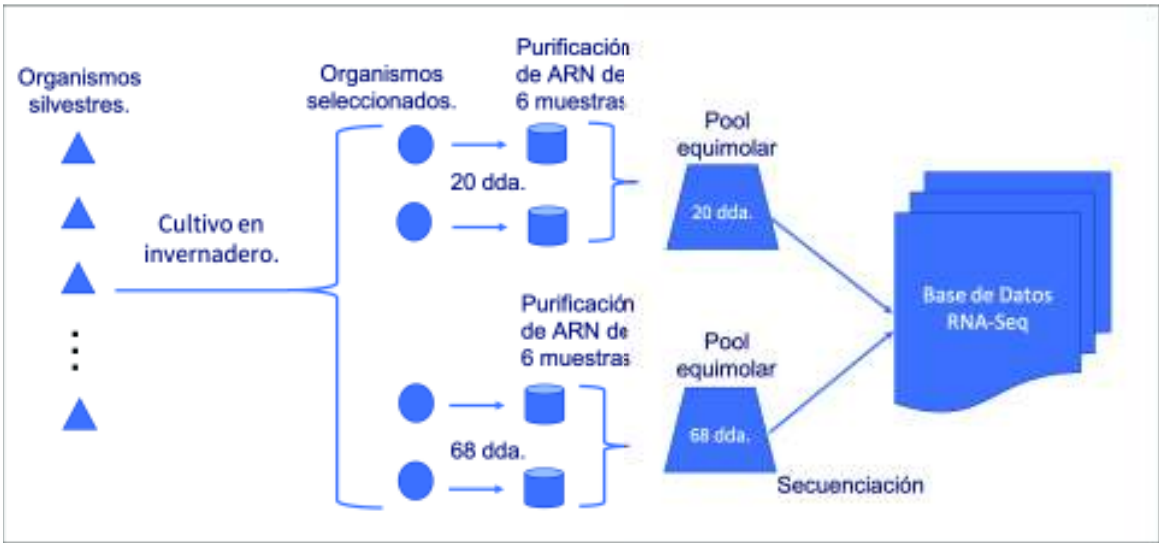


Figura 3. Diseño experimental de la fase 1 correspondiente a la obtención de material biológico, extracción de ARN y secuenciación de ARN.

IV.1. Fase 1: Obtención de material biológico, extracción y secuenciación de ARN

IV.1.1. Obtención del material biológico de chiltepín

Se recolectaron frutos maduros de plantas silvestres de chiltepín de poblaciones localizadas en la localidad de Tepoca en el municipio de Yécora, Sonora (Figura 4), México (28.43°N, -109.25°W). Se retiraron las semillas de los frutos secos, se desinfectaron y se sembraron en macetas de plástico con mezcla de tierra previamente esterilizada en el invernadero del Laboratorio Nacional de Genómica (LANGEBIO-CINVESTAV, Irapuato, México) y se germinaron bajo condiciones controladas en un diseño experimental aleatorio, de acuerdo al protocolo descrito por del Rosario Abraham-Juárez *et al.* (2008). Una vez que las plántulas desarrollaron hojas verdaderas fueron transplantadas a macetas de tamaño medio (5 litros) y cuando las plantas comenzaron el proceso de floración se etiquetaron flores de forma individual durante la antesis y se recolectaron frutos a 20 (fruto verde) y 68 (fruto maduro) días después de antesis (DDA) de forma aleatoria de diferentes plantas, los frutos fueron limpiados con etanol, colocados en nitrógeno líquido y almacenados en ultracongelador a -80°C para su posterior análisis.

IV.1.2. Extracción de ARN, preparación de librerías y secuenciación

Para la extracción de ARN se eligieron aleatoriamente 6 frutos del conjunto de frutos recolectados a 20 y 68 DDA, se molieron los frutos completos (pericarpio, placenta y semillas) utilizando un equipo QIAGEN TissueLyser con perlas e intervalos de disrupción de 1 min a 30 Hz (1,800 oscilaciones por minuto).

La extracción de ARN se realizó en cada fruto de chiltepín por separado utilizando el NucleoSpin RNA Plant kit (Macherey-Nagel) de acuerdo a las intrucciones del fabricante, las muestras fueron sometidas a tratamiento con DNase I (Macherey-Nagel) de acuerdo a las re-

Figura 4. Zona de recolección de frutos de poblaciones silvestres de chiltepín, localidad de Tepoca en el Municipio de Yécora, Sonora, México (28.43°N, -109.25°W).

comendaciones del fabricante, para remover ADN contaminante. Se crearon alícuotas de 100mg de cada extracción de ARN y posteriormente fueron combinadas equimolarmente en una sola muestra de ARN total. El proceso se repitió para cada una de las etapas de maduración para crear réplicas biológicas independientes y obtener un total de 4 muestras.

Se obtuvo la concentración de ARN total utilizando un espectrofotómetro Nanodrop ND-1000 (Thermo Scientific Nanodrop), se evaluó la integridad del ARN por electroforesis en gel de 1 % de agarosa, y en un equipo Agilent Bioanalyzer 2100.

Las cuatro muestras de ARN total fueron secuenciadas en el Laboratorio de Servicios Ge-

nómicos (CINVESTAV-LANGEBIO, Irapuato, México). Se utilizó un kit de preparación de muestras Illumina TruSeq RNA sample v2 de acuerdo a las instrucciones del fabricante para crear librerías de ADNc utilizando 12 μg de ARN total de cada réplica biológica, las cuatro genotecas resultantes fueron secuenciadas en ambos extremos (5' y 3') utilizando una plataforma Illumina NextSeq 500 de acuerdo a las instrucciones del fabricante, para producir alrededor de 260 millones de lecturas pareadas 2x150pb.

IV.2. Fase 2: Análisis bioinformático

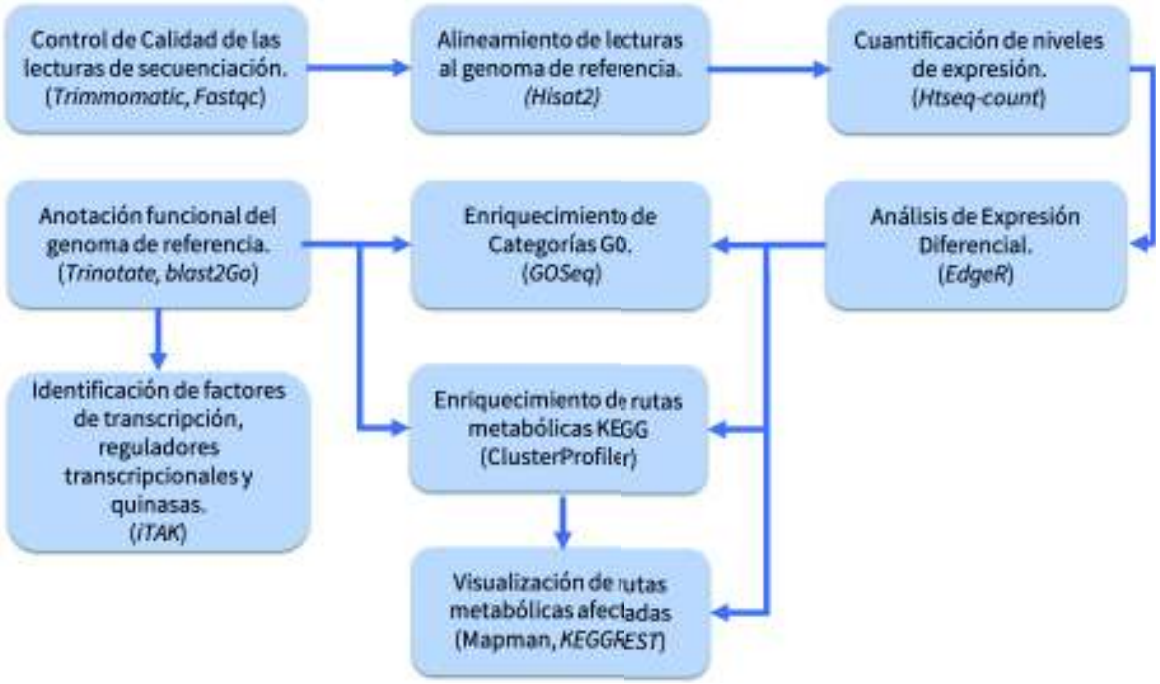


Figura 5. Diseño experimental de la fase 2 correspondiente al análisis bioinformático de los datos de secuenciación obtenidos en la fase 1 del experimento.

IV.2.1. Análisis y control de calidad de las lecturas de secuenciación

Las lecturas de secuenciación obtenidas de la fase 1 fueron sometidas a un proceso de análisis de calidad utilizando la herramienta FastQC V. 0.11.5 para verificar la calidad de las secuen-

cias por base, distribución de longitud de lecturas, contenido de GC por secuencias y evaluar la presencia de adaptadores de secuenciación. Una vez identificadas las áreas que podían ser mejoradas se utilizó el software Trimmomatic V. 0.36 (Bolger *et al.*, 2014) con los filtros correspondientes para eliminar adaptadores de secuenciación de Illumina presentes en las muestras, realizar un filtrado suave por ventana deslizante de 4 bases que recorta lecturas si la media de la calidad phred en las bases de la ventana disminuye de un valor definido en 15 (phred score). Por último, se agregó un filtro para eliminar lecturas de secuenciación que después del proceso de filtrado presentaran una longitud menor a 50 bases. Posterior al proceso de filtrado se realizó nuevamente el análisis de las lecturas procesadas utilizando FastQC, para comprobar que se obtuvieron los resultados deseados (apéndice A.1).

IV.2.2. Alineamiento al genoma de referencia y estimación de abundancias

Las lecturas de secuenciación previamente filtradas fueron alineadas al genoma de referencia de *Capsicum annuum* cv. CM334 v. 1.6 (Kim *et al.*, 2014) utilizando la herramienta Hisat2 V. 2.1.0 (Kim *et al.*, 2015) del reconocido protocolo 'new tuxedo', este programa permite analizar la expresión de genes, identificar genes novedosos y variantes de isoformas producidas por splicing alternativo. Se utilizó Hisat2 para extraer los sitios de splicing del archivo de anotación V. 2.0 de genoma de referencia (`http://peppergenome.snu.ac.kr`; último acceso junio, 2020), este archivo se utilizó al ejecutar Hisat2 para realizar el alineamiento de lecturas de secuenciación al genoma de referencia con el parámetro –known-splicesite-infile, en el resto de los parámetros se utilizaron los valores por defecto (apéndice A.2).

Se utilizó el programa informático HTseq-count V. 0.11.0 (Anders *et al.*, 2014) para estimar la abundancia de genes en función de los alineamientos de las lecturas de secuenciación al genoma de referencia, se utilizaron parámetros por defecto y la opción 'union', por lo cual fue necesario desechar los alineamientos de lecturas a más de un sitio en el genoma de referencia

utilizando el programa samtools (`http://www.htslib.org`, apéndice A.3).

IV.2.3. Anotación génica

La anotación génica consiste en la asignación de información sobre la función biológica de los genes, para esto se pueden utilizar diferentes enfoques, el más común es la transferencia por homología, que consiste en la búsqueda de las secuencias de interés en bases de datos públicas como UniprotKB, NCBI nr, entre otras (Conesa y Götz, 2008).

IV.2.3.1. Anotación Gene Ontology

La anotación de los genes de *Capsicum annuum* cv. CM334 se realizó utilizando la herramienta Blast2GO Conesa *et al.* (2005), la cual utiliza BLAST para realizar búsquedas de las secuencias de interés en las bases de datos curadas de UniProtKB/Swiss-Prot Consortium (2018) para identificar secuencias similares, también con Blast2GO se realizó la asignación de términos de Gene Ontology (GO) y de su versión generalizada slimGO para identificar las funciones moleculares, los componentes celulares a los que pertenecen y los procesos biológicos en los que están relacionados (Ashburner *et al.*, 2000).

IV.2.3.2. Identificación de genes homólogos con tomate

El mejor alineamiento recíproco (RBH por sus siglas en inglés) es la técnica más común para la detección de genes homólogos en dos especies, un RBH es identificado cuando las proteínas codificadas por dos genes en diferentes genomas son el mejor alineamiento una de otra. Sin embargo, en este caso el método de RBH no es el mejor enfoque ya que el genoma de chile ha pasado a través de eventos de duplicación génica que no ocurrieron en tomate, resultando en un genoma cuatro veces mayor que el de tomate (Kim *et al.*, 2014), para inferir relaciones de homología entre genes de tomate y chile se utilizó el enfoque de agrupamiento de ortólogos

implementado en el software Inparanoid V. 4.2 (Remm *et al.*, 2002), que permite identificar grupos de genes parálogos en un genoma A y su(s) correspondiente(s) ortólogo(s) en el genoma B.

Para la identificación de grupos de ortólogos entre tomate y chile se utilizaron las proteínas codificadas por el genoma de referencia de chile (*Capsicum annuum* cv. CM334 v. 1.6) y las secuencias de proteínas del genoma de referencia de tomate (ITAG Release 3.2), se utilizaron un valor de corte de 40 bits, un valor de confianza de 0.05, valor de corte para el traslape de secuencias de 0.5, valor de corte para la fusión de grupos de 0.5 y una matriz de puntajes BLOSUM62.

IV.2.3.3. Anotación KEGG

La Enciclopedia de Genes y Genomas de Kyoto (KEGG, `https://www.genome.jp/kegg/`) es una base de datos integral que permite realizar un enlace entre genes y diferentes niveles de funciones celulares de la célula y organismos (Kanehisa *et al.*, 2015). La asignación de números K (identificadores KEGG) se realizó a través de la plataforma del servidor de anotación automática de KEGG (KAAS) (Moriya *et al.*, 2007) y a través del método mejor alineamiento bi-direccional (BRH) con el genoma de referencia de *Capsicum annuum* cv. Zunla-1 (Qin *et al.*, 2014), el único genoma de chile que posee un registro en la base de datos de KEGG.

IV.2.3.4. Identificación de genes en zonas de barrido selectivo

Estudios previos han identificado regiones con señales de barrido selectivo en el genoma de *Capsicum annuum* cv. Zunla-1 (Qin *et al.*, 2014) a través de sus niveles de polimorfismo que son considerablemente más bajos que en el resto del genoma, esto sugiere que dichas regiones fueron modificadas durante el proceso de domesticación de chile y los genes que se encuentran en ellas estarían estrechamente relacionados con las diferencias entre plantas cultivadas y silvestres. Para

identificar los genes de Chiltepín y Serrano pertenecientes a regiones de barrido selectivo que se encuentran expresados durante el proceso de maduración de fruto se utilizó el método de mejor alineamiento bi-direccional (BRH) con el genoma de referencia de *Capsicum annuum* cv. Zunla-1 y se seleccionaron los genes de barrido selectivo reportados por Qin *et al.* (2014).

IV.2.4. Análisis de ARNs no codificantes

IV.2.4.1. Ensamblado de transcriptoma *ab initio*

Para realizar la identificación de ARNs no codificantes potenciales se utilizaron los archivos bam generados por Hisat2 como resultado del alineamiento de las lecturas de secuenciación al genoma de referencia, se utilizó el ensamblador de transcritos StringTie v1.3.4d (`https://ccb.jhu.edu/software/stringtie/`) para construir transcritos potenciales a partir del alineamiento de las lecturas de cada una de las librerías al genoma de referencia. Posteriormente se utilizó nuevamente Stringtie con la opción *merge* para fusionar los transcritos potenciales detectados previamente y construir el transcriptoma *ab initio* y *m* 200 para desechar transcritos con tamaño menor a 200 pb.

IV.2.4.2. Identificación de ARNs no codificantes potenciales

El transcriptoma *ab initio* ensamblado se comparó con genoma de referencia de *Capsicum annuum* cv. CM334 v. 1.6 utilizando la herramienta GffCompare v0.10.6 (`https://ccb.jhu.edu/software/stringtie/gffcompare.shtml`) con parámetros por defecto, durante su análisis GffCompare realiza la asignación de códigos de clase a los transcritos en función de su relación con los transcritos del genoma de referencia, se utilizaron comandos del sistema operativo UNIX para extraer la información de códigos de clase de los transcritos, obtener estadísticas básicas correspondientes y extraer la lista de transcritos potenciales ARN no codificante con códigos de clase *i* (intronic contained), *o* (exonic overlap), *u* (lincRNA), *j* (intronic overlap) y *x*

(exonic overlap antisense).

IV.2.4.3. Confirmación de ARNs no codificantes

Una vez obtenida la lista de transcritos potenciales ARN no codificantes, se extrajo su secuencia a partir del genoma de referencia y el transcriptoma ensamblado *ab initio* utilizando la herramienta GffRead versión 0.9.12 (`http://ccb.jhu.edu/software/stringtie/gff.shtml`) con parámetros por defecto. Posteriormente se identificaron los transcritos con potencial codificante utilizando la calculadora de potencial codificante CPC2 versión 0.1 (`http://cpc2.cbi.pku.edu.cn`) con parámetros por defecto y, la herramienta TransDecoder versión 5.5.0 (`https://github.com/TransDecoder/TransDecoder/wiki`) con la opción *longest.orfs* para obtener los marcos de lectura abiertos de los transcritos, los cuales fueron comparados con las bases de datos UniProtKB/Swiss-Prot y Pfam utilizando blast, estos resultados se ingresaron a TransDecoder con la opción *predict* para identificar los transcritos codificantes. Se filtraron todos los transcritos que presentaron potencial codificante de acuerdo a CPC2 o TransDecoder, para obtener una lista de transcritos ARN no codificantes.

IV.2.4.4. Anotación de ARNs no codificantes

Una vez filtrados los transcritos con potencial codificante, obtuvimos una lista de ARNs no codificantes, sus secuencias fueron utilizadas para su identificación con la herramienta Infernal v1.1.3 (`http://eddylab.org/infernal/`) que usa la base de datos Rfam (versión 14.1) que contiene modelos de covarianza y modelos ocultos de Markov de las secuencias y estructuras secundarias de familias de ARNs no codificantes.

IV.2.4.5. Identficación de micro ARNs

Para la identificación de microARNs se utilizaron las secuencias de ARNs no codificantes obtenidos previamente, las cuales fueron comparadas via blast con la base de datos MiRBase (release 22, `http://www.mirbase.org`) que contiene secuencias de micro ARNs maduros y secuencias precursoras de diferentes especies, incluyendo 73 especies de *viridiplantae*. Debido al tamaño pequeño de las secuencias incluídas en MiRBase se utilizó blastn con la opción *-task blastn-short*, además se incluyeron los parámetros *-perc_identity 0.9 -evalue 10* para filtrar los alineamientos de baja calidad.

IV.2.5. Análisis estadístico

IV.2.5.1. Análisis de expresión diferencial

Una de las principales aplicaciones de la transcriptómica es la identificación de genes expresados diferencialmente (DEGs) entre dos condiciones, es decir, genes que muestran una diferencia en sus niveles de expresión entre dichas condiciones (Soneson y Delorenzi, 2013). Se utilizó el paquete EdgeR v. 3.24.3 (Robinson *et al.*, 2009) de Bioconductor en R v. 3.5.3 (R Core Team, 2019) para identificar genes expresados diferencialmente en los contrastes entre las diferentes comparaciones de las dos accesiones: Chiltepín (silvestre) y Serrano (cultivado) y en los tiempos de maduración: 20 DDA y 60(68) DDA. Para esto se realizó un diseño con un total de 5 contrastes (Tabla I) que incluye comparaciones entre ambas accesiones en las mismas etapas de maduración (Ch20-St20 y Ch68-St60), la comparación entre las dos etapas de maduración en cada accesión (Ch20-Ch68 y St20-St60) y la comparación de estas dos últimas (Ch2068-St2060).

Utilizando EdgeR se filtraron genes con abundancia estimada nula y se descartaron genes con una media del logaritmo de las abundancias por millón menor a 1 (Ave log count per million $<$ 1) en todas las librerías. Posteriormente, con el fin de corregir el sesgo por diferencia de profun-

Tabla I. Diseño de contrastes creados para el análisis de expresión diferencial con EdgeR, se descartaron posibles comparaciones que carecen de importancia biológica.

Contraste	Accesión	Tiempo de muestreo		Accesión	Tiempo de muestreo
Ch20-St20	Chiltepin	20	vs	Serrano	20
Ch68-St60	Chiltepin	68	vs	Serrano	60
Ch20-Ch68	Chiltepin	20	vs	Chiltepin	68
St20-St60	Serrano	20	vs	Serrano	60
Ch2068-St2060	(Chiltepin 20 vs Chiltepin 68)		vs	(Serrano 20 vs Serrano 60)	

didad de secuenciación se realizó una normalización *trimmed mean of M values* (TMM) para calcular factores de normalización para cada librería, se calcularon las dispersiones: *common, trended* y *tagwise* y se implementó un modelo linear generalizado (GLM) con una distribución binomial negativa para ajustar las abundancias de los genes. Se utilizaron las funciones decideTestsDGE y glmTreat para identificar genes DE tomando en cuenta la ejecución de pruebas múltiples, se seleccionaron genes con *fold change* absoluto ≥ 4 y un *False discovery rate* (FDR) < 0.01 (apéndice A.4).

IV.2.5.2. Análisis de enriquecimiento de términos GO

Analizar los genes expresados diferencialmente (DEGs) en un experimento de RNA-Seq puede ser un trabajo complejo, debido al uso de secuenciación masiva para realizar el perfil transcriptómico y dependiendo de la profundidad de secuenciación un experimento puede contener desde cientos hasta miles de DEGs. Para ello, existen algunos recursos como la base de datos de Gene Ontology (GO, `http://geneontology.org/`) que consiste en un vocabulario jerárquico de términos con definiciones lógicas que describen relaciones entre genes y sus funciones moleculares, procesos biológicos en los que participan y componentes celulares a los que pertenecen, este tipo de anotación permite analizar los DEGs agrupándolos en grupos a través de los niveles jerárquicos de los términos GO, otro paso importante es el análisis de términos GO enriquecidos, esto consiste en enfocarse en los términos GO que estadísticamente se encuentran 'enriquecidos' (a los cuales pertenecen más DEGs) en contraste con el total de genes expresados

en el experimento (Pomaznoy *et al.*, 2018). Para este trabajo de investigación se realizó el análisis de los términos GO enriquecidos utilizando el paquete GOSeq v. 1.38.0 (Young *et al.*, 2010) de Bioconductor, que implementa una distribución de Wallenius para aproximar una distribución nula de los términos GO y evita el sesgo por longitud de secuencia inherente a los datos de RNA-Seq, realiza pruebas para evaluar si los términos GO se encuentran sobre-representados o bajo-representados en comparación con el conjunto total de genes expresados diferencialmente. Se clasificaron como términos GO enriquecidos estadísticamente relevantes aquellos con un *False Discovery Rate* (FDR) < 0.05.

IV.2.5.3. Análisis de enriquecimiento de rutas metabólicas KEGG

El análisis de rutas metabólicas enriquecidas se llevó a cabo utilizando la librería ClusterprofileR v. 3.10.1 (Yu, 2018) de Bioconductor, para identificar posibles procesos biológicos afectados por las condiciones del experimento (accesión y tiempo de muestreo), para cada contraste se creó una lista de rutas metabólicas KEGG enriquecidas filtrando por *False Discovery Rate* (FDR < 0.05).

IV.3. Datos de RNA-Seq de Serrano 'Tampiqueño 74'

Los datos de Serrano 'Tampiqueño 74' fueron obtenidos de un trabajo de investigación actual en el cual hemos participado. Las semillas utilizadas se obtuvieron del germoplasma del grupo de trabajo, las semillas fueron germinadas, las flores fueron etiquetadas y los frutos recolectados y el ARN fue extraído siguiendo el mismo protocolo que se utilizó en los frutos de chiltepín. La secuenciación se realizó enviando las muestras de ARN a Novogene (`https://en.novogene.com/`) donde se crearon librerías que fueron secuenciadas en una plataforma Illumina NovaSeq para generar lecturas pareadas de 150pb. La empresa Novogene proporcionó una matriz de abundancias por librería, la cual se obtuvo realizando el alineamiento

de las lecturas de secuenciación al genoma de referencia de *Capsicum annuum* cv. CM334 v. 1.6 (Kim *et al.*, 2014) con la herramienta Tophat2 (predecesor de Hisat2) y las abundancias de transcritos fueron obtenidas utilizando la herramienta HTSeq en modo 'union'.

V. RESULTADOS

V.1. Secuenciación de ARN

Se obtuvieron cuatro muestras de chiltepín silvestre (CH; *Capsicum annuum var. glabriusculum*) y el chile cultivado Serrano Tampiqueño (ST; *Capsicum annuum* L. cv. 'Tampiqueño 74') recolectados en dos etapas: fruto verde en desarrollo a 20 días después de antesis (DDA) y fruto rojo maduro a 60 DDA (Serrano) y 68 DDA (Chiltepín Korkmaz *et al.*, 2014). Todas las muestras fueron secuenciadas y produjeron entre 18.8 y 38.8 millones de lecturas *paired-end* de 150 bases de longitud por muestra, para un total de 260,680,166 lecturas de secuenciación en bruto (Tabla II).

Tabla II. Resultado de la secuenciación de las 4 muestras de chiltepín, cantidad de lecturas de secuenciación en bruto por muestra.

Librería	Descripción	Lecturas de secuenciación en bruto	Porcentaje (%)
1	CH 20 DDA, réplica 1	76,679,402	29.41
2	CH 20 DDA, réplica 2	68,923,334	26.43
3	CH 68 DDA, réplica 1	37,520,926	14.39
4	CH 68 DDA, réplica 2	77,556,504	29.75
	Total:	260,680,166	100

V.2. Análisis y control de calidad de las lecturas de secuenciación

V.2.1. Análisis de calidad de lecturas en bruto

Una vez obtenidas las lecturas de secuenciación fueron sometidas a un análisis de calidad utilizando la herramienta FastQC, en la tabla III podemos observar las estadísticas básicas de las

lecturas de secuenciación en bruto, en 8 archivos, 2 por muestra correspondientes a las lecturas forward y reverse, se obtuvieron de 18 a 38 millones de lecturas con entre 41 y 43 % de GCs e intervalos de tamaño de lectura de 35-151 pb.

Tabla III. Resumen del resultado de análisis de calidad de lecturas de secuenciación en bruto de muestras de chiltepín realizado con FastQC, cantidad de lecturas de secuenciación por cada uno de los archivos de lecturas por muestra (forward y reverse), porcentaje de GC e intervalo de tamaño de lecturas de secuenciación.

Librería	Descripción	Orientación	No. de lecturas	GC (%)	tamaño (pb)
1	CH 20 DDA, réplica 1	Forward	38,339,701	42	35-151
1	CH 20 DDA, réplica 1	Reverse	38,339,701	42	35-151
2	CH 20 DDA, réplica 2	Forward	34,461,667	43	35-151
2	CH 20 DDA, réplica 2	Reverse	34,461,667	42	35-151
3	CH 68 DDA, réplica 1	Forward	18,760,463	41	35-151
3	CH 68 DDA, réplica 1	Reverse	18,760,463	41	35-151
4	CH 68 DDA, réplica 2	Forward	38,778,252	42	35-151
4	CH 68 DDA, réplica 2	Reverse	38,778,252	42	35-151
		Total:	260,680,166		

FastQC es una herramienta diseñada para evaluar diferentes aspectos de lecturas de secuenciación genómica, algunas de estas métricas son útiles para la evaluación de lecturas de RNA-Seq, por ejemplo, en la figura 6 se muestra la calidad por base de las lecturas de secuenciación de las cuatro muestras de Chiltepín, en el eje X se muestra la posición y en el eje Y se muestra una serie de gráficas de cajas y bigotes con la calidad media en azul, podemos observar que en los cuatro casos la gráfica de las lecturas forward (izquierda) presentan una calidad media por encima del valor phred 28 considerado comúnmente como el límite inferior de un intervalo de calidad bueno; sin embargo, existen algunos bigotes en los extremos 3' donde la calidad disminuye hasta intervalos de calidad 'adecuada' (20-28). En el caso de los archivos correspondientes a las secuencias reverso (derecha), se puede observar una calidad media similar a la de sus parejas; sin embargo, en este caso la calidad de las lecturas en el extremo 3' disminuye significativamente hasta el intervalo considerado no deseable (0-20), esto es normal en la secuenciación en plataformas Illumina y es corregible con filtros de calidad. Existen algunas

métricas como el contenido por base de secuencia, los niveles de duplicación de secuencias y el contenido de k-meros que de acuerdo a la herramienta FastQC necesitan ser corregidos en todas las muestras, no obstante, este comportamiento es común debido a la naturaleza de los experimentos RNA-Seq, por lo cual no es necesario realizar un procesamiento para corregirlo.

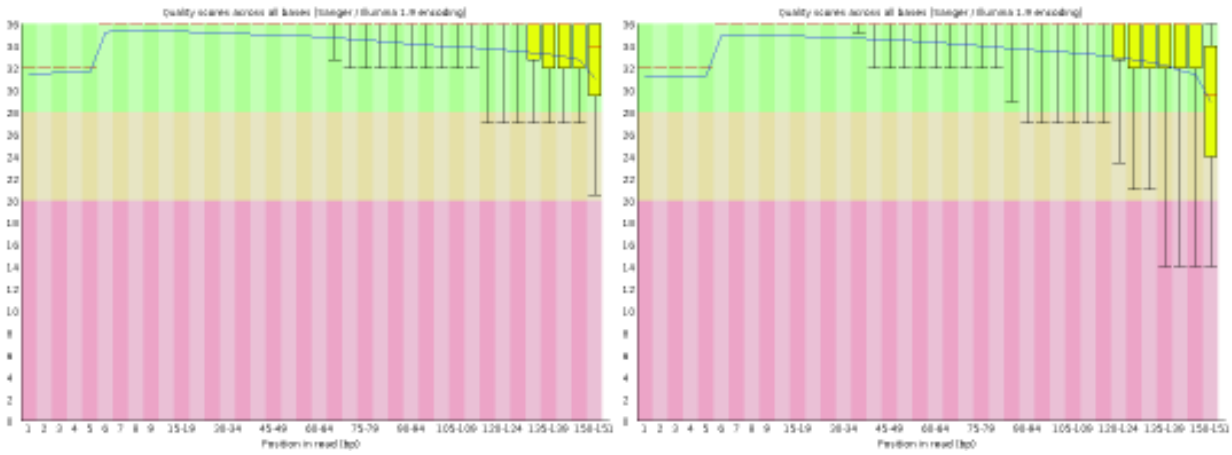
V.2.2. Filtrado de calidad

Después de realizar el análisis de calidad de las lecturas de secuenciación, se eligieron los filtros de trimmomatic correspondientes, debido a que no se encontraron aspectos a corregir solamente se utilizó un filtrado suave por ventana deslizante que eliminó regiones de las lecturas donde la calidad media en la ventana era menor de 17 (phred score). Se agregó un filtro para eliminar lecturas con un tamaño final menor a 50 bases y un filtro para cortar adaptadores de Illumina ya que el reporte de FastQC solo los reporta si se encuentran en más de un 1 % de las lecturas de secuenciación.

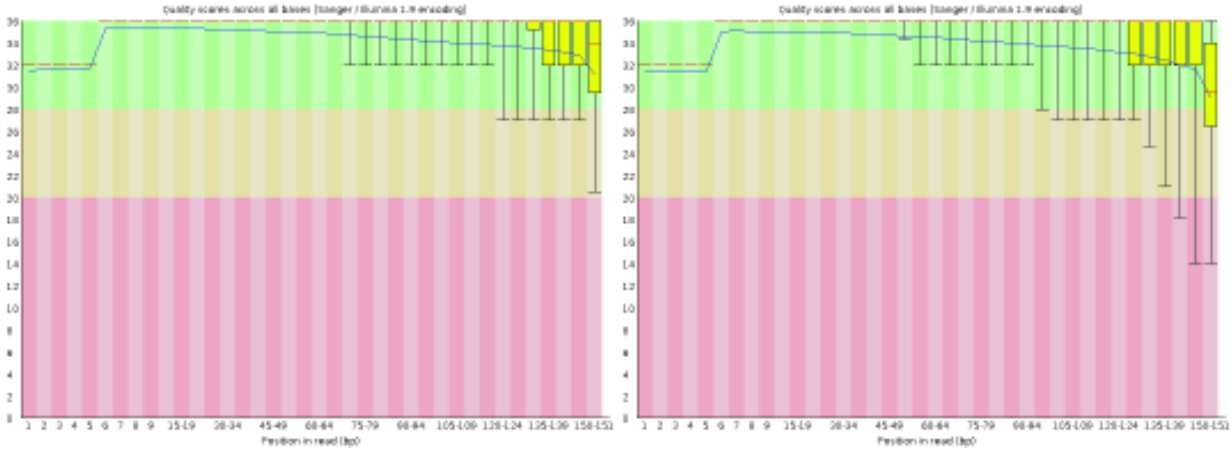
La tabla IV muestra las estadísticas de los resultados producidos por el filtrado de lecturas de secuenciación realizado con trimmomatic, se obtuvieron de 95.87 a 96.2 % de lecturas limpias pareadas por librería, se descartaron entre 0.53 y 0.64 % de lecturas por librería debido a su baja calidad o tamaño corto, entre 2.55 y 2.86 % de lecturas resultaron huérfanos de su pareja reverse y entre 0.67 y 0.8 % de su pareja *forward*.

Tabla IV. Estadísticas del resultado del filtrado de calidad de las lecturas de secuenciación de Chiltepín realizado con trimmomatic.

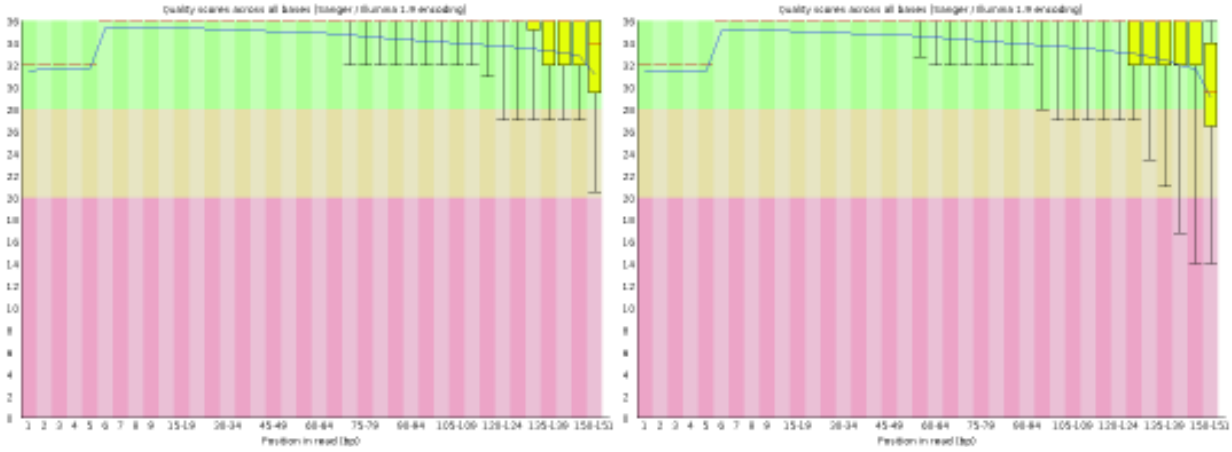
			Porcentaje de lecturas filtradas			
Librería	Descripción	No. de lecturas	Pareadas	en *forward*	en *reverse*	Eliminadas
1	CH 20 DDA	38,339,701	95.87 %	2.82 %	0.69 %	0.62 %
2	CH 20 DDA	34,461,667	96.20 %	2.55 %	0.67 %	0.59 %
3	CH 68 DDA	18,760,463	96.07 %	2.59 %	0.80 %	0.53 %
4	CH 68 DDA	38,778,252	95.72 %	2.86 %	0.78 %	0.64 %



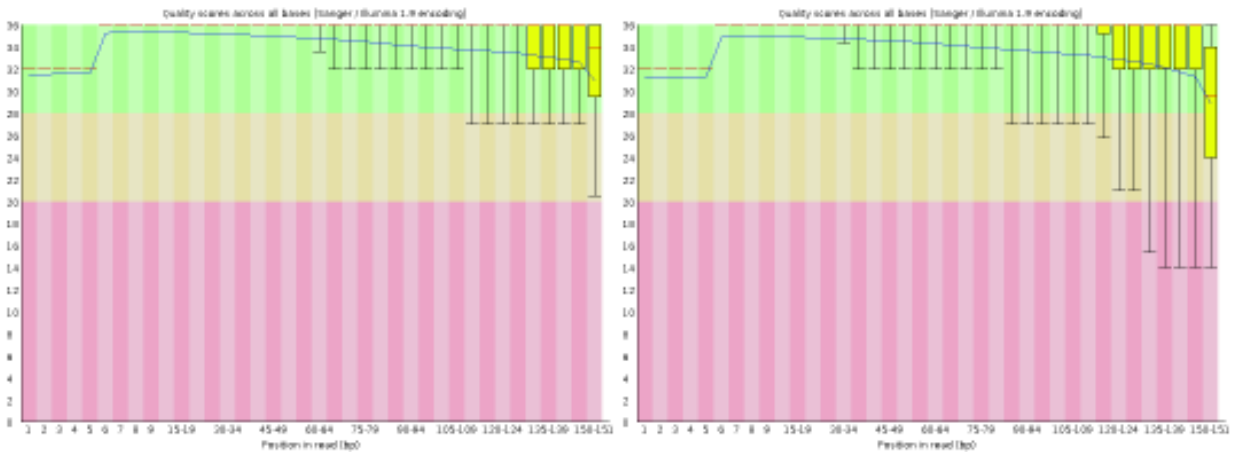
(a)



(b)



(c)



(d)

Figura 6. Calidad por base de las lecturas de RNA-Seq en bruto de Chiltepín: a) frutos recolectados a 20 DDA réplica 1, b) frutos recolectados a 20 DDA réplica 2, c) frutos recolectados a 68 DDA réplica 1, d) frutos recolectados a 68 DDA réplica 2. En todos los casos, la gráfica izquierda muestra el resultado de FastQC para el archivo de lecturas forward en bruto y la gráfica derecha el resultado de FastQC para el archivo de lecturas reverse en bruto.

V.2.3. Análisis de calidad post-filtrado

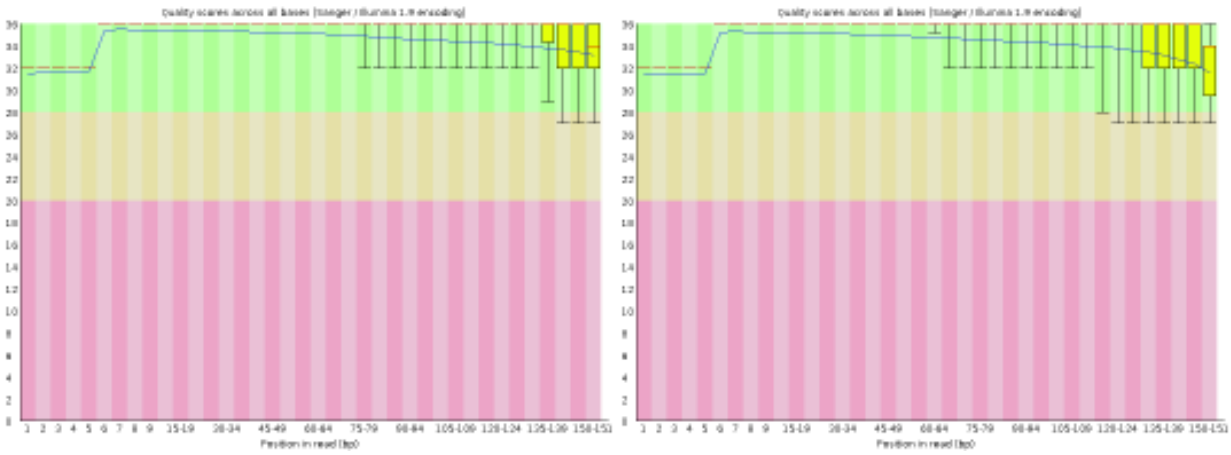
Posterior al filtrado de calidad se utilizó nuevamente la herramienta FastQC verificar la calidad de las lecturas filtradas, la tabla V muestra las estadísticas básica de los cuatro pares de archivos de lecturas de secuenciación generadas por FastQC, se obtuvieron un total de 250,094,380 lecturas de secuenciación pareadas limpias (84G de datos), lo cual representa un 95.93 % de las lecturas de secuenciación en bruto, el porcentaje de GC ahora se encuentra entre 41 y 42 y el intervalo de lecturas de secuenciación cambió, siendo 50 bases el tamaño mínimo de lectura de secuenciación actual.

La figura 7 muestra la gráfica de calidad por base de secuencia generado por FastQC para las lecturas de secuenciación previamente filtradas, nuevamente la calidad media por base se encuentra por encima del phred-score 31 en todos los casos; sin embargo, los intervalos de calidad de las posiciones en el extremo 3' aumentaron y ahora se encuentran por encima del valor de phred 26, dentro del intervalo considerado adecuado para un análisis bioinformático

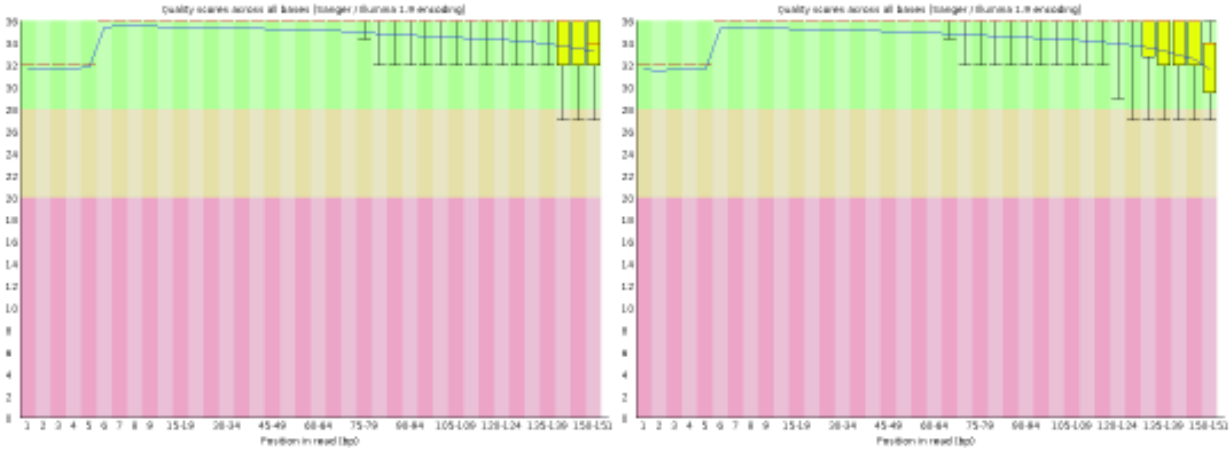
Tabla V. Resumen del resultado de análisis de calidad de lecturas de secuenciación de chiltepín filtradas, realizado con FastQC; cantidad de lecturas de secuenciación por cada uno de los archivos de lecturas por muestra (forward y reverse), porcentaje de GC e intervalo de tamaño de lecturas de secuenciación.

Librería	Descripción	Orientación	No. de lecturas	GC (%)	tamaño (pb)
1	CH 20 DDA, réplica 1	Forward	36,755,068	42	50-151
1	CH 20 DDA, réplica 1	Reverse	36,755,068	42	50-151
2	CH 20 DDA, réplica 2	Forward	33,150,884	42	50-151
2	CH 20 DDA, réplica 2	Reverse	33,150,884	42	50-151
3	CH 68 DDA, réplica 1	Forward	18,023,217	41	50-151
3	CH 68 DDA, réplica 1	Reverse	18,023,217	41	50-151
4	CH 68 DDA, réplica 2	Forward	37,118,021	42	50-151
4	CH 68 DDA, réplica 2	Reverse	37,118,021	42	50-151
		Total:	250,094,380		

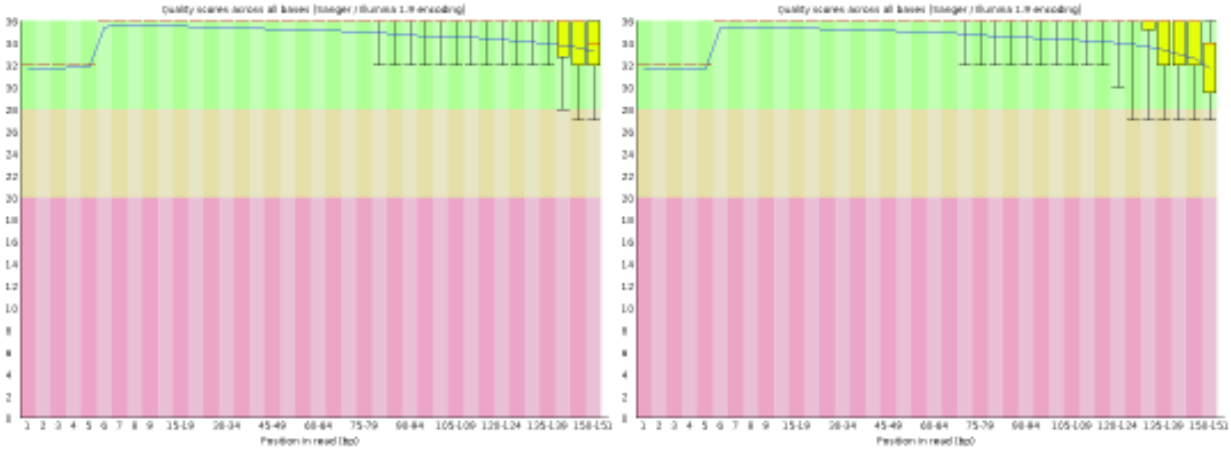
subsecuente.



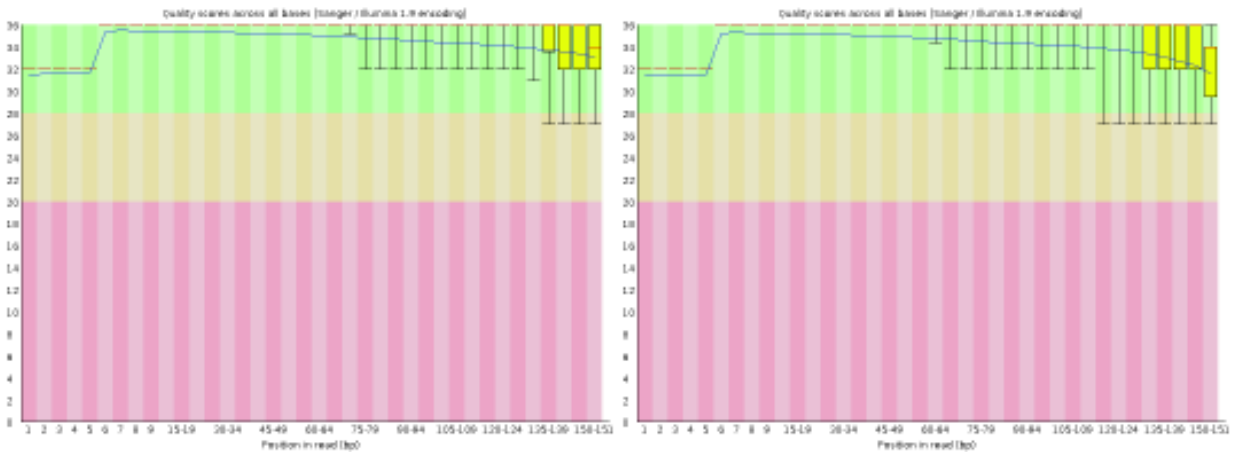
(a)



(b)



(c)



(d)

Figura 7. Calidad por base de las lecturas de RNA-Seq de Chiltepín filtradas: a) frutos recolectados a 20 DDA réplica 1, b) frutos recolectados a 20 DDA réplica 2, c) frutos recolectados a 68 DDA réplica 1, d) frutos recolectados a 68 DDA réplica 2. En todos los casos, la gráfica izquierda muestra el resultado de FastQC para el archivo de lecturas forward en bruto y la gráfica derecha el resultado de FastQC para el archivo de lecturas reverse.

V.3. Alineamiento al genoma de referencia

Una vez realizado el proceso de filtrado y habiendo obtenido secuencias libres de adaptadores y de regiones de baja calidad, se procedió a alinearlas al genoma de referencia de *Capsicum annuum* (cv. 'Criollo de Morelos 334') utilizando la herramienta Hisat2. La tabla VI muestra las estadísticas de alineamiento obtenidas, se obtuvo un promedio de 83.79 % (80.94 % - 86.63 %) de alineamiento pareados concordantes (ambas lecturas a distancia razonable y con orientación correcta) a un solo sitio del genoma, un promedio de 3.61 % (2.85 %-5 %) de alineamientos pareados concordantes a más de un sitio del genoma, entre 2.24 % y 4.92 % (media 3.77 %) alineamientos pareados discordantes (ambas lecturas alineadas pero a distancia no razonable o en el mismo sentido) a un solo sitio del genoma y un promedio de alineamiento general (incluyendo alineamiento de lecturas individuales) de 91.91 %.

Tabla VI. Estadísticas de alineamiento de las lecturas de secuenciación de Chiltepín al genoma de referencia de *Capsicum annuum* (cv. 'Criollo de Morelos 334').

		Alineamientos pareados			
Librería	Descripción	Concordante a 1 sitio	Concordante a >1 sitio	Discordante	General
1	CH 20 DDA, réplica 1	83.57 %	3.39 %	3.87 %	91.57 %
2	CH 20 DDA, réplica 1	80.94 %	5.00 %	2.24 %	89.48 %
3	CH 68 DDA, réplica 1	86.63 %	2.85 %	4.08 %	93.10 %
4	CH 68 DDA, réplica 1	86.29 %	3.21 %	4.92 %	93.49 %

V.4. Estimación de abundancias

Tras haber realizado el alineamiento de las lecturas de secuenciación de Chiltepín al genoma de referencia, se utilizaron los archivos sam con los resultados de los alineamientos para calcular la estimación de abundancias con la herramienta HTSeq en modo unión. El resultado es una matriz de abundancias con genes en las filas y librerías en las columnas (la tabla VII muestra las primeras 10 filas de la matriz). Los datos de chile Serrano se obtuvieron en este formato y se fusionaron con los datos de Chiltepín para producir una matriz de abundancias única.

Tabla VII. Primeras 10 líneas de la matriz de abundancia de genes obtenido por el alineamiento de lecturas de secuenciación de muestras de Chiltepín al genoma de referencia de *Capsicum annuum*, calculadas con la herramienta HTSeq en modo 'union'.

gene ID	Ch20dda r1	Ch 20dda r2	Ch68dda r1	Ch68dda r2
CA.PGAv.1.6.scaffold1.1	0	0	0	0
CA.PGAv.1.6.scaffold1.2	12	48	14	14
CA.PGAv.1.6.scaffold1.3	0	0	0	0
CA.PGAv.1.6.scaffold1.4	0	0	0	0
CA.PGAv.1.6.scaffold1.5	0	0	0	0
CA.PGAv.1.6.scaffold1.6	0	0	0	0
CA.PGAv.1.6.scaffold1.7	1	11	2	1
CA.PGAv.1.6.scaffold1.8	0	0	0	0
CA.PGAv.1.6.scaffold1.9	0	0	0	0

V.5. Anotación génica

V.5.1. Anotación Gene Ontology

Se utilizó la anotación automática de los genes de *Capsicum annuum* utilizando la herramienta Blast2GO, la cual consiste en 5 pasos: 1) una búsqueda con BLAST a las bases de datos elegida (en este caso las bases de datos curadas de Swissprot/Uniprot) para identificar secuencias similares a la de nuestro interés, 2) el mapeo de los genes homólogos identificados con BLAST con el fin de obtener los identificadores GO correspondientes, 3) el siguiente paso es la anotación y consiste en la selección de los términos GO del conjunto obtenido en el paso anterior, de modo que se obtenga la anotación más específica con el mayor nivel de confianza; 4) posteriormente el análisis estadístico permite realizar estudios de enriquecimiento GO entre un par de juegos de datos, 5) el último paso consiste en la visualización de los resultados a través de gráficas y tablas de datos.

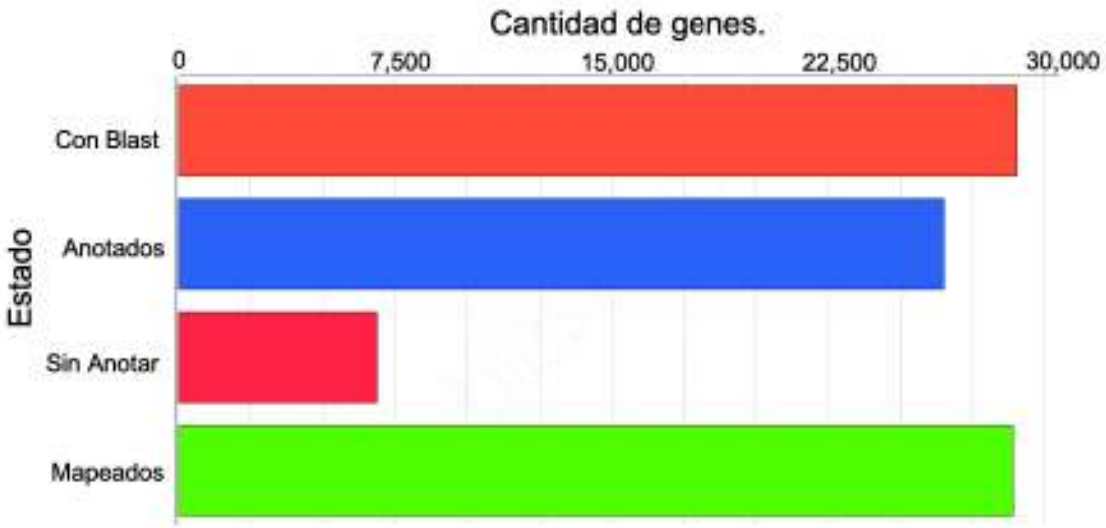


Figura 8. Estadística de genes de *Capsicum annuum* L. cv. 'Criollo de Morelos' anotados por BLAST2GO, en cada una de sus etapas, en el eje X se muestran la cantidad de genes y en el eje Y la etapa de anotación.

La Figura 8 muestra los resultados obtenidos por Blast2GO, de un total de 35,885 genes, 6,856 (19.10 %) no encontraron un alineamiento con BLAST; 29,028 (80.89 %) encontraron un

homólogo en la base de datos de referencia, 28,915 (80.57 %) obtuvo identificadores GO del homólogo encontrado y 26,511 (73.87 %) obtuvieron una anotación funcional estadísticamente significativa, la distribución de E-valores (apéndice B.38) mostró que alrededor de un 75 % de las secuencias anotadas tuvieron alineamientos con E-valores $\leq 6E^{-33}$.

La Figura 9 muestra información sobre los organismos utilizados para realizar la anotación de los genes de *Capsicum annuum*, 460 organismos fueron utilizados, de estos, seis especies contribuyeron con un 83.9 % de los alineamientos utilizados en la anotación, empezando con *Arabidopsis thaliana* que contribuyó con 18,416 (58.15 %) de los alineamientos, seguida por *Oryza sativa subsp. japonica, Nicotiana tabacum, Solanum lycopersicum, Solanum demissum, Solanum tuberosum* con 1,248, 728, 578 y 441 alineamientos, respectivamente. La categoría 'no identificado' corresponde a alineamientos a secuencias cuya especie no fue indicada en la base de datos, mientras que la categoría 'otros' contiene el número de alineamientos acumulado del resto (454) de los organismos utilizados para llevar a cabo la anotación.

V.5.2. Anotación KEGG

La anotación de genes a través de KEGG (Kyoto Encyclopedia of Genes and Genomes) permite asociar los genes expresados con rutas metabólicas, facilitando el entendimiento de su función, existen varios enfoques para realizar este procedimiento, el más común consiste en el uso de herramientas provistas por el servidor de KEGG, con ellas se puede realizar la anotación KO (asignación de identificador K) basadas en búsquedas con BLAST, GHOSTX o HMM, otra opción consiste en obtener identificadores de genes de un organismo anotado en la base de datos de KEGG por homología, es decir, realizar búsquedas con BLAST de los genes de interés en una base de datos local de un organismo anotado en KEGG. En este caso, utilizamos la segunda opción, comparamos e identificamos genes ortólogos con otro genoma de chile (cv Zunla-1) que se encuentra registrado en la base de datos de KEGG, utilizamos la herramienta Blastp e

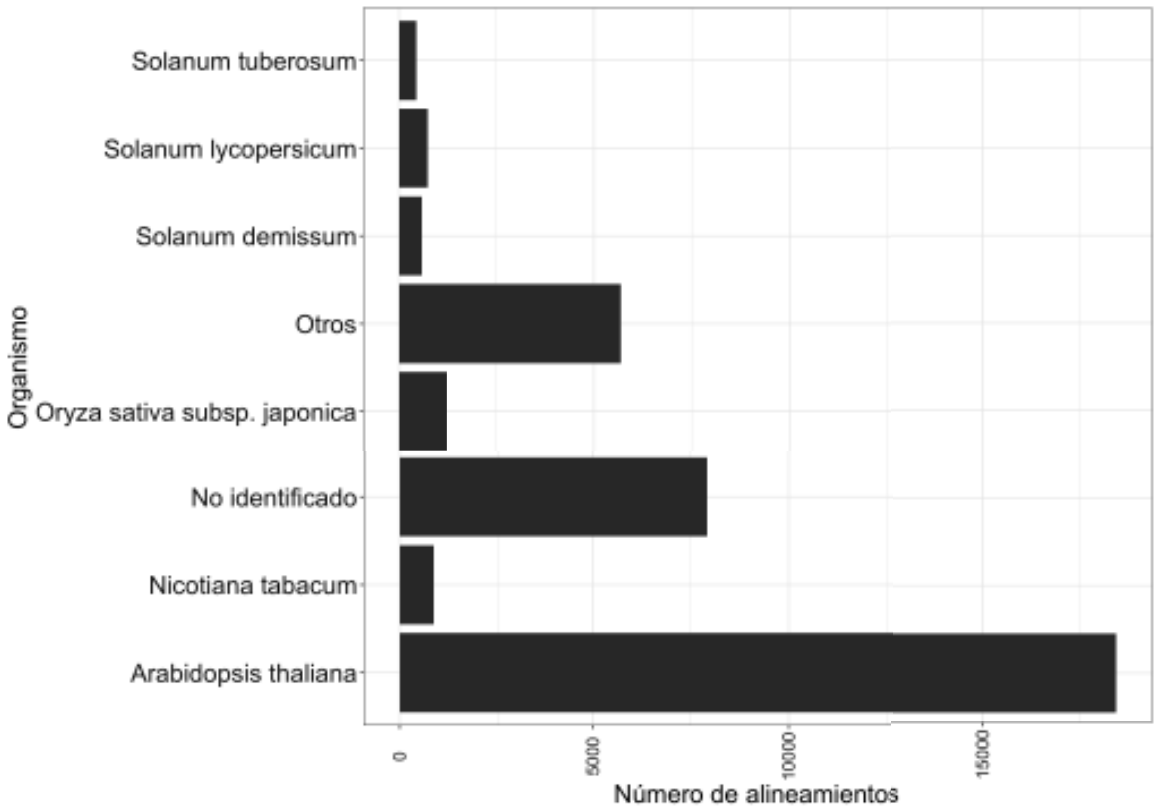


Figura 9. Organismos utilizados para realizar la anotación de los genes de chile con BLAST2GO, en el eje Y se encuentra el nombre del organismos y en el eje X el número de alineamientos con genes de chile.

identificamos un total de 35,093 ortólogos con un bitscore ≥ 50.

V.5.3. Identificación de genes homólogos con tomate

Tomate (*Solanum lycopersicum*) y chile son parientes cercanos; sin embargo, el tamaño del genoma de chile es tres veces más grande que el de tomate, por lo cual esperábamos encontrar múltiples copias de homólogos en el genoma de chile, tal como ha sido mostrado en estudios anteriores (Dubey *et al.*, 2019), por lo cual, utilizamos el software Inparanoid que es capaz de identificar grupos de ortólogos y parálogos en dos genomas. Se encontraron 16,671 grupos de ortólogos, 18,797 parálogos en chile y 18,109 parálogos en tomate.

V.6. Análisis de expresión génica en Chiltepín y Serrano

La matriz de abundancias única fue analizada en la herramienta estadística R, se descartaron genes con expresión nula y genes con una media del logaritmo de expresión por millón menor a 1 (Ave log count per million < 1) en todas las librerías. Posteriormente se calcularon estadísticas básicas sobre los genes expresados, se identificaron 19,811 genes expresados en al menos una muestra de Chiltepín o Serrano, el número de genes expresados por condición fue similar en todos los casos; de 18,590 genes expresados en Serrano a 60 DDA a 19,525 genes expresados en Chiltepín a 20 DDA. La Figura 10 muestra un diagrama de Venn con los genes expresados en cada condición y sus correspondientes interacciones, se encontraron 18,067 (91.2 %) genes expresados en todas las muestras, 443 genes (2.24 %) genes se encontraron expresados solamente a 20 DDA (en Chiltepín, Serrano o ambos) de los cuales 6 se identificaron en Serrano solamente y 31 en Chiltepín. Se hallaron 43 genes (0.22 %) expresados solamente en el estado maduro (en Chiltepín, Serrano o ambos), 2 de ellos se identificaron solamente en Serrano y 12 en Chiltepín. El número de genes expresados fue mayor tanto en Serrano como en Chiltepín a 20 DDA, estudios previos han reportado resultados similares y sugieren que podría deberse a que

los genes comienzan a 'apagarse' conforme los frutos maduros pasan a la etapa de senescencia (Martínez-López *et al.*, 2014).

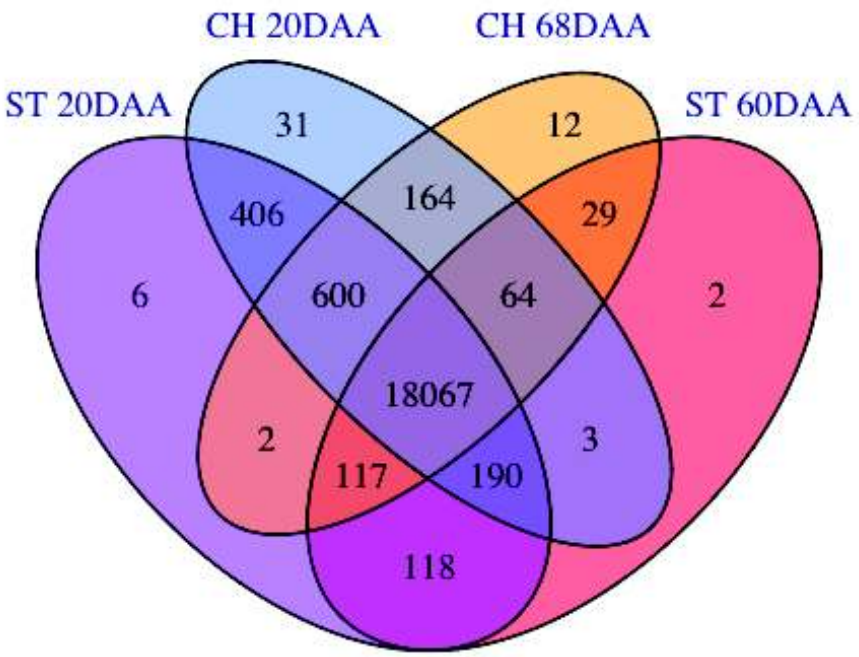
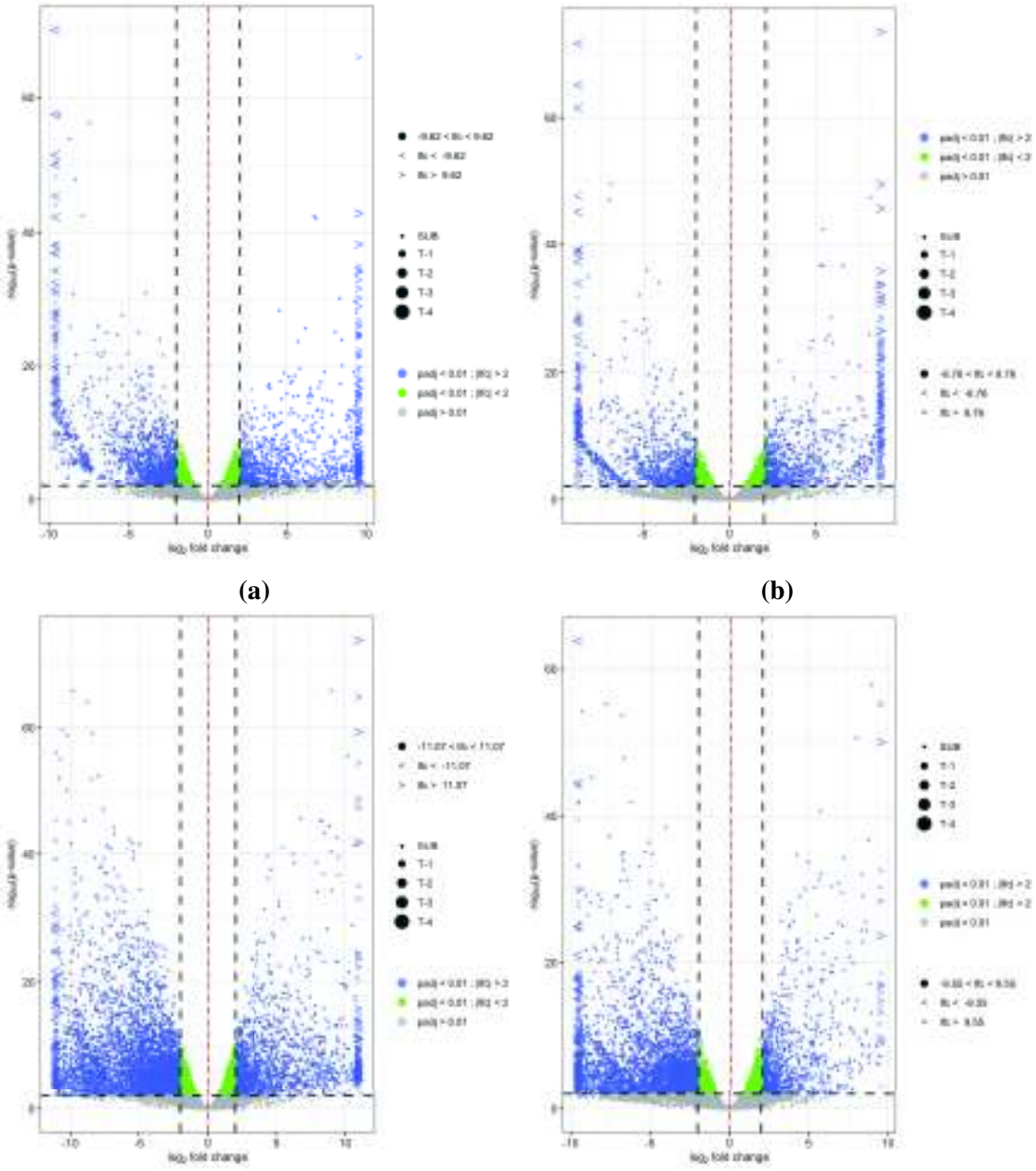


Figura 10. Diagrama de Venn del número de genes expresados en cada etapa de desarrollo de fruto de chiltepín (CH) y Serrano (ST), los números en cada intersección representan el número de genes compartidos en el conjunto de muestras corrspondientes a la intersección.

V.6.1. Genes expresados diferencialmente

Posteriormente se procedió a realizar el análisis de expresión diferencial con el paquete EdgeR de R, se utilizaron los contrastes definidos en la tabla I del capítulo de Materiales y métodos.

La figura 12 muestra un diagrama Venn de los genes expresados diferencialmente (DEGs) identificados en cada uno de los contrastes y sus respectivas intersecciones (DEGs identificados en más de un contraste). Se identificaron un total de 1,008 genes expresados diferencialmente en al menos una de las comparaciones, el número de DEGs por contraste osciló entre 43 en el



(a) **(b)**

(c) **(d)**

Figura 11. Gráficas volcano de los genes expresados en 4 contrastes: a) Chiltepín 20 DDA contra Serrano 20 DDA, b) Chiltepín 68 DDA contra Serrano 60 DDA, c) Chiltepín 20 DDA contra Chiltepín 68 DDA, d) Serrano 20 DDA contra Serrano 60 DDA. El eje X representa el logaritmo base 2 de la razón de cambio (log2fold change), el eje Y representa el logaritmo negativo del p valor de las comparaciones entre niveles de expresión de los genes en los grupos experimentales contrastados, puntos en gris representan genes no significativos estadísticamente, puntos en verde indican genes estadísticamente significativos pero con valor log2fold <2, puntos azules representan genes expresados diferencialmente (log2fodl >|2| y FDR <0.01).

contraste entre las dos variedades a 20 DDA (Ch20-St20) y 882 DEGs en la comparación de Chiltepín a 20 DDA contra 68 DDA (Ch20-Ch68). No se encontraron genes expresados diferencialmente comunes a todos los contrastes; sin embargo, 278 DEGs fueron identificados en más de un contraste y 730 DEGs se encontraron exclusivamente en un contraste, la mayoría de los genes expresados diferencialmente exclusivos a un contraste (639, 63.39 %) se encontraron en la comparación entre 20 y 68DDA de Chiltepín, seguido de 72 (7.14 %) DEGs encontrados solamente en la comparación entre ambas etapas de maduración de Serrano (St20-St60), 13 (1.29 %) genes expresados diferencialmente exclusivamente en el contraste de fruto maduro entre ambas variedades (Ch68-St60), y 6 (0.6 %) DEGs identificados solamente en la comparación de fruto verde de ambas variedades (Ch20-St20).

En un estudio previo de Martínez-López *et al.* (2014) mostraron que la mayoría de los genes expresados diferencialmente en etapas de maduración de fruto se encontraban reprimidos, es decir, los genes se expresaban en las etapas de desarrollo de fruto y se 'apagaban' al alcanzar la etapa de fruto maduro y sugieren que podría deberse a la proximidad de la senescencia, de igual manera, en este estudio se encontró que la mayoría de los genes expresados diferencialmente al comparar las etapas de fruto maduro contra fruto verde en Chiltepín (Ch20-Ch68) y Serrano (St20-St60), se encuentran reprimidos. En el contraste Ch20-Ch68 se encontraron 639 (72.45 %) genes reprimidos y solamente 243 (27.55) DEGs inducidos; en el contraste St20-St60 se identificaron 214 (69.94 %) DEGs reprimidos y 92 (30.06 %) DEGs inducidos. Curiosamente este mismo comportamiento se repitió en los contrastes donde se comparan ambas variedades en la misma etapa de maduración, en Ch20-St20 se encontraron 27 (62.79 %) DEGs reprimidos y 16 (37.21 %) inducidos, mientras que en el contraste Ch68-St60 se encontraron 32 (54.24 %) DEGs reprimidos y 27 (45.76 %) DEGs inducidos (Figura 13).

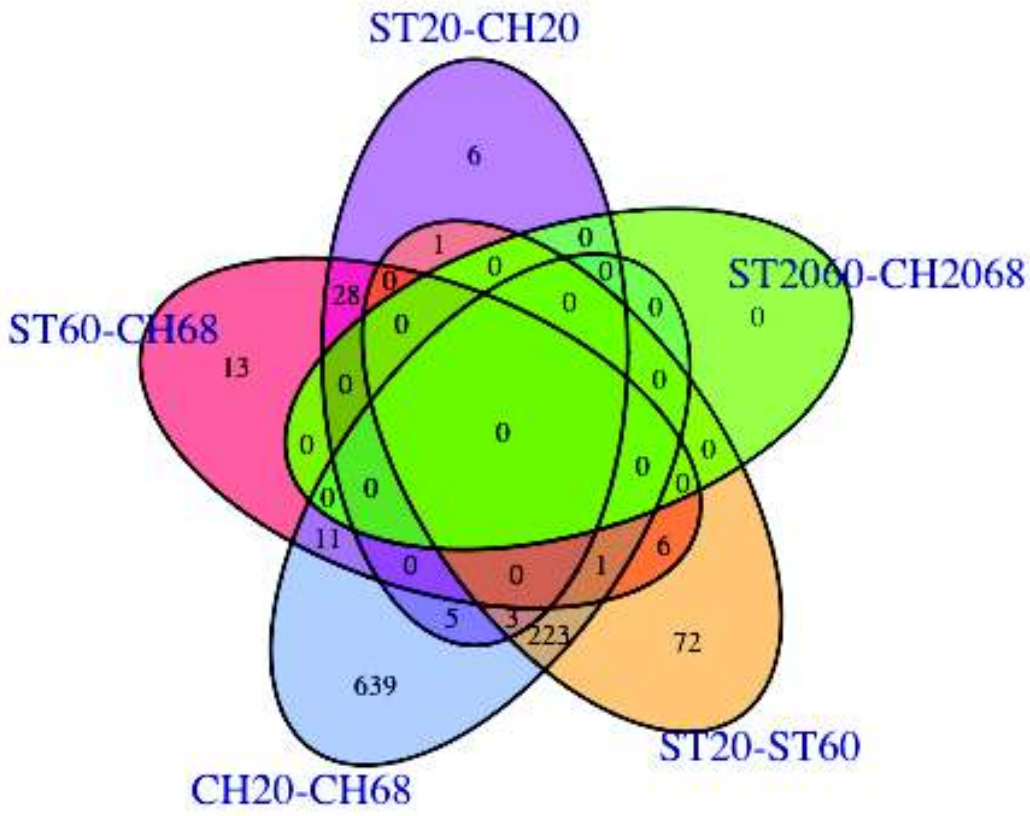


Figura 12. Diagrama de Venn del número de genes expresados diferencialmente (DEGs) por contraste en Serrano (ST) y Chiltepín (CH), los números en las intersecciones representan el total (genes inducidos y reprimidos) de genes expresados diferencialmente compartidos por los contrastes correspondientes a la intersección.

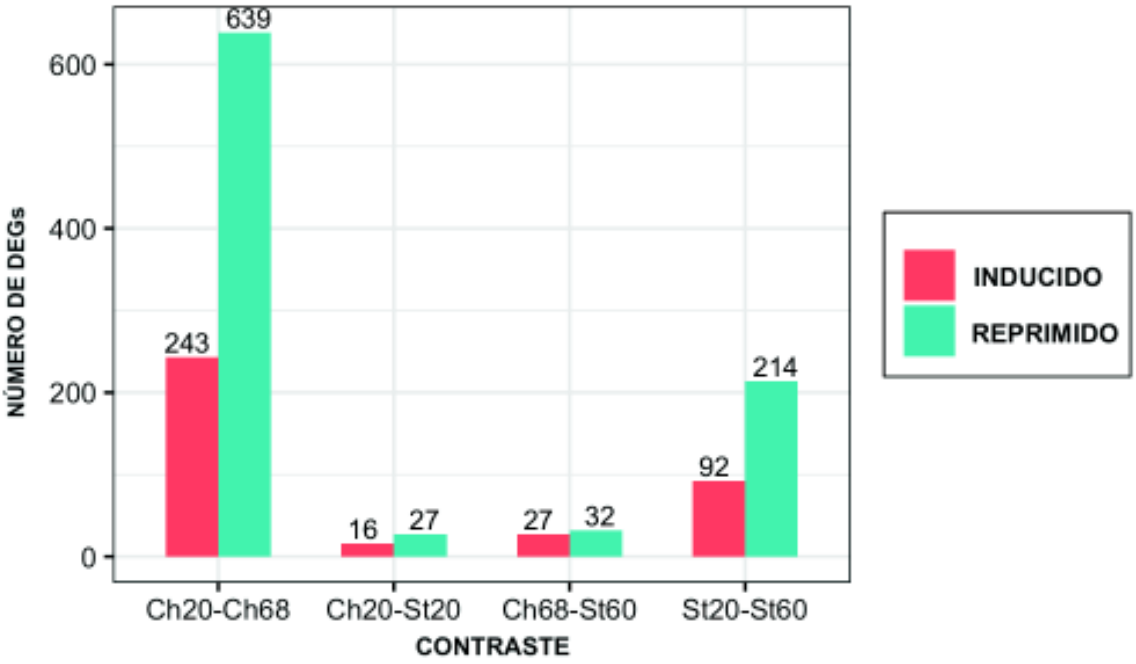


Figura 13. Cantidad de genes expresados diferencialmente en los cinco contrastes diseñados para las muestras de Chiltepín (Ch) y Serrano (St) a 20 y 68(60) DDA, en el eje X se muestran los contrastes, en el eje Y se muestra la cantidad de genes expresados diferencialmente y el color de las barras indica si son genes reprimidos o inducidos.

V.6.1.1. Análisis de enriquecimiento de términos de Gene Ontology

Después de realizar la asignación de términos GO al genoma de chile, se realizó la anotación de los genes presentes en el conjunto de datos de este experimento. En total se asignaron 235,293 términos de proceso biológico, 104,106 de componente celular y 87,767 de función molecular, de acuerdo a la base de datos slimGO.

El análisis de enriquecimiento de términos GO puede ayudar a determinar las funciones de los genes expresados diferencialmente, ya que permite visualizar a diferentes niveles jerárquicos, los procesos biológicos, componentes celulares y funciones moleculares posiblemente afectados por las diferencias de expresión en las condiciones contrastadas. Para realizar el análisis de enriquecimiento de términos GO se utilizó GOSeq, una herramienta que toma en cuenta el tamaño de los genes para evitar el sesgo por tamaño, común en experimentos de RNA-Seq, y compara los términos GO asignados al conjunto de genes expresados diferencialmente en un contraste con los asignados al conjunto de genes de fondo (todos los genes expresados en el contraste), para identificar categorías enriquecidas estadísticamente significativas.

La Figura 14 muestra los resultados del análisis de enriquecimiento de términos GO obtenido de los genes expresados diferencialmente en los 5 contrastes de Chiltepín y Serrano en dos etapas de maduración de fruto, los contrastes en los que se comparan ambas variedades en la misma etapa de maduración (Ch20-St20 y Ch68-St60) no presentaron términos GO enriquecidos estadísticamente significativos, probablemente debido a la cantidad reducida de genes expresados diferencialmente en dichos contrastes. La Figura 14,A muestra los términos GO enriquecidos en el contraste St20-St60 y 14,B muestra los términos GO enriquecidos en el contraste Ch20-Ch68, en ambos casos la mayoría de los términos enriquecidos corresponden a la ontología de Proceso Biológico con 7 y 17 términos, respectivamente; seguido de la ontología de Componente Celular con 2 y 4 términos, respectivamente y función molecular con 0 y 2 términos, respectivamente.

En ambos contrastes (Ch20-St20 y Ch68-St60), las categorías más representadas correspon-

den a términos de la ontología de Proceso biológico: respuesta a estrés con 159 y 420 genes, respectivamente; respuesta a químicos con 157 y 421 genes, respectivamente; respuesta a estímulos endógenos con 114 y 313 genes, respectivamente; respuesta a estímulo externo con 116 y 292 genes, respectivamente y respuesta a estímulo biótico con 101 y 257 genes, respectivamente. De forma similar, en la ontología de componente celular, existen dos términos GO enriquecidos en común en los dos contrastes: vacuola con 61 y 148 genes, respectivamente y pared celular con 117 y 41 genes, respectivamente.

Todos los términos GO enriquecidos en el contraste St20-St60 se encuentran también enriquecidos en el contraste Ch20-Ch68; sin embargo, este último contraste presentó términos GO enriquecidos exclusivos, en la ontología de proceso biológico: se encontraron 150 genes en la categoría de proceso celular, 121 genes relacionados con el proceso de metabolismo de lípidos, 108 genes relacionados con el desarrollo floral, 91 genes responsables del desarrollo celular, 89 genes pertenecientes al proceso metabólico de carbohidratos, 59 genes asociados con el crecimiento, 30 genes vinculados con diferentes aspectos de la fotosíntesis, 28 genes relacionados con el tropismo, 19 genes de abscisión y 11 genes asociados a la maduración de frutos. En la ontología de componente celular se encontraron los términos: membrana plasmática con 292 genes y región extracelular con 119 genes; mientras que en la ontología de función molecular se encontraron los términos de actividad catalítica con 146 genes y actividad de receptor de señales con 39 genes (Figura 14,B).

V.6.1.2. Análisis de enriquecimiento de rutas metabólicas (KEGG)

La Enciclopedia de Genes y Genomas de Kyoto (KEGG por sus siglas en inglés), es una base de datos de mapas de rutas metabólicas utilizada para obtener información de las redes metabólicas correspondientes al conjunto de genes de nuestro experimento.

Se utilizaron los homólogos con *C. annuum* cv. Zunla-1 para realizar la identificación de los

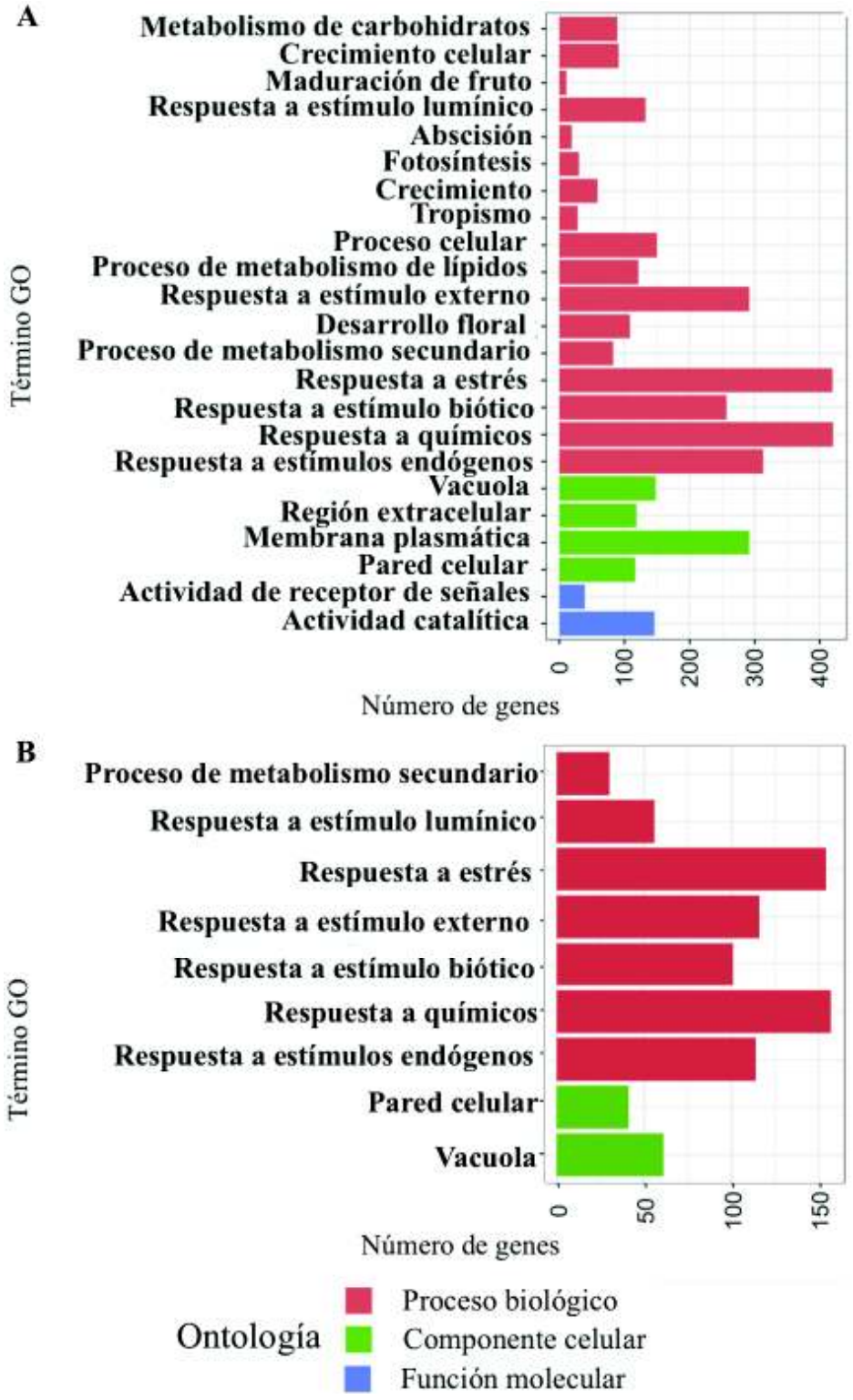


Figura 14. Análisis de enriquecimiento de términos de ontología génica (GO) de los genes expresados diferencialmente (DEGs), el eje X indica el número de DEGs correspondientes al término GO, el color de las barras representa la ontología a la que pertenece el correspondiente término GO. A) Términos GO enriquecidos en el contraste Ch20-Ch68 (Chiltepín 20 DDA vs Chiltepín 68 DDA), B) Términos GO enriquecidos en el contraste St20-St60 (Serrano 20 DDA vs Serrano 60 DDA).

genes expresados diferencialmente en la base de datos de mapas de rutas metabólicas de KEGG, del total de 18,067 19,811 genes expresados se identificaron 9,401 enzimas (52.03 %) pertenecientes a 135 rutas metabólicas. Posteriormente, se llevó a cabo un análisis de enriquecimiento de rutas metabólicas con el paquete de R ClusterProfiler, que implementa una distribución hipergeométrica para detectar rutas metabólicas enriquecidas por genes expresados diferencialmente, comparándolas contra el universo de genes expresados.

Al igual que en el análisis de enriquecimiento de términos GO, los contrastes donde se comparan ambas variedades en la misma etapa de maduración (Ch20-St20 y Ch68-St60) no presentaron rutas metabólicas enriquecidas, probablemente debido a la pequeña cantidad de genes expresados diferencialmente en cada una de ellas. La Figura 15 muestra las rutas metabólicas enriquecidas (EMP) estadísticamente significativas detectadas en: A) el contraste Ch20-Ch68 y B) St20-St60, las rutas metabólicas de *Biosíntesis de fenilpropanoides*, *Fotosíntesis - proteínas de antena* y *Biosíntesis de cutina, suberina y ceras* se encuentran enriquecidas en ambos contrastes; sin embargo, muestran diferencias de expresión en genes relacionados con las enzimas de dichas rutas metabólicas.

La ruta de *Biosíntesis de fenilpropanoides* es la EMP con el mayor número de genes expresados diferencialmente en el contraste Ch20-Ch68 con 21 DEGs (Figura 17 y el segundo en el contraste St20-St60 con 8 DEGs (Figura 16. Además de las diferencias en el número de genes expresados diferencialmente, en St20-St60 los 8 DEGs pertenecientes a la EMP se encuentran reprimidos, mientras que en el contraste Ch20-Ch68, 8 genes relacionados con las enzimas *cinnamyl-alcohol dehydrogenase* y *caffeoyl shikimate esterase* se encuentran inducidos.

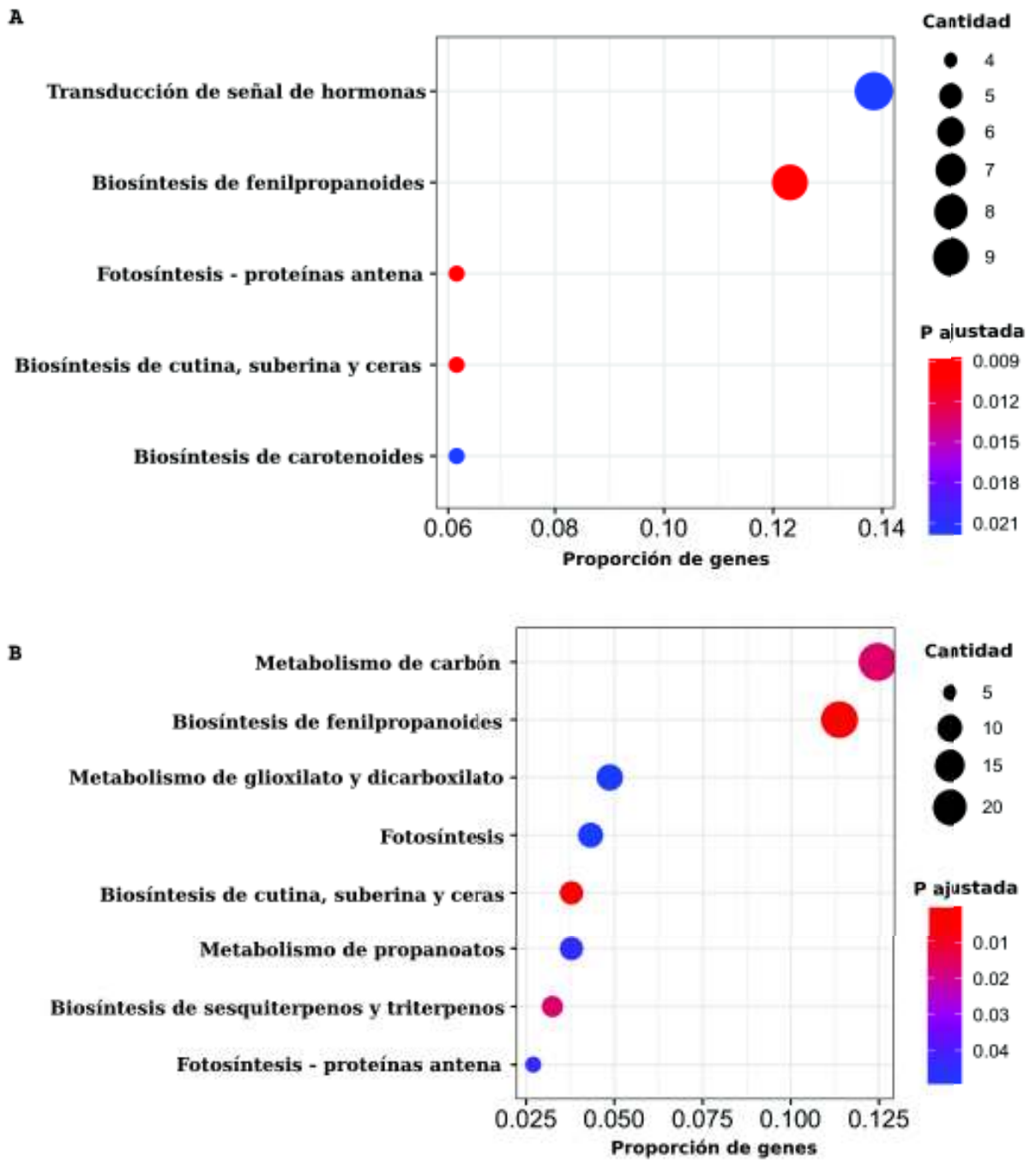


Figura 15. Análisis de enriquecimiento de rutas metabólicas (KEGG) de los genes expresados diferencialmente (DEGs), el eje X indica la proporción (el cociente del número de DEGs y el número de genes en el fondo), el tamaño de las burbujas representa el número de DEGs en la ruta metabólica correspondiente, el color de las burbujas indica el q valor del enriquecimiento de la ruta metabólica correspondiente. A) Rutas metabólicas enriquecidas en el contraste St20-St60 (Serrano 20 DDA vs Serrano 60 DDA), B) Rutas metabólicas enriquecidas en el contraste Ch20-Ch68 (Chiltepín 20 DDA vs Chiltepín 68 DDA).

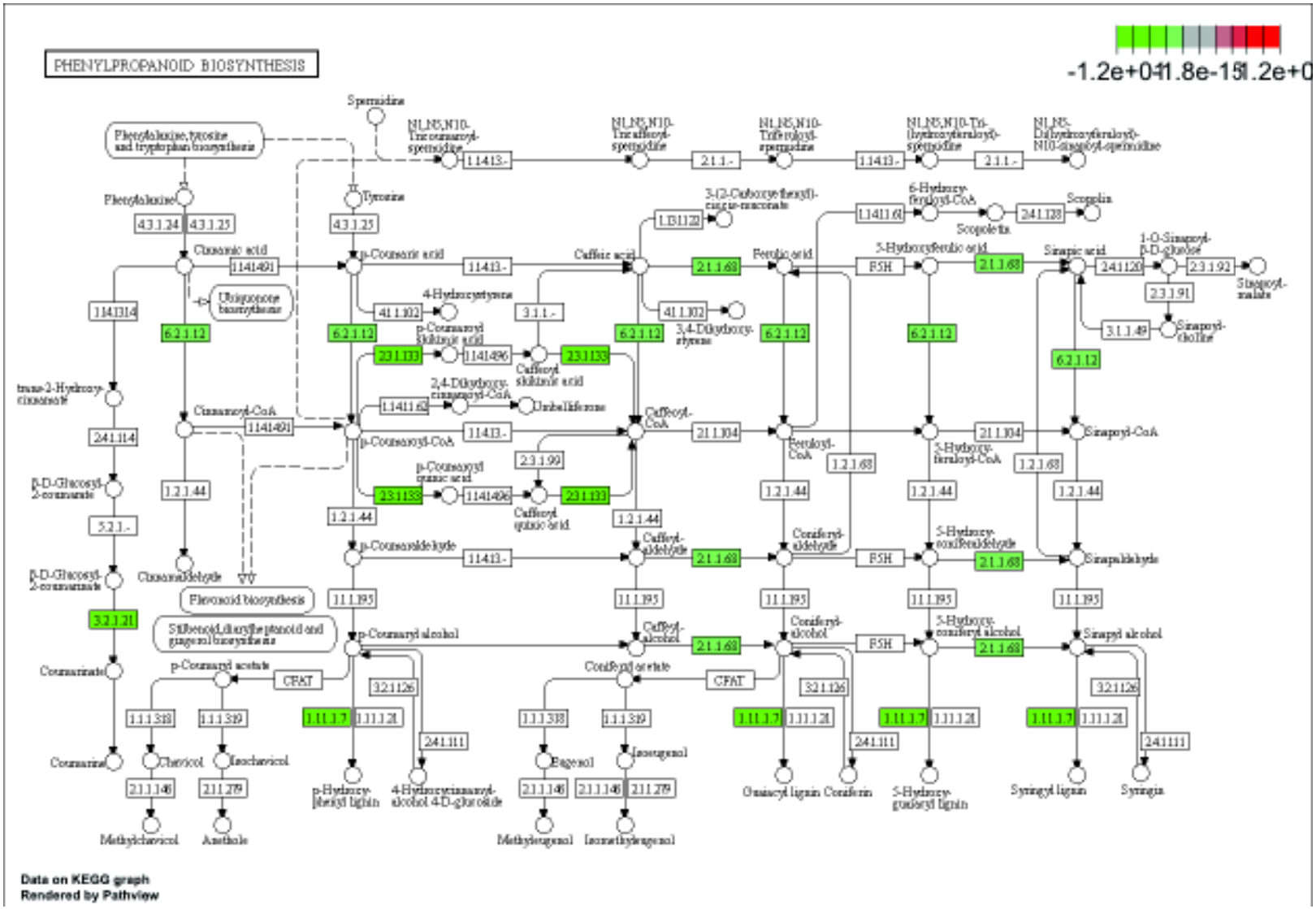


Figura 16. Ruta metabólica *Biosíntesis de fenilpropanoides* enriquecida en el contraste St20-St60, en verde enzimas cuyos genes se encuentran reprimidos, en rojo enzimas cuyos genes se encuentran inducidos en el contraste.

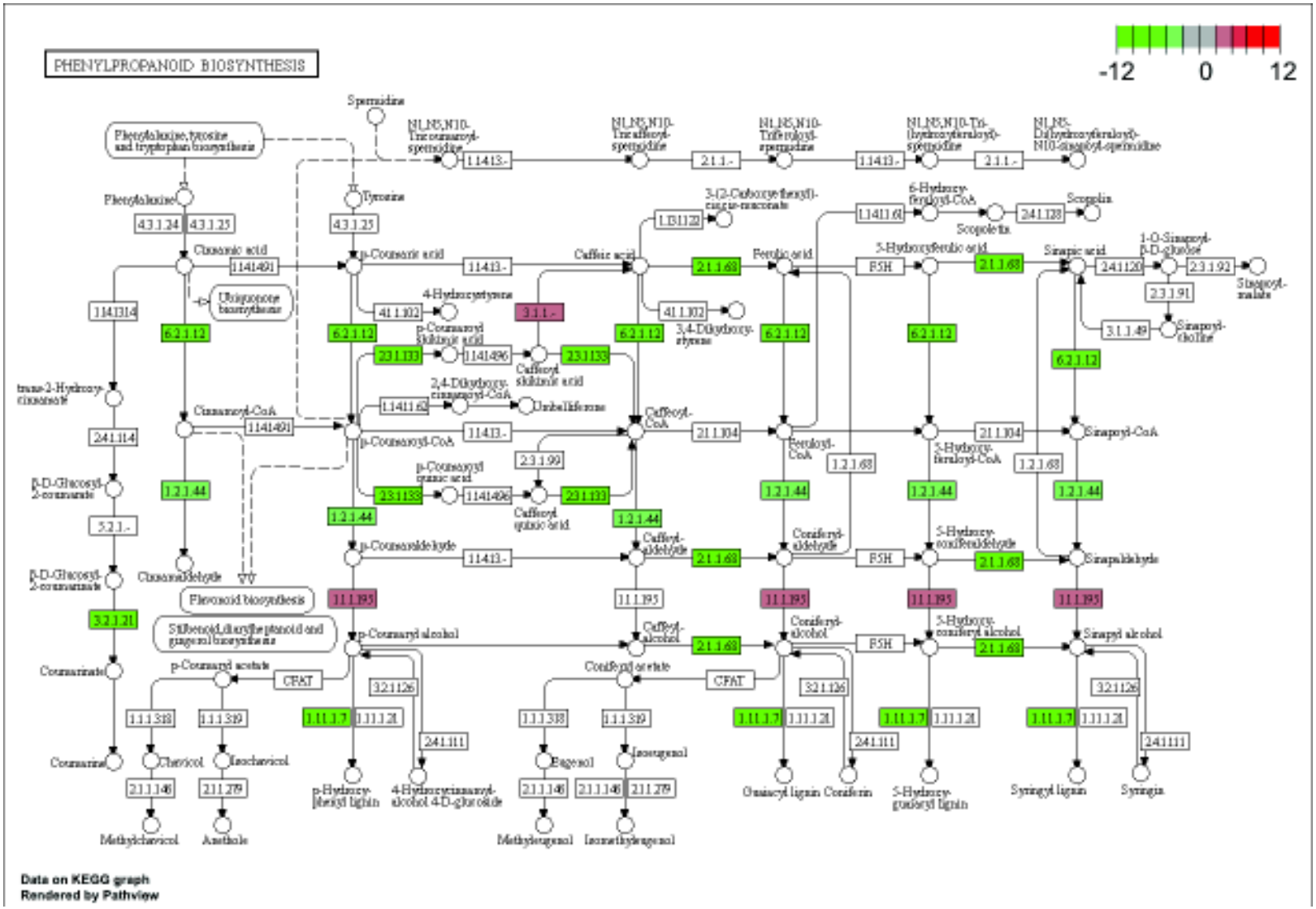


Figura 17. Ruta metabólica *Biosíntesis de fenilpropanoides* enriquecida en el contraste Ch20-Ch68, en verde enzimas cuyos genes se encuentran reprimidos, en rojo enzimas cuyos genes se encuentran inducidos en el contraste.

La segunda ruta metabólica enriquecida en ambos contraste es la de *Biosíntesis de cutina, suberina y ceras*, que presenta diferencias en los genes correspondientes a las enzimas *fatty acid omega-hydroxy dehydrogenase* inducidos solamente en el contraste Ch20-Ch68, *CYP94A5* que se encuentran inducidos solamente en el contraste St20-St60 y la enzima *CYP77A6* cuyos genes se encuentran reprimidos solamente en el contraste Ch20-Ch68; las tres enzimas se encuentran involucradas en la biosíntesis de ácidos grasos insaturados, un elemento crucial de los polímeros que componen la pared celular.

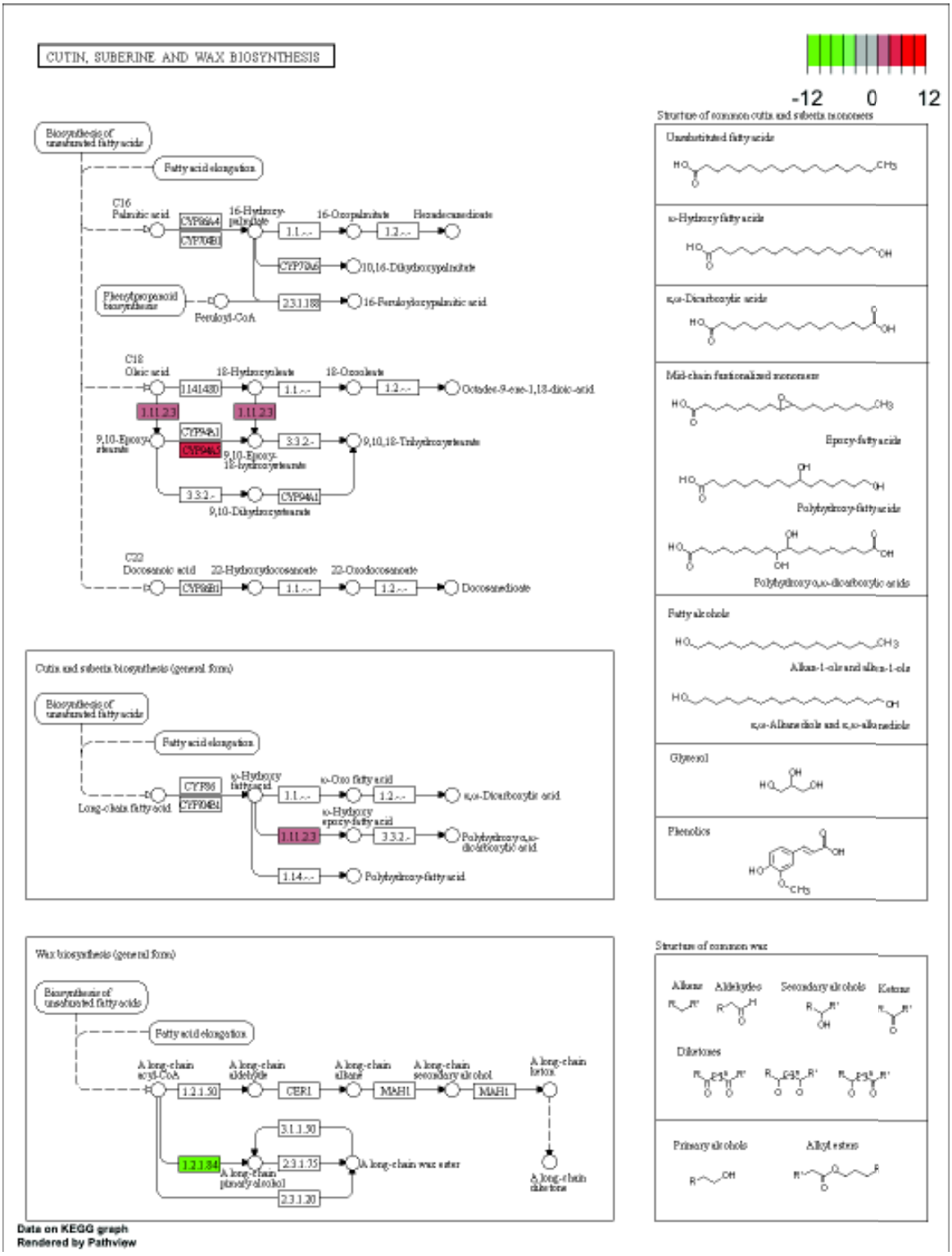
La tercera ruta metabólica enriquecida en ambos contrastes es la de *Fotosíntesis - proteínas antena*, en la cual se muestran diferencias solamente en el número de genes expresados diferencialmente en cada contraste; sin embargo, todos los genes correspondientes a las enzimas de la ruta metabólica se encuentran reprimidos en ambos contrastes; de forma consistente con la degradación de clorofila que sucede durante el proceso de maduración de fruto (Figura 19).

Se encontraron dos rutas metabólicas enriquecidas exclusivamente en el contraste St20-St60, la encargada de la *biosíntesis de carotenoides* y *Transducción de señales de hormonas*. La ruta metabólica de *biosíntesis de carotenoides* contiene 4 genes expresados diferencialmente, correspondientes a las enzimas: *15-cis-phytoene synthase* (inducido), *9-cis-epoxycarotenoid dioxygenase* (inducido), *xanthoxin dehydrogenase* (reprimido) encargada de la producción de fitoenos; y *9-cis-epoxycarotenoid dioxygenase* (reprimido), encargada de la biosíntesis de xantoxina y aldehído abscísico (Figura 20).

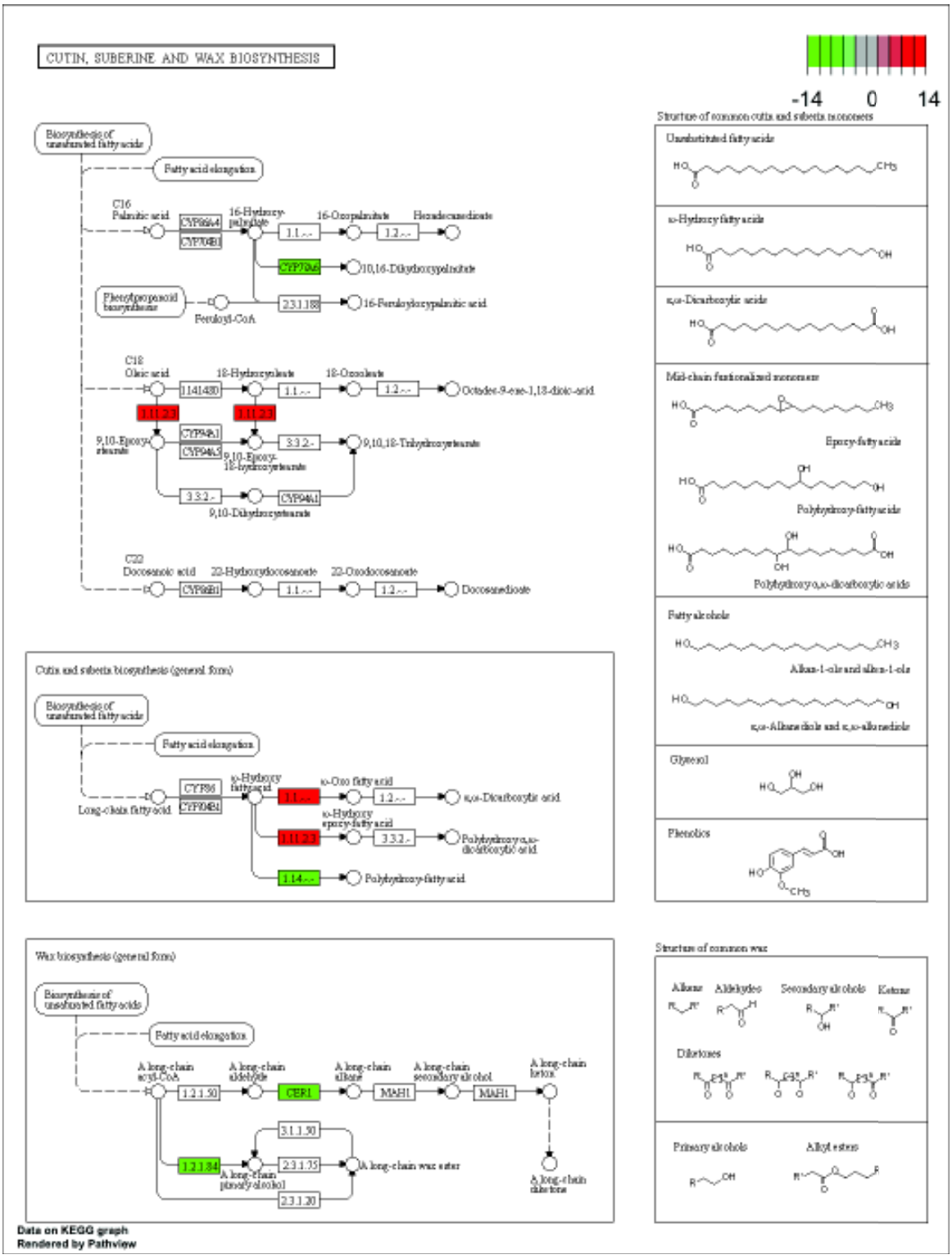
La ruta metabólica *Transducción de señales de hormonas* es la EMP con el mayor número de genes expresados diferencialmente en el contraste St20-St60 con 9 DEGs, entre los cuales se encuentran: PP2C (inducido) y SnRK2 (reprimido), componentes importantes en el señalamiento de ácido abscísico (ABA) relacionados con el cierre de estomas, dormancia de semillas y biosíntesis de carotenoides; el gen ETRF1/2 (inducido), parte de la señalización por etileno (ET) uno de los principales controladores del proceso de maduración de fruto y senescencia;

por último, el gen AUXIAA (reprimido) y SAUR (inducido) parte de la ruta de señalización relacionada con la elongación celular y crecimiento de plantas (Figura 21).

Por otro lado, se identificaron 5 rutas metabólicas enriquecidas exclusivamente en el contraste Ch20-Ch68 que incluye la *Biosíntesis de sesquiterpenos y triterpenos* que incluye 6 genes expresados diferencialmente y que codifican a las enzimas *germacrene D synthase* (2 genes, reprimido e inducido), *alpha-farnesene synthase* (inducido), *squalene monooxygenase*(reprimido), *(3S,6E)-nerolidol synthase* (reprimido),*beta-amyrin synthase* (reprimido); responsables de la producción de compuestos germacreno D, α farneseno, estructura de terpenos, nerolidol y β amirina, respectivamente y que son utilizados como defensa ante diferentes tipos de estrés (Figura 22. Otras rutas metabólicas enriquecidas en este contraste son el *Metabolismo de carbono* que contiene 23 genes expresados diferencialmente, *Metabolismo de propanoatos* con 7 genes expresados diferencialmente, *Metabolismo de glioxilato y dicarboxilato* con 9 DEGs y la ruta metabólica de *Fotosíntesis* que contiene 9 genes expresados diferencialmente.

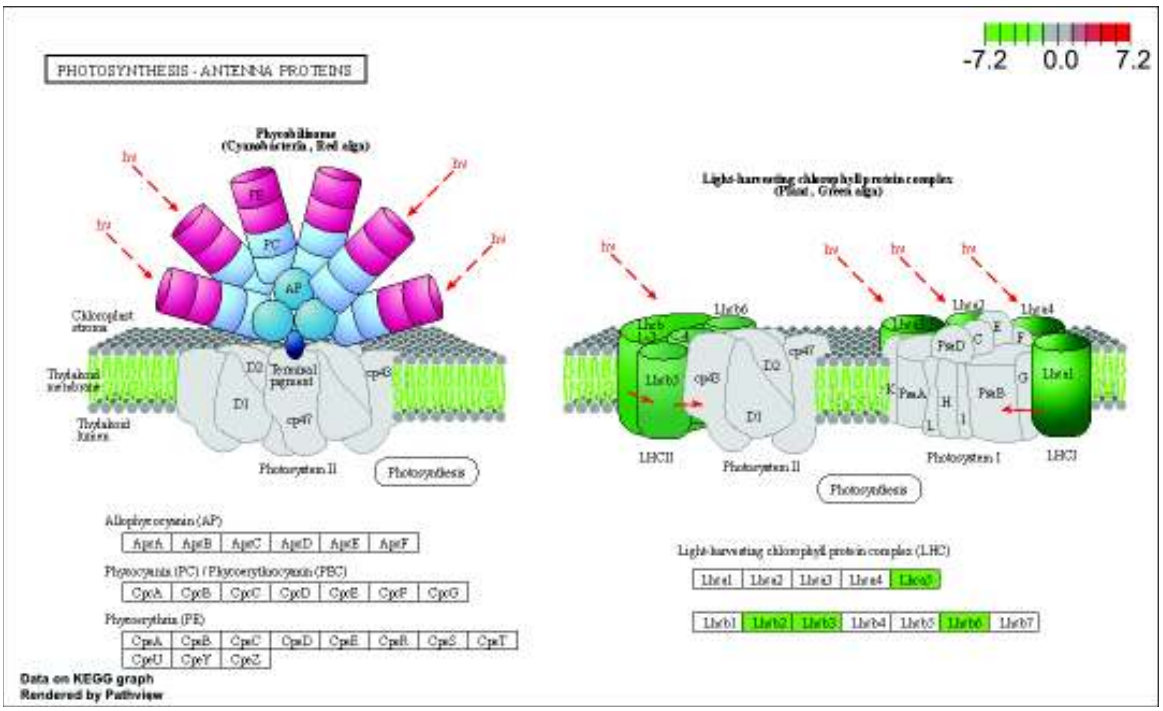

CUTIN, SUBERINE AND WAX BIOSYNTHESIS
-12
0
12
Structure of common cutin and suberin monomers
Unsubstituted fatty acids
ω-Hydroxy fatty acids
α,ω-Dicarboxylic acids
Mid-chain functionalized monomers
Epoxy-fatty acids
Polyhydroxy-fatty acids
Polyhydroxy α,ω-dicarboxylic acids
Fatty alcohols
Alkan-1-ols and alken-1-ols
α,ω-Alkanediols and α,ω-alkenediols
Glycerol
Phenolics
Biosynthesis of unsaturated fatty acids
Fatty acid elongation
C16 Palmitic acid
16-Hydroxy-palmitate
16-Oxopalmitate
Hexadecanedioate
10,16-Dihydroxypalmitate
Phenylpropanoid biosynthesis
Feruloyl-CoA
16-Feruloyloxypalmitic acid
C18 Oleic acid
18-Hydroxyoleate
18-Oxooleate
Octadec-9-ene-1,18-dioic acid
9,10-Epoxy-stearate
9,10-Epoxy-18-hydroxystearate
9,10,18-Trihydroxystearate
9,10-Dihydroxystearate
C22 Docosanoic acid
22-Hydroxydocosanoate
22-Oxodocosanoate
Docosanedioate
Cutin and suberin biosynthesis (general form)
Long-chain fatty acid
ω-Hydroxy fatty acid
ω-Oxo fatty acid
α,ω-Dicarboxylic acid
ω-Hydroxy epoxy-fatty acid
Polyhydroxy α,ω-dicarboxylic acid
Polyhydroxy-fatty acid
Wax biosynthesis (general form)
Biosynthesis of unsaturated fatty acids
Fatty acid elongation
A long-chain acyl-CoA
A long-chain aldehyde
A long-chain alkane
A long-chain secondary alcohol
A long-chain ketone
CER1
MAH1
A long-chain primary alcohol
A long-chain wax ester
Structure of common wax
Alkane
Aldehydes
Secondary alcohols
Ketone
Diketones
Primary alcohols
Alkyl esters
Data on KEGG graph
Rendered by Pathview

(a)

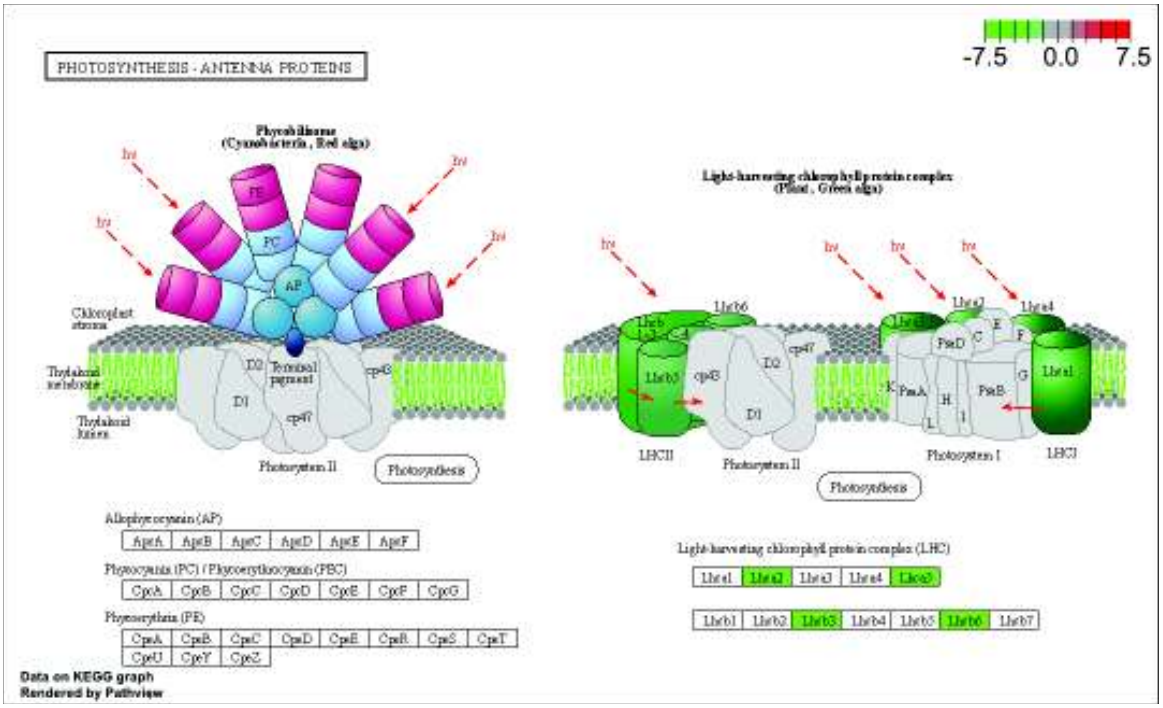


(b)

Figura 18. Ruta metabólica de *Biosíntesis de cutina, suberina y ceras* enriquecida en los contrastes: a) Serrano 20 DDA contra Serrano 60 DDA. b) Chiltepín 20 DDA contra Chiltepín 68 DDA. En rojo se muestran genes inducidos, en verde genes reprimidos, genes sin color indica que no se encontraron en los datos del experimento.



(a)



(b)

Figura 19. Ruta metabólica *Fotosíntesis - proteínas antena* enriquecida en los contrastes: a) Serrano 20 DDA contra Serrano 60 DDA, b) Chiltepín 20 DDA contra Chiltepín 68 DDA. En rojo se muestran genes inducidos, en verde genes reprimidos, genes sin color indica que no se encontraron en los datos del experimento.

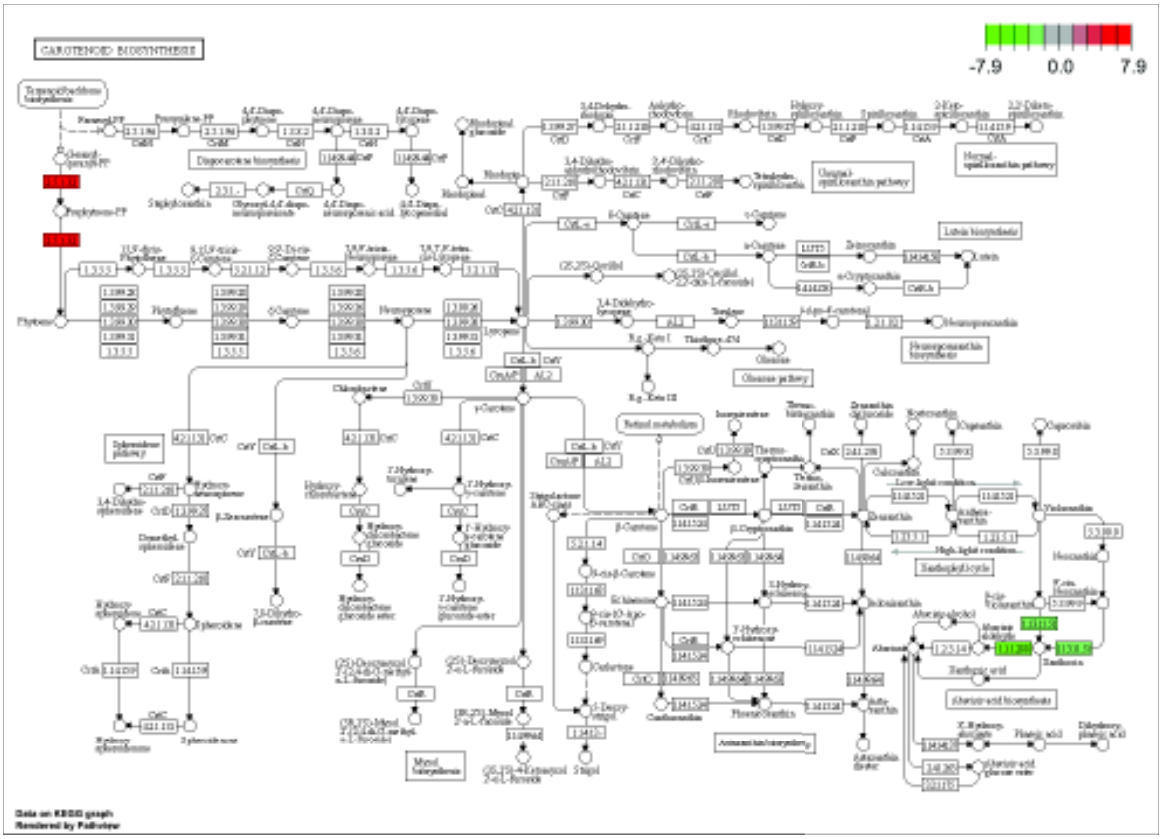


Figura 20. Ruta metabólica de *biosíntesis de carotenoides* enriquecida en el contraste Serrano 20 DDA contra Serrano 60 DDA. En rojo se muestran genes inducidos, en verde genes reprimidos, genes sin color indica que no se encontraron en los datos del experimento

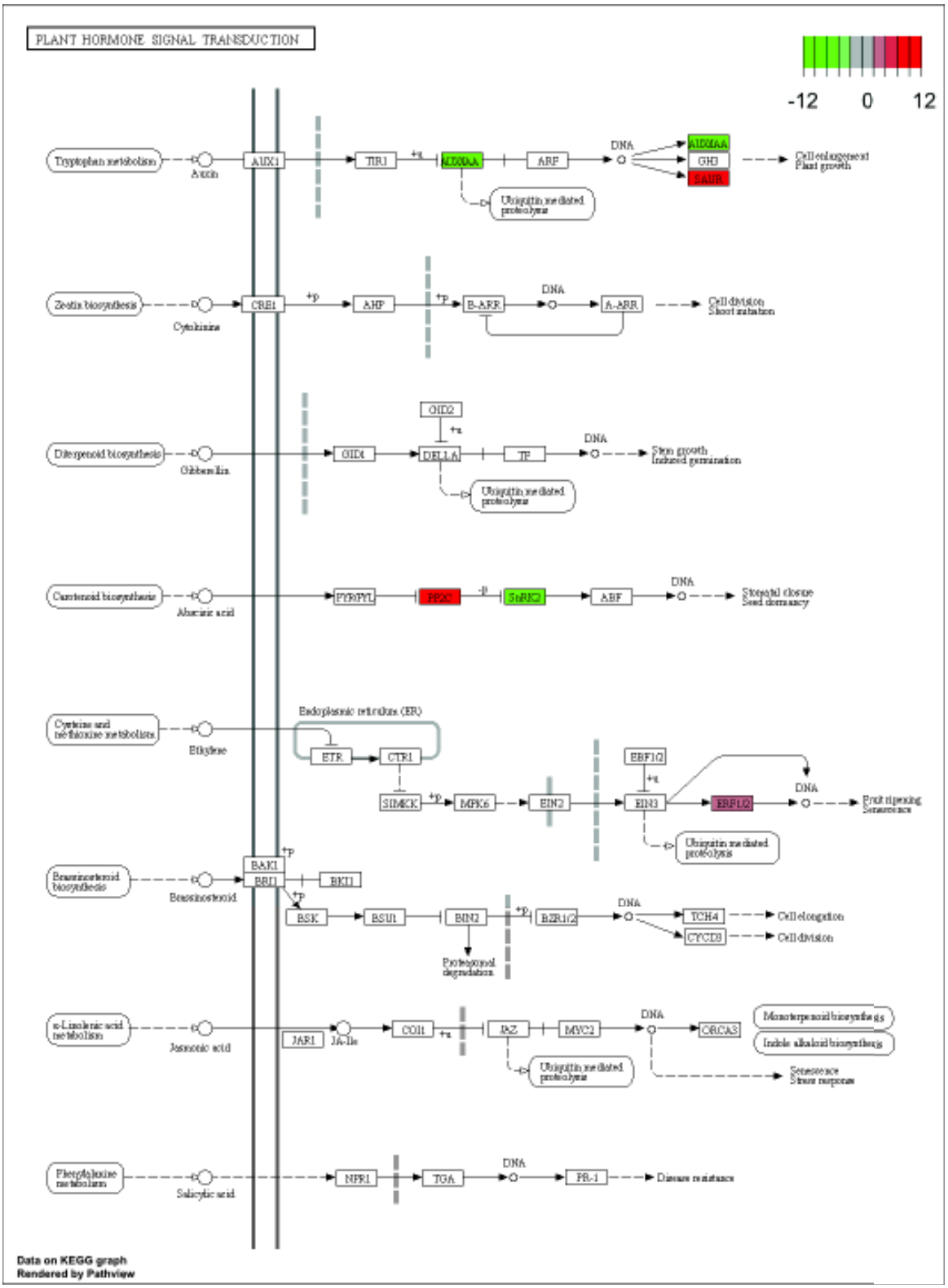


Figura 21. Ruta metabólica de *Transducción de señales de hormonas* enriquecida significativamente en el contraste de Serrano 20 DDA contra Serrano 60 DDA. En rojo se muestran genes inducidos, en verde genes reprimidos, genes sin color indica que no se encontraron en los datos del experimento

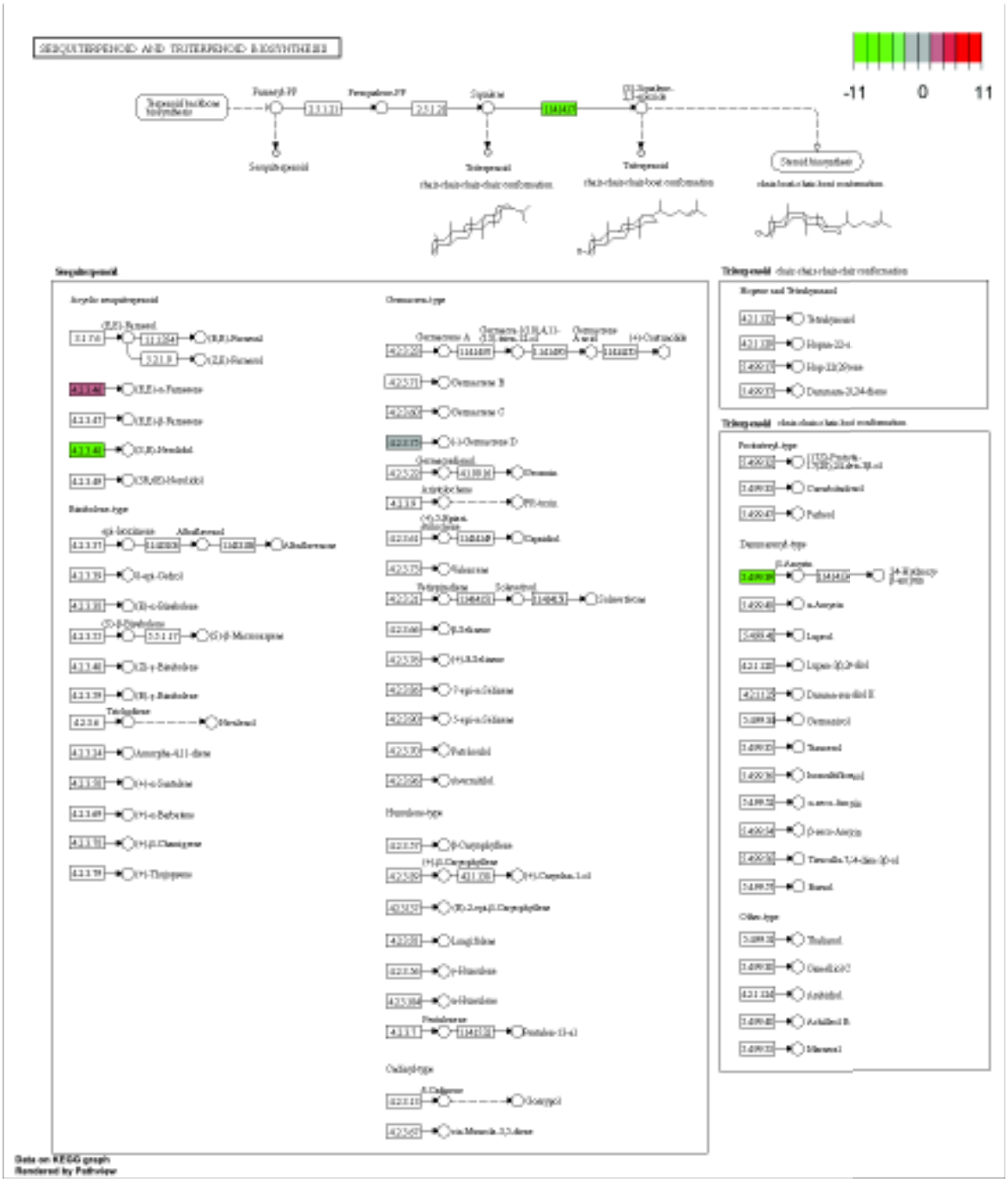


Figura 22. Ruta metabólica *Biosíntesis de sesquiterpenos y triterpenos* enriquecida significativamente en el contraste Chiltepín 20 DDA contra Chiltepín 68 DDA. En rojo se muestran genes inducidos, en verde genes reprimidos, genes sin color indica que no se encontraron en los datos del experimento.

V.7. Análisis de genes expresados diferencialmente con MapMan

Mapman (https://mapman.gabipd.org) es una herramienta que permite la visualización de perfiles de expresión en función de conocimiento existente, posee su propia anotación génica en categorías funcionales (BINs) con una estructura jerárquica definida y que posteriormente puede ser visualizada en diagramas de procesos biológicos (mapas) a través de códigos de color.

Para el análisis de MapMan, se utilizaron todos los genes expresados diferencialmente de cada contraste para identificar grupos de genes asignados a procesos biológicos de interés, en el contraste Ch20-St20 las categorías más abundantes fueron la de estrés y proteína, en el contraste Ch68-Ch60 las categorías con más DEGs fueron estrés, señalización, ARN y hormonas; mientras que las categorías más prevalentes en los contrastes Ch20-Ch68 y St20-St60 fueron: ARN, señalización, proteína, célula, desarrollo, transporte y estrés.

Las figuras 23, 24, 25 y 26 muestran un gráfico resumen de los mapas de MapMan más relevantes en cada contraste. En el mapa de MapMan para el contraste Ch20-St20 (Figura 23) se muestra que la mayoría de los genes se encuentran reprimidos, indicando que estos genes presentan al menos cuatro veces el nivel de expresión en Chiltepín que en Serrano a 20 DDA. La mayoría de los genes corresponden a diferentes tipos de respuesta a estrés, incluyendo el factor de transcripción WRKY44 asociado a la producción de antocianinas, catequinas y taninos, el gen (E)-β-ocimeno sintasa responsable de la producción de ocimeno, un monoterpeno producido como respuesta a ataques herbívoros (Tohge *et al.*, 2005); las proteínas de señalización COP9 involucradas en la señalización hormonal y en el desarrollo; el gen GBP, una GTPasa involucrada en la protección contra patógenos; una proteína BTB/POZ involucrada en la regulación de expresión génica por medio de la mediación de la conformación de la cromatina; y una proteína F-box que actúa como detector de dianas para las ligasas ubiquitina-proteína SCF E3s, responsable de la degradación de proteínas, permitiendo que la célula responda de forma rápida a los cambios en el ambiente. En este mismo contraste se encontraron genes inducidos (con al menos

4 veces el nivel de expresión en Serrano que en Chiltepín a 20 DDA), incluyendo el gen hsp70 relacionado con la modificación de proteínas y protección contra estrés oxidativo y térmico; el gen Ycf1, parte del complejo intermediario de translocación involucrado en la importación de proteínas a los cloroplastos.

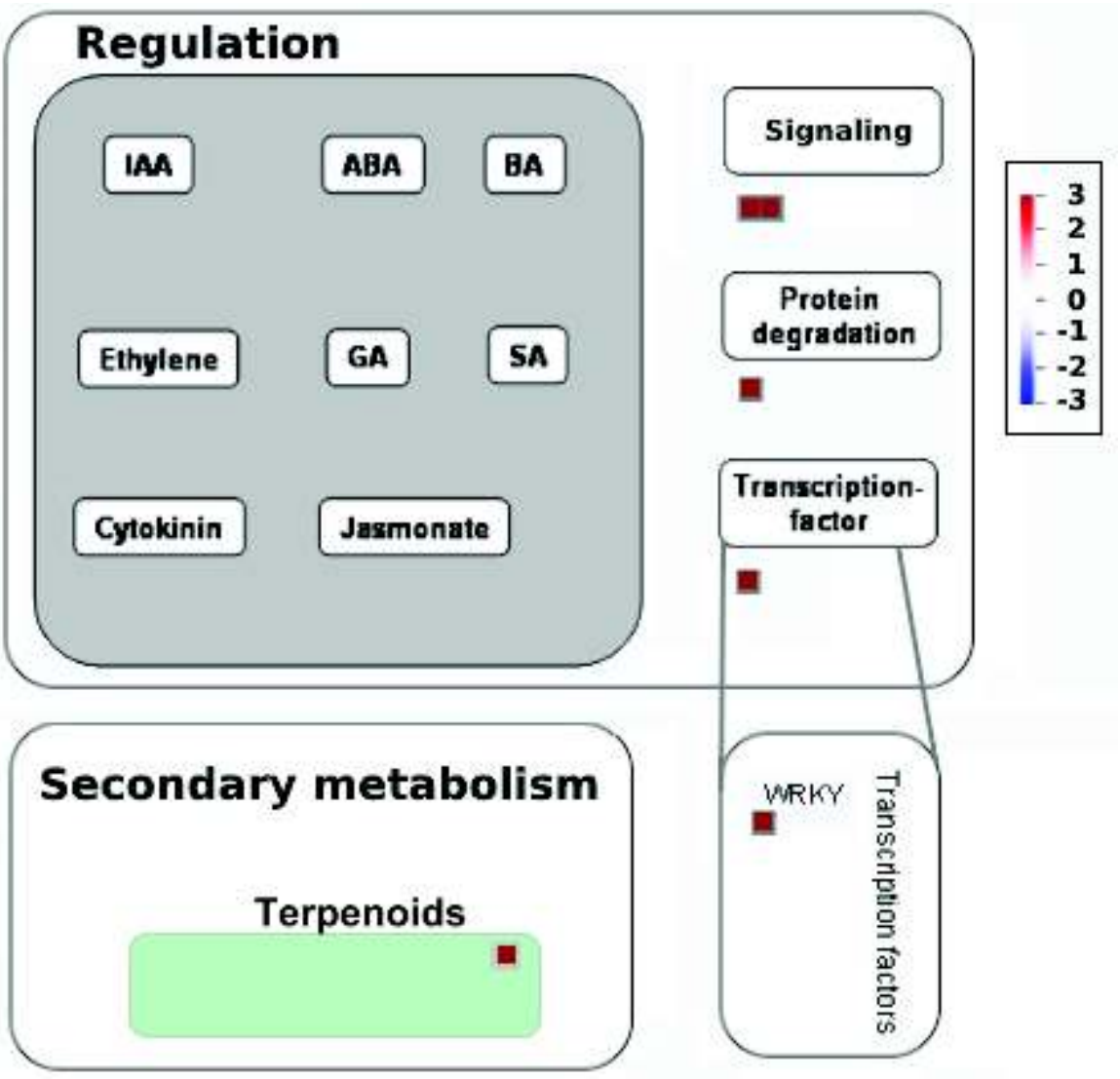


Figura 23. Resumen de mapas de MapMan de los genes expresados diferencialmente (DEGs) en el contraste Ch20-St20, con los mapas de regulación: hormonas, señalización, modificación y degradación de proteínas, factores de transcripción; metabolismo secundario: terpenos, fenil-propanoides, flavonoides, carotenoides, glucosinolatos, etc. Cada recuadro representa un DEG, el color rojo indica que se encuentra inducido y el color azul que se encuentra reprimido.

La figura 24 muestra la imagen resumen de los mapas de MapMan de interés para el con-

traste Ch68-St60, los genes reprimidos indican un nivel de expresión al menos 4 veces mayor en Chiltepín que en Serrano en la etapa 60(68) DDA; los genes inducidos, por lo tanto, indican un nivel de expresión al menos 4 veces mayor en Serrano que en Chiltepín a 60(68) DDA. Entre los genes reprimidos se encuentra el factor de transcripción BH, involucrado en la regulación de etileno; el gen SAUR32 (Small Auxin Responsive Protein), relacionado con el desarrollo de gancho apical y respuesta a deficiencia de hierro; el gen chalcona isomerasa (CHI), elemento clave en la biosíntesis de flavonoides; ILK1 un gen tipo MAPKKK que actúa como elemento de señalización en el proceso de resistencia a patógenos y nuevamente se encontraron los genes GBP y COP9 reprimidos. Entre los genes inducidos se encuentran el receptor de Etileno 1 (ETR1), importante regulador en la ruta de transducción de señales por etileno; el factor de transcripción FUL2, una proteína MADS-Box que sirve como regulador de diferentes procesos de maduración como ablandamiento de fruto, acumulación de carotenoides y compuestos volátiles y modificación de la pared celular y el gen SR1|P1, que codifica a una proteína de respuesta fototrópica involucrada en la defensa de plantas por ubiquitinación.

La figura 25 muestra el gráfico resumen de las rutas metabólicas de MapMan del contraste St20-St60, la categoría con el mayor número de genes expresados diferencialmente fue la de *factores de transcripción* con 29 DEGs, de los cuales solo 4 se encontraron inducidos y 25 se encuentran reprimidos. Algunos genes de esta categoría se encuentran relacionados con el proceso de desarrollo y maduración de fruto: el gen DA1 es un receptor de ubiquitina que limita el tamaño final de las semillas al regular la proliferación celular, se encontraron 3 factores de transcripción IAA (IAA8, IAA27 y IAA32) reprimidos, involucrados en el silenciamiento de los genes de auxina tempranos en concentraciones bajas de auxina (Liscum y Reed, 2002); se encontraron 2 factores de transcripción WRKY (reprimidos), relacionados con la respuesta a estrés y la producción de taninos y antocianinas; 3 factores de transcripción Homeobox (2 reprimidos y 1 inducido) que podrían estar relacionados con la regulación negativa de la elongación

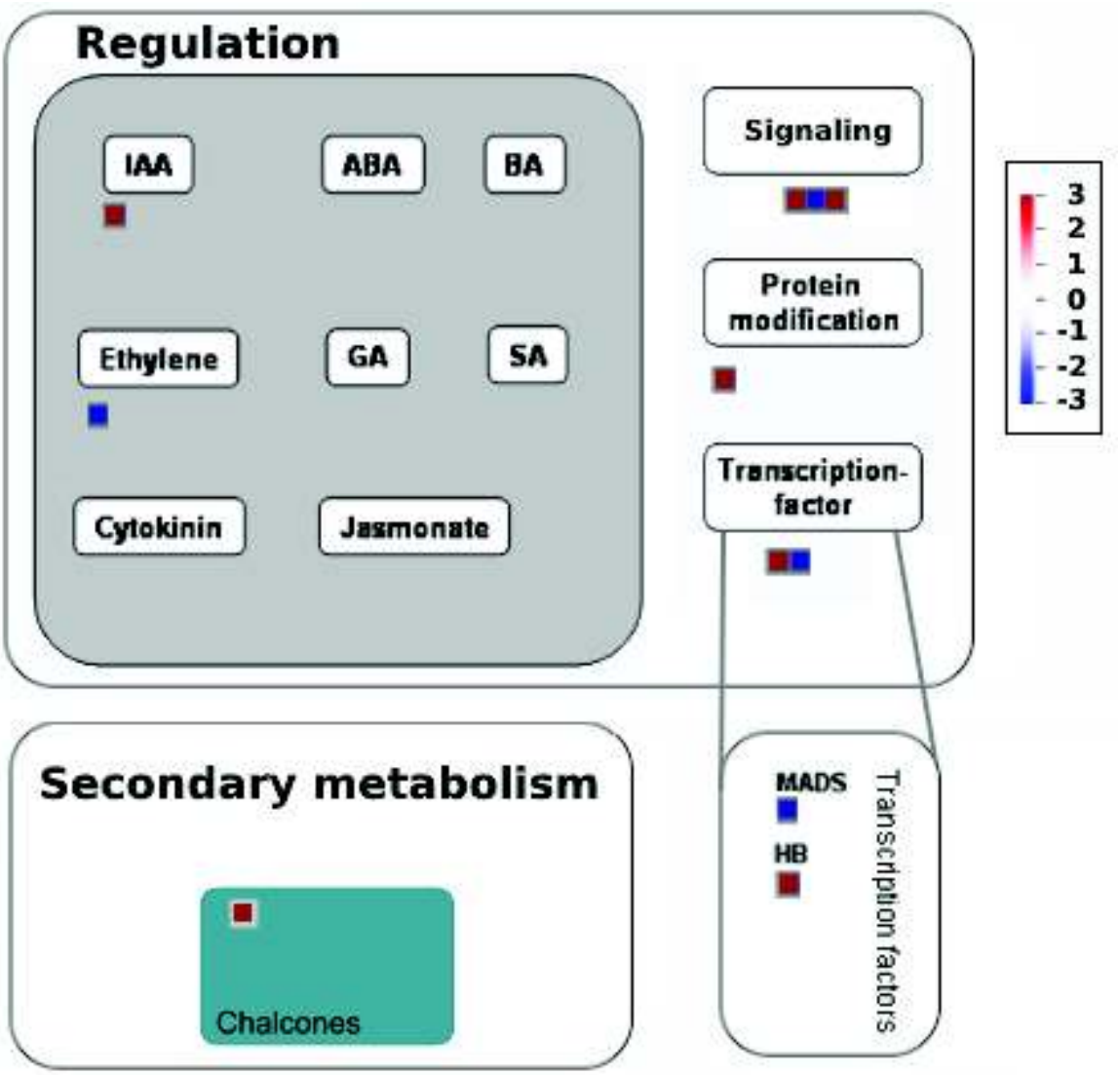


Figura 24. Resumen de mapas de MapMan de los genes expresados diferencialmente (DEGs) en el contraste Ch68-St60, con los mapas de regulación: hormonas, señalización, modificación y degradación de proteínas, factores de transcripción; metabolismo secundario: terpenos, fenil-propanoides, flavonoides, carotenoides, gluscosinolatos, etc. Cada recuadro representa un DEG, el color rojo indica que se encuentra inducido y el color azul que se encuentra reprimido.

celular y proliferación; 2 factores de transcripción bZIP que podrían estar relacionados con la regulación del crecimiento bajo condiciones de estrés hídrico.

Otra categoría con un vasto número de genes expresados diferencialmente es la de *señalización*, con 20 genes reprimidos, entre los cuales se encuentra el gen LSH10, involucrado en la estimulación del crecimiento celular en respuesta a estímulo lumínico; el gen THE1, que se encuentra asociado a la elongación celular durante el crecimiento vegetativo; el gen NPY2, esencial en organogénesis mediada por auxina; el gen TMK1, involucrado en la transducción de señales por AUX y la regulación de la expansión y proliferación celular.

En la categoría de *hormonas* identificamos 8 genes expresados diferencialmente, 6 de ellos reprimidos y los 2 restantes inducidos, incluyendo: 3 factores de transcripción relacionados con el Ácido Abscísico (2 reprimidos, 1 inducido) que codifican a proteínas fosfatasas 2C (PP2C) que se encuentran involucradas en la expansión celular por el mecanismo de crecimiento ácido; 2 miembros de la familia PIN (ambos reprimidos), facilitadores del flujo de auxina, relacionados con el transporte de AUX; los genes ACCO2 (reprimido) y ERF1 (inducido) involucrados en la cascada de biosíntesis de etileno y un gen SNAK1 (inducido) que codifica a una proteína relacionada con giberelinas con actividad antimicrobiana.

La figura 26 resume las categorías MapMan de interés para el contraste Ch20-Ch68, en ella se muestran 73 genes relacionados con señalización, 7 de ellos inducidos, incluyendo el gen MAPKKK18, un transductor de ABA en el contexto de estrés abiótico. También se encuentran 66 genes reprimidos, incluyendo ALE2 involucrado en la diferenciación del protodermis en brotes, epidermis y cutícula; el gen GHR1, un receptor de ABA necesario para la regulación de movimiento estomal; los genes TMK1 y TMK3, involucrados en la transducción y regulación de la expansión y proliferación celular; el gen LSH que es un promotor del crecimiento celular como respuesta a estímulo lumínico; el gen NPY2, involucrado en la organogénesis mediada por auxina; el gen THE1, un receptor requerido para la expansión celular y 8 genes relacionados

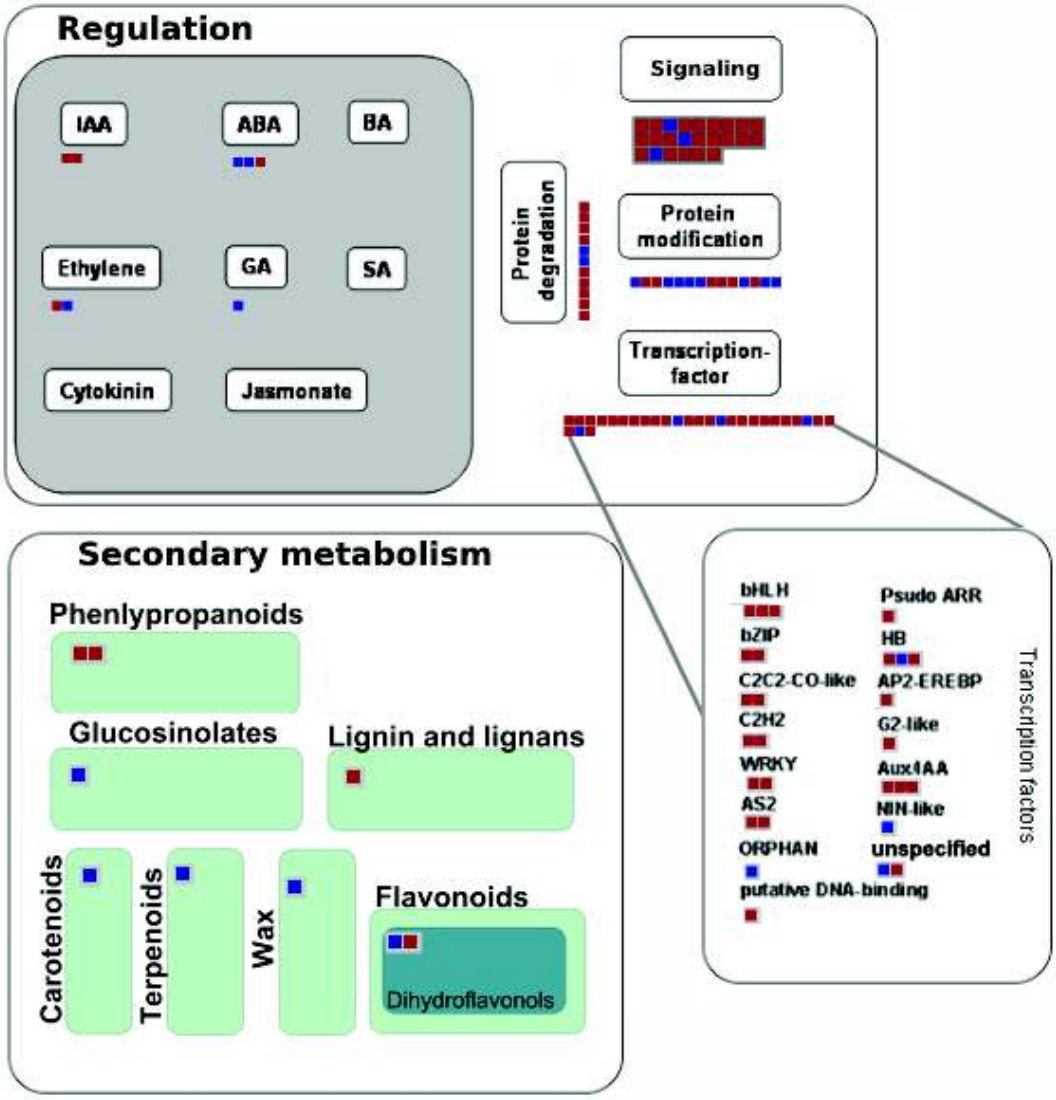


Figura 25. Resumen de mapas de MapMan de los genes expresados diferencialmente (DEGs) en el contraste St20-St60, con los mapas de regulación: hormonas, señalización, modificación y degradación de proteínas, factores de transcripción; metabolismo secundario: terpenos, fenil-propanoides, flavonoides, carotenoides, glucosinolatos, etc. Cada recuadro representa un DEG, el color rojo indica que se encuentra inducido y el color azul que se encuentra reprimido.

con la defensa (SR1P1, M3K5G, NIK1,GLR34, NCL1, EDR2L, IQM3, PICBP).

Otra categoría importante en el contraste es la de *hormonas de plantas* con 25 genes expresados diferencialmente, 17 de ellos reprimidos y 8 inducidos, en esta categoría se encuentran 3 genes LOX (LOX21 y LOX15 reprimidos y LOX15 inducido) que están relacionados con diferentes procesos de las plantas, como crecimiento y desarrollo, resistencia ante patógenos y senescencia; también se encontraron 2 genes ACCO (ACCO1 y ACCO2 ambos reprimidos), relacionados con la biosíntesis de etileno, ACCO1 es un gen clave en el proceso de maduración de frutos climatéricos; 5 genes expresados como una respuesta a estrés (SNAK2, SC5D y AHK4 reprimidos, SNAK1 y PTI5 inducidos). Finalmente, otra categoría con una cantidad enorme de genes expresados diferencialmente es la de *factores de transcripción* con 87 DEGs asignados de los cuales 70 se encuentran reprimidos y 17 inducidos, algunos tipos de factores de transcripción fueron muy frecuentes, por ejemplo, se encontraron 10 factores de transcripción C2H2 (8 reprimidos y 2 inducidos), la mayoría de ellos relacionados con la respuesta a estrés, pero, algunos se encuentran relacionados en otros procesos como homeostasis de hierro y regulación de ABA; 7 factores de transcripción MYB reprimidos, involucrados en un conjunto diverso de aspectos de la fisiología de plantas, como crecimiento y desarrollo, respuesta a estrés y regulación de genes responsable de la biosíntesis y regulación de ABA; 8 factores de transcripción bzip (5 reprimidos y 3 inducidos) la mayoría de los cuales funcionan como activadores transcripcionales, regulando el crecimiento en condiciones de estrés hídrico; 7 factores de transcripción MADS (4 reprimidos y 3 inducidos), relacionados con el desarrollo floral y maduración de fruto, incluyendo la acumulación de carotenoides, modificación de pared celular y senescencia, especialmente FUL1, FUL2 y AGL1; 4 factores de transcripción WRKY (3 reprimidos y 1 inducido), todos ellos relacionados con la respuesta a estrés y algunos de ellos con la biosíntesis de antocianinas y taninos.

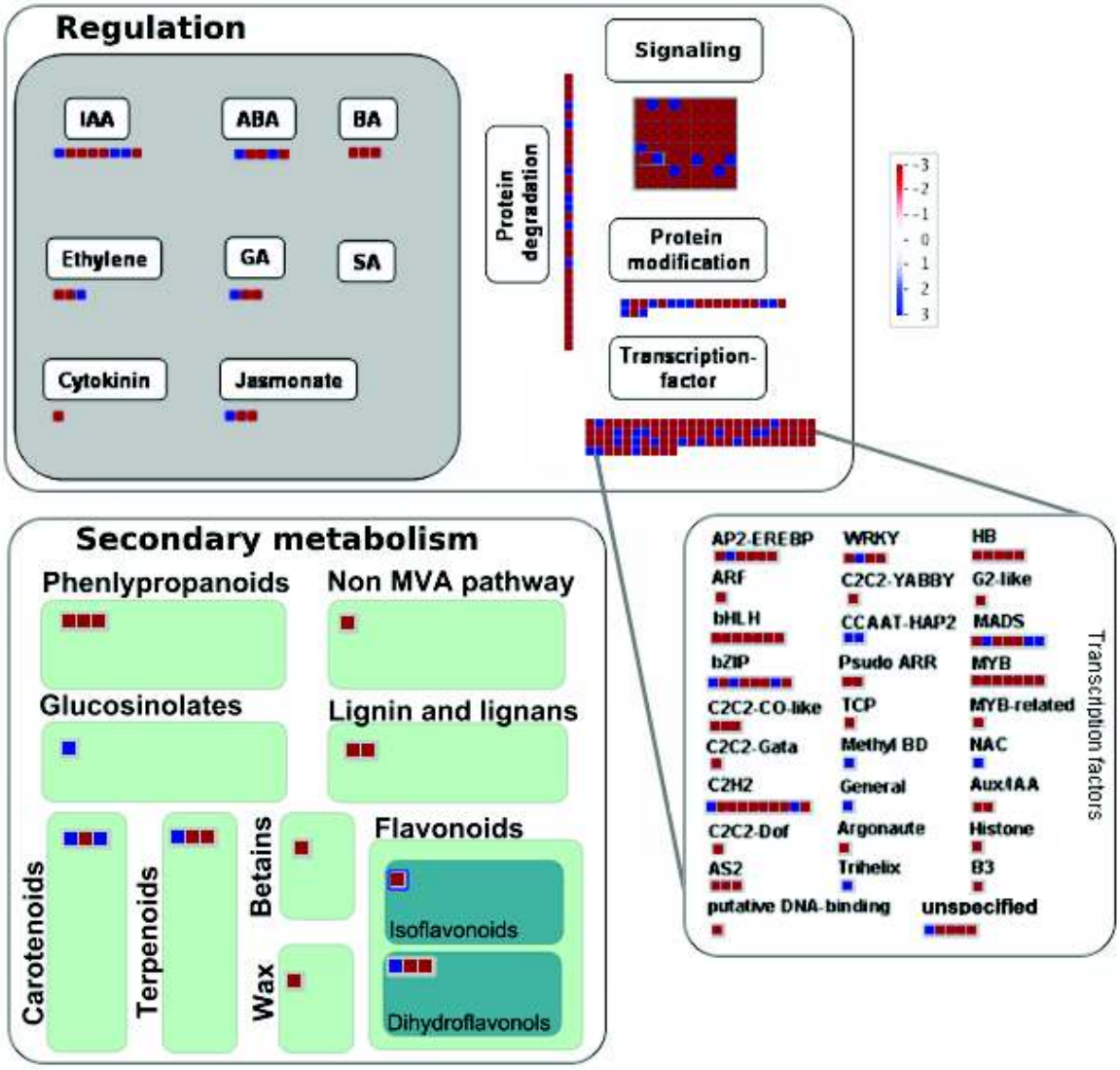


Figura 26. Resumen de mapas de MapMan de los genes expresados diferencialmente (DEGs) en el contraste Ch20-Ch68, con los mapas de regulación: hormonas, señalización, modificación y degradación de proteínas, factores de transcripción; metabolismo secundario: terpenos, fenilpropanoides, flavonoides, carotenoides, glucosinolatos, etc. Cada recuadro representa un DEG, el color rojo indica que se encuentra inducido y el color azul que se encuentra reprimido.

V.8. Genes involucrados en el desarrollo y maduración de fruto

El chile es un fruto no-climatérico, aún no se ha dilucidado el mecanismo molecular responsable del proceso de desarrollo y maduración de fruto; sin embargo, el tomate, que también es parte de la familia de las solanáceas y es un pariente cercano del género *Capsicum*, es un fruto climatérico cuyo proceso de maduración ha sido muy estudiado utilizando diferentes mutantes de maduración como el no-madura (nor), nunca-madura (nr), no-madura sin color (cnr), inhibidores de maduración (rin). En este trabajo de investigación seleccionamos 105 genes de tomate involucrados en diferentes partes del proceso de desarrollo y maduración de fruto, estos genes se utilizaron para buscar genes homólogos en el genoma de chile. En total se identificaron 778 genes candidatos en el genoma de chile, de los cuales 53 se encontraron expresados diferencialmente en al menos uno de los contraste usados.

La figura 27 muestra un mapa de calor de los 53 genes ortólogos con tomate expresados diferencialmente, se agruparon los datos por genes y por muestras basandose en los niveles de expresión, las muestras de Chiltepín y Serrano son perfectamente agrupadas de acuerdo a la etapa de maduración (20 DDA y 60/68 DDA), mientras que por gen se crearon 7 grupos.

El grupo número 1 contiene solo un gen ERF6 que es un factor de transcripción que regula negativamente la biosíntesis de etileno y carotenoides (Lee *et al.*, 2012) y, se encuentra expresado en las dos etapas de maduración de Serrano con un pico en 60DAA. El grupo 2 contiene 13 genes expresados diferencialmente con un patrón de expresión similar: baja expresión a 20 DDA y mayor expresión a 60 (68) DDA, aunque, su expresión tiene un patrón similar en ambas variedades, la mayoría de los genes se encuentra expresados diferencialmente solo en el contraste Ch20-Ch68, excepto por el gen PSY1, que es un elemento clave en la biosíntesis de carotenoides (Bird *et al.*, 1991) y el gen PG2 que cataliza la depolimerización de las pectinas en el prericarpio (Lincoln y Fischer, 1988) (inducido en los contrastes Ch20-Ch68 y St20-St60); en este cluster también se pueden encontrar el factor de transcripción AGAMOUS-LIKE 1 (TAGL1), regulador

Tabla VIII. Genes de tomate relacionados con el desarrollo y maduración de fruto y sus correspondientes genes ortólogos en chile, E-valor y bitscore de los alineamientos realizados con BLASTp para identificarlos.

Gen en Tomate	Gen en *C. annuum* CM334	E-valor	Score(bits)
Alpha-mannosidase	CA.PGAv.1.6.scaffold423.45	0	2052
AOX1a	CA.PGAv.1.6.scaffold1731.3	0	670
AP2	CA.PGAv.1.6.scaffold606.27	3.83E-175	69.2
ARF4	CA.PGAv.1.6.scaffold631.7	0	1652
ASR1	CA.PGAv.1.6.scaffold264.4	1.76E-97	288
Beta-hex	CA.PGAv.1.6.scaffold1361.13	0	1192
CGS	CA.PGAv.1.6.scaffold979.1	0	1088
CGS	CA.PGAv.1.6.scaffold991.4	0	484
CHS1	CA.PGAv.1.6.scaffold305.2	4.62E-55	175
CHS1	CA.PGAv.1.6.scaffold765.1	0	809
ERF6	CA.PGAv.1.6.scaffold14.17	9.19E-153	427
ETR4	CA.PGAv.1.6.scaffold374.1	0	1503
FUL1	CA.PGAv.1.6.scaffold587.54	0	503
FUL2	CA.PGAv.1.6.scaffold1592.5	0	512
GDH1	CA.PGAv.1.6.scaffold960.48	0	848
GLK2	CA.PGAv.1.6.scaffold782.26	0	565
Hydroquinone glucosyltransferase	CA.PGAv.1.6.scaffold239.6	1.76E-41	145
LeSPL-CNR	CA.PGAv.1.6.scaffold529.23	1.01E-86	249
Lutescent 2	CA.PGAv.1.6.scaffold128.144	0	976
MADS-box protein 1	CA.PGAv.1.6.scaffold1592.4	1.20E-176	485
MADS-RIN	CA.PGAv.1.6.scaffold1021.30	0	504
MADS-RIN	CA.PGAv.1.6.scaffold1091.1	0	501
NAC1	CA.PGAv.1.6.scaffold1861.9	0	630
NSGT1	CA.PGAv.1.6.scaffold92.26	5.36E-110	328
PG2	CA.PGAv.1.6.scaffold128.13	0	504
PSY1	CA.PGAv.1.6.scaffold404.3	0	860
SGR1	CA.PGAv.1.6.scaffold981.58	0	507
SlDML2	CA.PGAv.1.6.scaffold1138.12	0	3378
SlscADH1	CA.PGAv.1.6.scaffold815.5	0	549
spdsyn	CA.PGAv.1.6.scaffold615.4	0	689
TAGL1	CA.PGAv.1.6.scaffold585.57	3.75E-172	475
TCTR1	CA.PGAv.1.6.scaffold1138.13	0	715
TCTR1	CA.PGAv.1.6.scaffold1138.14	0	837
TCTR1	CA.PGAv.1.6.scaffold824.28	0	1517
TomloxC	CA.PGAv.1.6.scaffold649.19	0	1773
TomloxC	CA.PGAv.1.6.scaffold649.21	3.76E-83	261
TomloxC	CA.PGAv.1.6.scaffold460.8	0	632
XTH5	CA.PGAv.1.6.scaffold700.22	0	719

de la acumulación de carotenoides, del grosor de pericarpio y la biosíntesis de etileno y uno de los principales reguladores de los genes PSY y PC (Vrebalov *et al.*, 2009).

El tercer grupo contiene un conjunto de genes interesantes con expresión en Serrano en ambas etapas pero en Chiltepín solamente a 20 DDA, los genes APETALA2 (reprimido en el contraste Ch20-Ch68) es un elemento en respuesta a etileno que modula la maduración a través de los genes involucrados en la biosíntesis de etileno y carotenoides (Chung *et al.*, 2010); dos genes reprimidos en el contraste Ch20-Ch68 e inducidos en Ch68-St60 mientras que se mantiene alrededor de la misma expresión en las dos etapas de Serrano: el gen FUL2 que regula positivamente varios aspectos de la maduración de fruto como el ablandamiento de fruto, la acumulación de carotenoides, el metabolismo de sucrosa, las modificaciones de pared celular, formación de cutícula y la biosíntesis de etileno (Verweij *et al.*, 2016) y el gen ACO3, una de las enzimas principales en la biosíntesis de etileno (Barry *et al.*, 2000).

El cuarto grupo contiene solo el gen FSM1 expresado solamente en Chiltepín a 20 DDA, este factor de transcripción parece regular varios aspectos del crecimiento de fruto tierno (fase de crecimiento II) (Barg *et al.*, 2005).

El grupo 5 contiene 4 genes expresados diferencialmente reprimidos en el contraste Ch20-Ch68, incluidos un gen tipo OVATE, que es un importante regulador negativo de la elongación de fruto y es responsable de la forma redonda en los frutos silvestres de tomate (Wu *et al.*, 2015); también se encuentra el factor de transcripción MYB12 que es un regulador de las enzimas *chalcona sintasa* y *flavonol sintasa* que juegan un importante rol en la biosíntesis de flavonoides (Mehrtens *et al.*, 2005); el gen FAS, que controla el número de lóculos y regula la expresión de varios genes involucrados en procesos como desarrollo de flores y meristemos, actividad de acoplamiento de microtúbulos y biosíntesis de esteroles Chu *et al.* (2019) y una copia del gen TAGL1.

El grupo 6 y 7 contiene genes expresados a 20 DDA pero no en la etapa madura (60/68

DDA) en ambas variedades, todos los genes en el grupo 6 se encuentran reprimidos en el contraste Ch20-Ch68, e incluyen genes relacionados con etileno, auxinas y regulación de ácido abscísico. El grupo 7 contiene 3 genes expresados diferencialmente en el contraste St20-St60 (reprimido), dos copias de la proteína de respuesta a auxina 9 (IAA9), un factor transcripcional que actúa como represor de los genes de respuesta temprana a auxina (Liscum y Reed, 2002), y el Inhibidor de Actividad Meristémica (IMA) una proteína mini dedo de zinc que regula la actividad meristémica y el proceso de desarrollo de flores y óvulos Bollier *et al.* (2018). Hay 10 genes expresados diferencialmente en más de uno de los contrastes en el grupo 7, todos con un patrón de expresión similar y ambos se encuentran reprimidos en los contrastes Ch20-Ch68 y St20-St60.

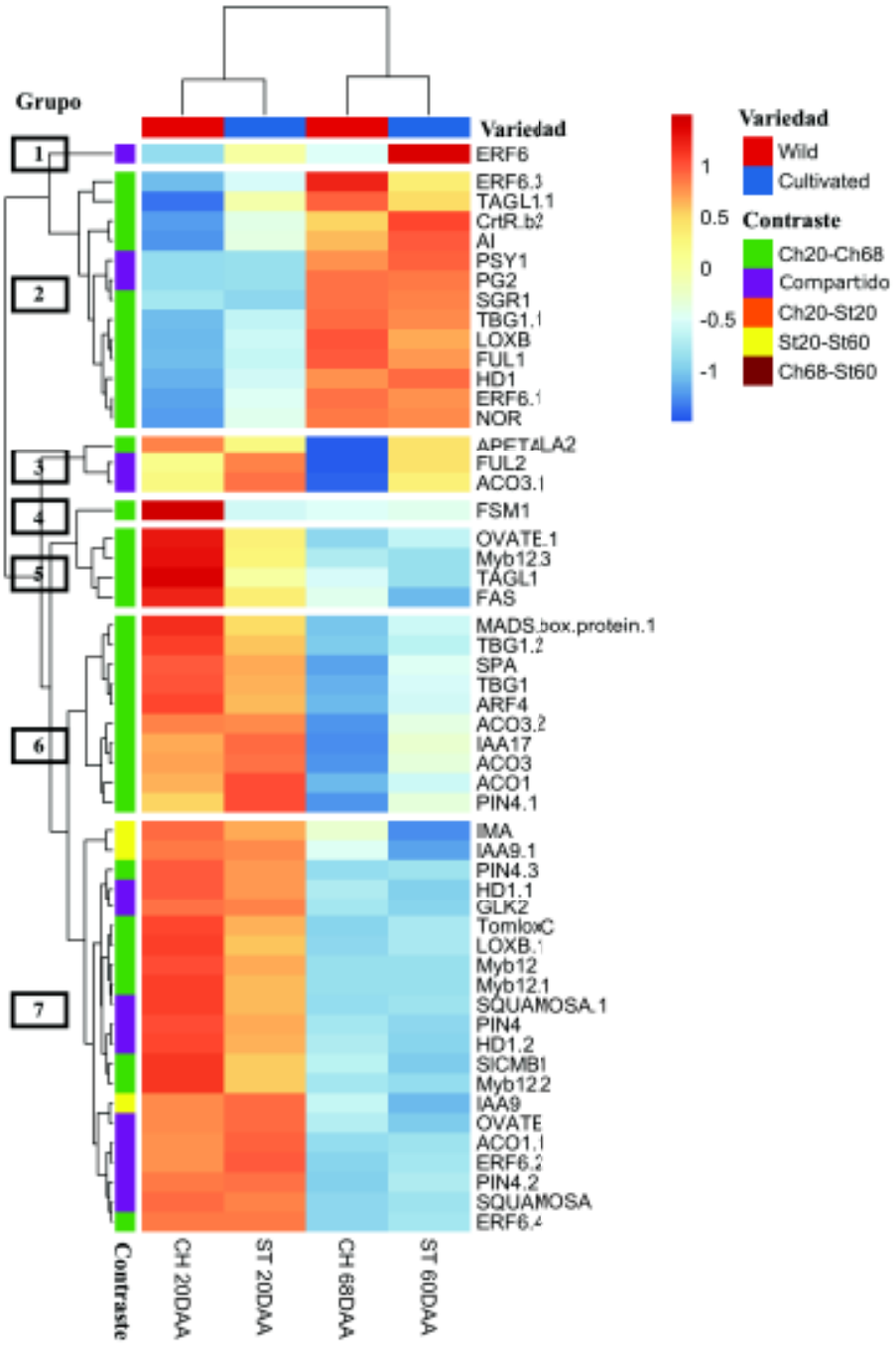


Figura 27. Representación en mapa de calor (Heatmap) de los genes expresados diferencialmente (DEGs) de chile, ortólogos con los genes de tomate relacionados con el desarrollo y maduración de fruto, agrupados por patrón de expresión en muestras (columnas) y genes (filas), el color de las cajas a la izquierda de cada fila indica el contraste en el cual se encuentra expresado diferencialmente (DE) el correspondiente gen, una caja púrpura (compartidos) indica que el gen correspondiente se encuentra DE en dos o más contrastes. La escala de color indica el nivel de expresión de los genes en la muestra correspondiente, expresados en conteos por millón normalizados y escalados con transformada Z.

V.9. Genes localizados en regiones con señales de barrido selectivo

Cuando aparece un alelo con una mutación beneficiosa en una población se esparce por selección natural hasta fijarse en la población, cuando esto ocurre se produce una disminución de la variabilidad en las secuencias de nucleótidos cercanos al alelo fijado, este efecto es conocido como barrido selectivo. Durante el proceso de domesticación se lleva a cabo una selección artificial que produce una fijación rápida del alelo beneficioso, y con ello también se fijan alelos ligados a él (Stephan, 2019).

En un estudio previo, Qin *et al.* (2014) identificaron 115 regiones con señales de barrido selectivo con 511 genes en el genoma de *Capsicum annuum* cv. Zunla-1. En este estudio se identificaron 475 genes en el genoma de *Capsicum annuum* cv. CM334 equivalentes a los genes de regiones con señal de barrido selectivo reportados para el cultivar Zunla-1. La figura 28,A muestra un mapa de calor de 234 genes asociados a barrido selectivo que se encontraron expresados en las librerías de Chiltepín y Serrano; la figura 28,B muestra un mapa de calor de 19 genes de barrido selectivo que se encontraron expresados diferencialmente en alguno de los contrastes de Serrano y Chiltepín, 9 DEGs se encontraron exclusivamente en Chiltepín: el gen PTI5 activador de genes de defensa, el gen transportador de lípidos NLPT1, el gen GAPA relacionado con la resistencia a choque térmico, el gen de respuesta a estrés abiótico NCL1 y dos copias del gen RPP13 relacionado con la resistencia a enfermedades. 4 genes de encontraron expresados diferencialmente exclusivamente en Serrano, los factores de transcripción IWS1 relacionados con definición del complejo de elongación de la ARN polimerasa II y la regulación de la maduración de ARN; y EFR1B, elemento clave en la señalización de etileno y jasmonato como respuesta de defensa ante estrés; el gen de resistencia a enferemedades RPP13 y un componente del complejo SWR1 que controla la regulación transcripcional a través de la remodelación de la cromatina.

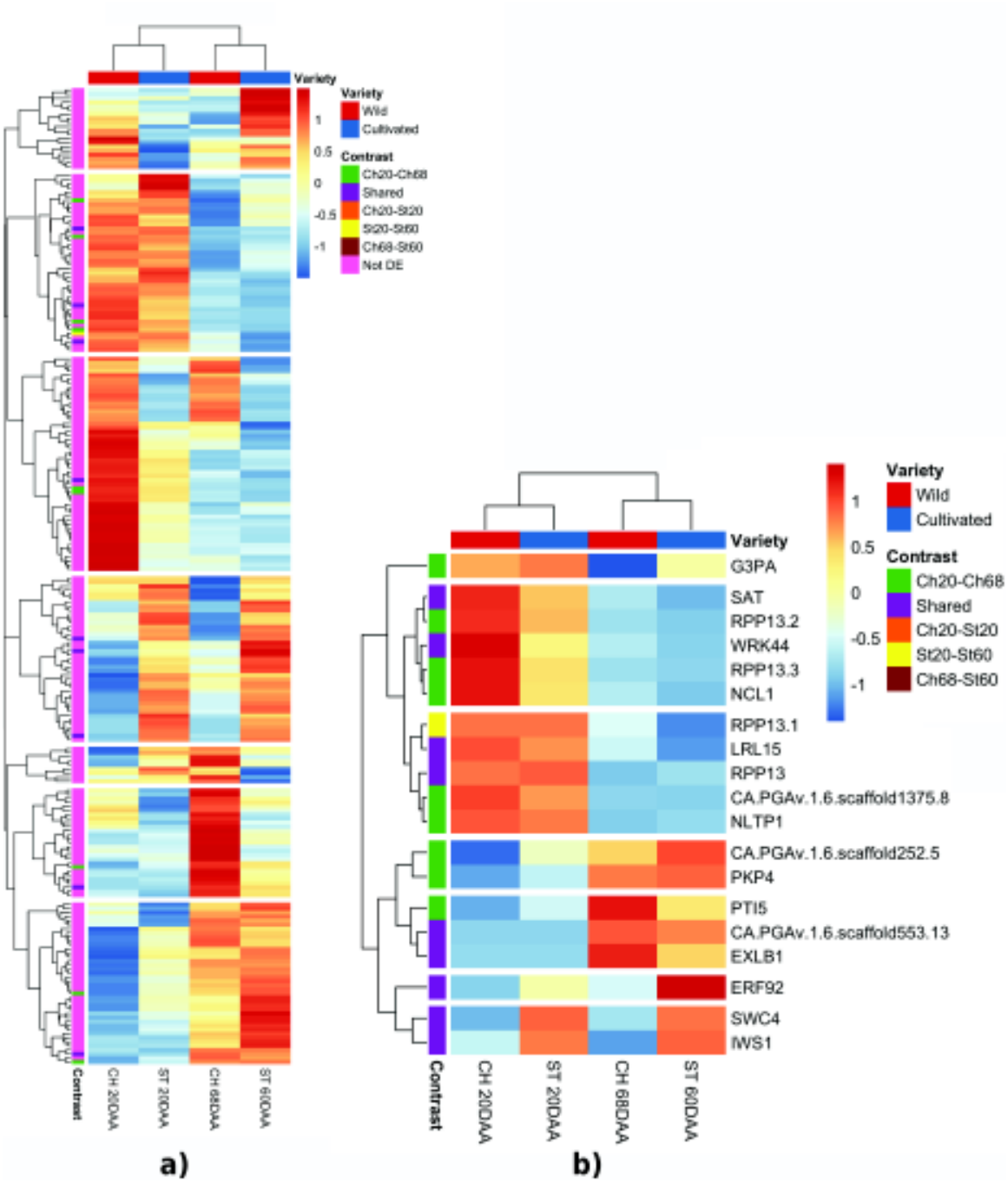


Figura 28. Representación en mapa de calor (Heatmap) de los genes de chile asociados a regiones de barrido selectivo, agrupados por patrón de expresión en muestras (columnas) y genes (filas), el color de las cajas a la izquierda de cada fila indica el contraste en el cual se encuentra expresado diferencialmente (DE) el correspondiente gen, una caja púrpura (compartidos) indica que el gen correspondiente se encuentra DE en dos o más contrastes, una caja roja indica que el gen no se encuentra expresado diferencialmente. La escala de color indica el nivel de expresión de los genes en la muestra correspondiente, expresados en conteos por millón normalizados y escalados con transformada Z. a) totalidad de genes expresados en las 4 muestras de Serrano (ST) y Chiltepín (CH), b) genes expresados diferencialmente en al menos uno de los contrastes de Serrano (ST) y Chiltepín (CH) utilizados.

V.10. Análisis de ARNs no codificantes en Chiltepín

V.10.1. Ensamblado *ab initio* de Chiltepín

Para realizar la identificación de genes con potencial ARN no codificante se realizó un ensamblado *ab initio* utilizando como referencia el genoma de Chile (cv. CM334) y las lecturas de secuenciación de Chiltepín previamente alineadas a él, el ensamblador Stringtie utiliza la información de los alineamientos al genoma para identificar genes y transcritos anotados, es decir, que ya se encuentran reportados en el archivo de anotación (GFF) del genoma; de igual manera, identifica elementos novedosos que aún no se encuentran en el archivo de anotación. La figura 29 muestra la cantidad de genes y transcritos ensamblados por Stringtie, se ensamblaron un total de 45,780 genes de los cuales 34,834 (76.09 %) corresponden con la anotación del genoma (anotados), 9,514 (20.78 %) genes son novedosos pues no se encontraron en el archivo de anotación y 1,432 (3.12 %) genes se encuentran parcialmente anotados, es decir, una parte de ellos se encuentra en genes anotados pero no su totalidad (figura 29,A). De igual manera, se ensamblaron un total de 69,630 transcritos de los cuales 35,884 (51.53 %)corresponden con la anotación del genoma, 13,218 (18.98 %) genes son novedosos y 20,528 (29.48 %) genes se encuentran parcialmente anotados (figura 29,B).

Una vez ensamblado el transcriptoma *ab initio* se compararon los transcritos ensamblados con el genoma de referencia para obtener los códigos de clase asignados por GffCompare, esto nos permitió identificar 24,930 ARNnc potenciales, los cuales se clasificaron de acuerdo a su posición respecto a los genes codificantes a proteínas en: transcritos intergénicos largos (lincRNAs, código de clase *u*) posicionados en regiones entre dos genes, transcritos intrónicos (código de clase *i*) formados por intrones de un gen codificante a proteína, transcritos con traslape intrónico (código de clase *j*) que pueden ser isoformas no reportadas previamente, transcritos de traslape exónico (código de clase *o*) que coinciden parcialmente con exones anotados en el genoma de

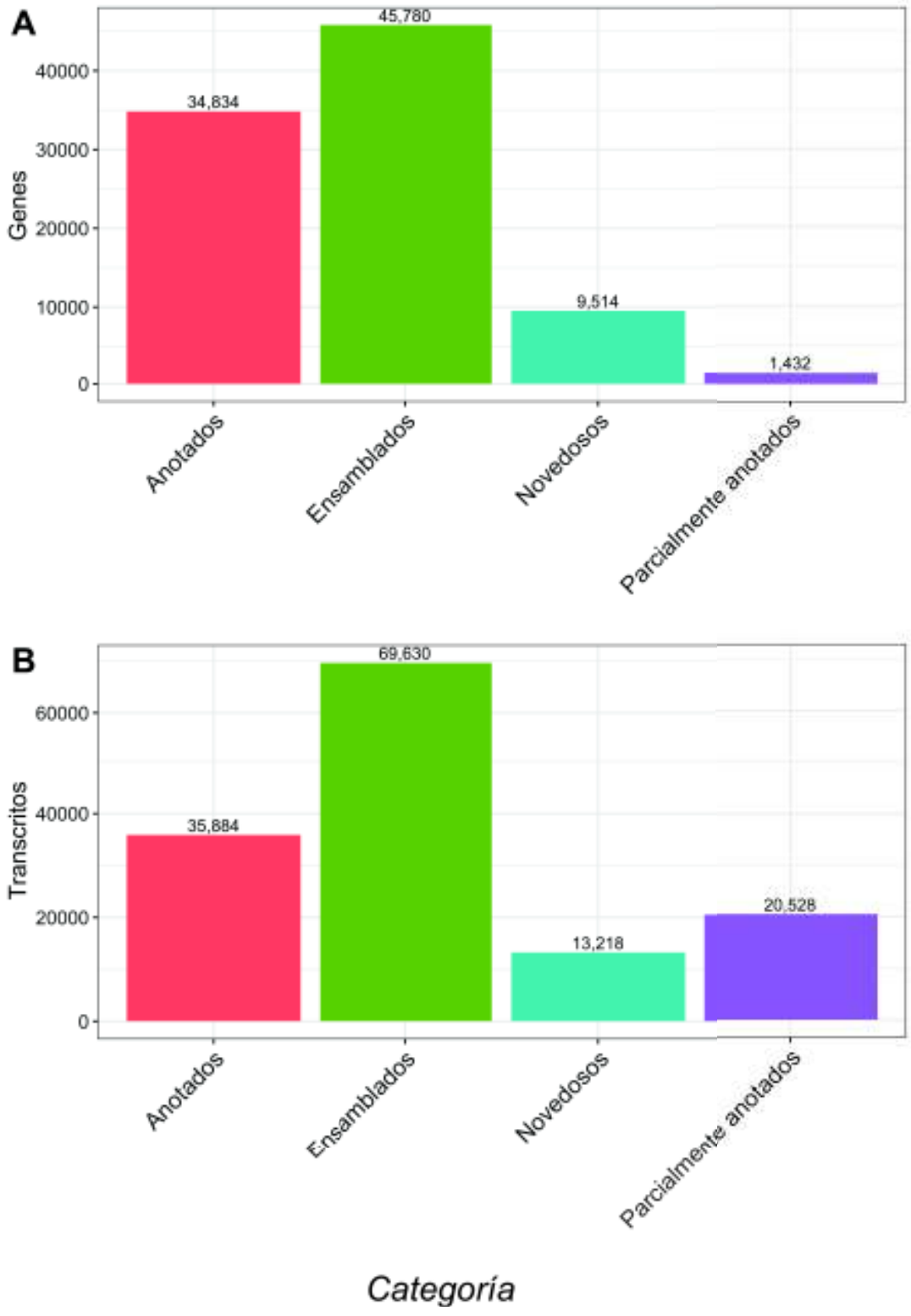


Figura 29. Cantidad de genes y transcritos pertenecientes al transcriptoma ensamblado *ab initio* por Stringtie con las lecturas de secuenciación de Chiltepín y el genoma de referencia de Chile cv. CM334.

referencia y transcritos de traslape exónico antisentido (código de clase *x*) que como su nombre lo indica son aquellos que coinciden parcialmente con exones anotados previamente, pero se encuentran en la cadena de ADN opuesta.

La figura 30 muestra la cantidad de transcritos potenciales ARN no codificantes por categoría de acuerdo a su posición, la mayoría de los transcritos (53.02 %) son lincARNs, que en organismos como el ser humano éstos representan la mitad de todos los ARN no codificantes largos (Ransohoff *et al.*, 2018); la segunda categoría con mayor cantidad de transcritos fue traslape intrónico (36 %), seguido de translape exónico antisentido (5.33 %), traslape exónico (4.7 %) y transcritos contenidos intrónicamente (9.8 %).

V.10.2. Identificación de transcritos con potencial codificante

Después de filtrar los transcritos de interés y obtener los posibles ARNs no codificantes, es necesario evaluar cuales de estos transcritos tienen las características necesarias para llegar a ser traducidos a un proteína, con el fin de eliminarlos de la lista de transcritos de interés y obtener una lista de ARNs no codificantes.

Un método común para identificar transcritos con potencial codificante es el uso de herramientas como la calculadora de potencial codificante (CPC) que analiza 6 diferentes propiedades biológicas de las secuencias y las clasifica a través de una máquina soporte vector en codificantes y no codificantes; sin embargo, estudios previos sugieren el uso de al menos dos herramientas para identificar transcritos codificantes a proteínas, CPAT es otra herramienta comúnmente utilizada pero se encuentra enfocado en organismos modelos y no incluye plantas, por lo cual fue descartado. El software Transdecoder permite identificar posibles regiones codificantes (marcos de lectura abiertos, ORFs) en los transcritos ensamblados y toma en cuenta el tamaño medio del marco de lectura abierto (ORF); además, es posible utilizar la información de alineamientos de los ORFs identificados a bases de datos como Pfam para identificar la presencia de dominios de

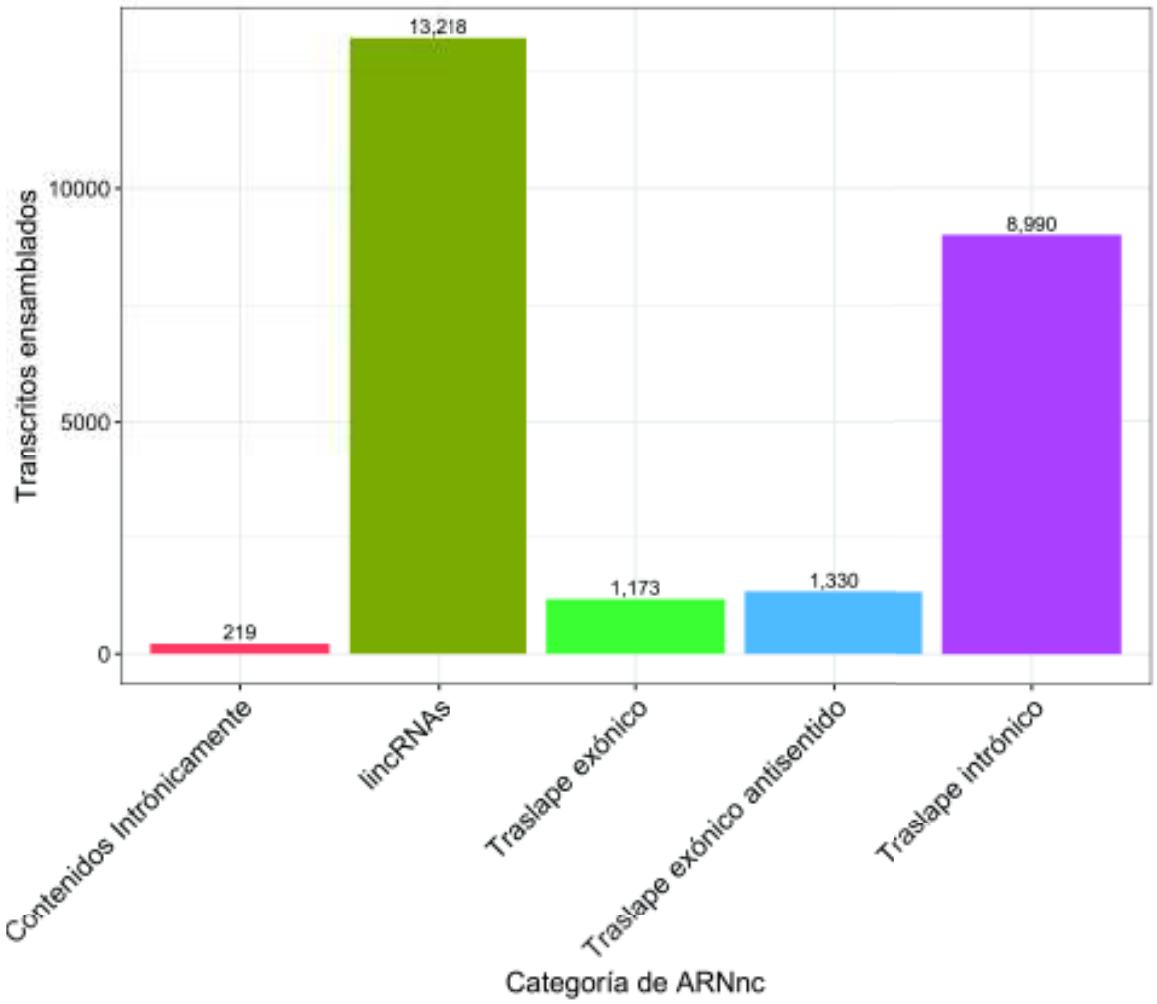


Figura 30. ARNs no codificantes largos potenciales clasificados de acuerdo a la posición en el genoma respecto a los genes codificantes de proteínas.

proteínas y desechar marcos de lectura abiertos no significativos.

La figura 31 muestra el resultado de la calculadora de potencial codificante (CPC) y transdecoder en los transcritos filtrados por código de clase (*i,o,u,j* y *x*), 51,464 transcritos fueron etiquetados como codificantes por ambas herramientas, 555 transcritos fueron identificados como potenciales codificantes a proteínas por CPC solamente; mientras que, Transdecoder con los alineamientos a las bases de datos Pfam y UniProtKB/Swiss-Prot identificó un total de 4,570 transcritos codificantes que no fueron identificados por CPC. Finalmente, se eliminaron 56,589 transcritos etiquetados como codificantes, lo que resultó en la identificación de 9,817 transcritos no codificantes que fue utilizada en los análisis posteriores.

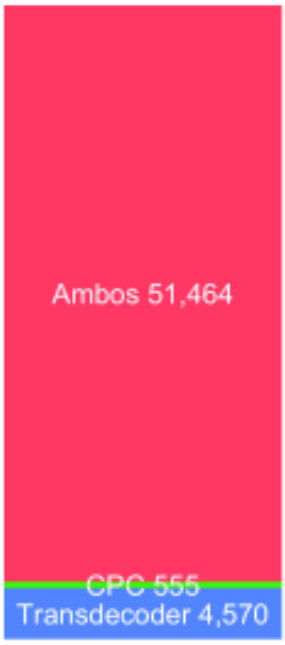


Figura 31. Cantidad de transcritos con potencial codificante detectados por las herramientas CPC2, Transdecoder y por ambas herramientas.

Una vez filtrados los transcritos codificantes, se realizó nuevamente la clasificación de los transcritos no codificantes a través de su código de clase asignado en función de su posición en el genoma, el resultado se muestra en la figura 32, donde se puede observar que los ARNs

no codificantes largos intergénicos (lincRNAs) se mantienen como la principal categoría con 7,861 (60.28 %), la categoría de transcritos contenidos intrónicamente perdió la menor cantidad de elementos y ahora contiene 172 (de 219) transcritos; mientras que la categoría de traslape intrónico fue la que tuvo una mayor reducción, ya que ahora contiene solamente 728 transcritos (5.58 %), de manera similar las categorías de traslape exónico con 192 (1.47 %) transcritos y traslape exónico antisentido con 864 (6.62 %) transcritos, la cantidad de transcritos codificantes detectados en estas últimas tres categorías pueden estar relacionados con la naturaleza de su composición, ya que los transcritos en estas tres categorías están compuestos de fragmentos de intrones o exones y podrían corresponder a isoformas no anotadas en el genoma de referencia actual.

V.10.3. Anotación de ARNs no codificantes de cadena larga de Chiltepín

Rfam es una base de datos de perfiles de ARNs no codificantes creados a partir del alineamiento múltiple de ARNnc de diversas especies, cada uno de ellos contiene una serie de patrones o motivos conservados y puede ser utilizada con herramientas como Infernal para realiza la anotación de ARNs no codificantes en función de dichos perfiles.

En este análisis se utilizó la lista de ARNs no codificantes confirmados por Transdecoder y CPC y se realizaron alineamientos a la base de datos de RFam con la herramienta Infernal, posteriormente se filtraron para eliminar alineamientos de baja calidad y obtener una lista de 178 transcritos anotados como ARN no codificantes, que fueron agrupados de acuerdo al término de su anotación en 19 diferentes grupos. La figura 33 muestra la cantidad de transcritos anotado por cada término anotado; la mayor cantidad de transcritos (162) fueron anotados como ARN ribosomales (rRNA), este tipo de ARN representa la mayor cantidad de ARN en cualquier muestra aislada (de 80 % a 90 %) y, aunque se utilizan métodos para eliminar el rRNA en los análisis de RNA-Seq, es posible encontrar pequeñas porciones de ellos. Como resultado de la anotación, en-

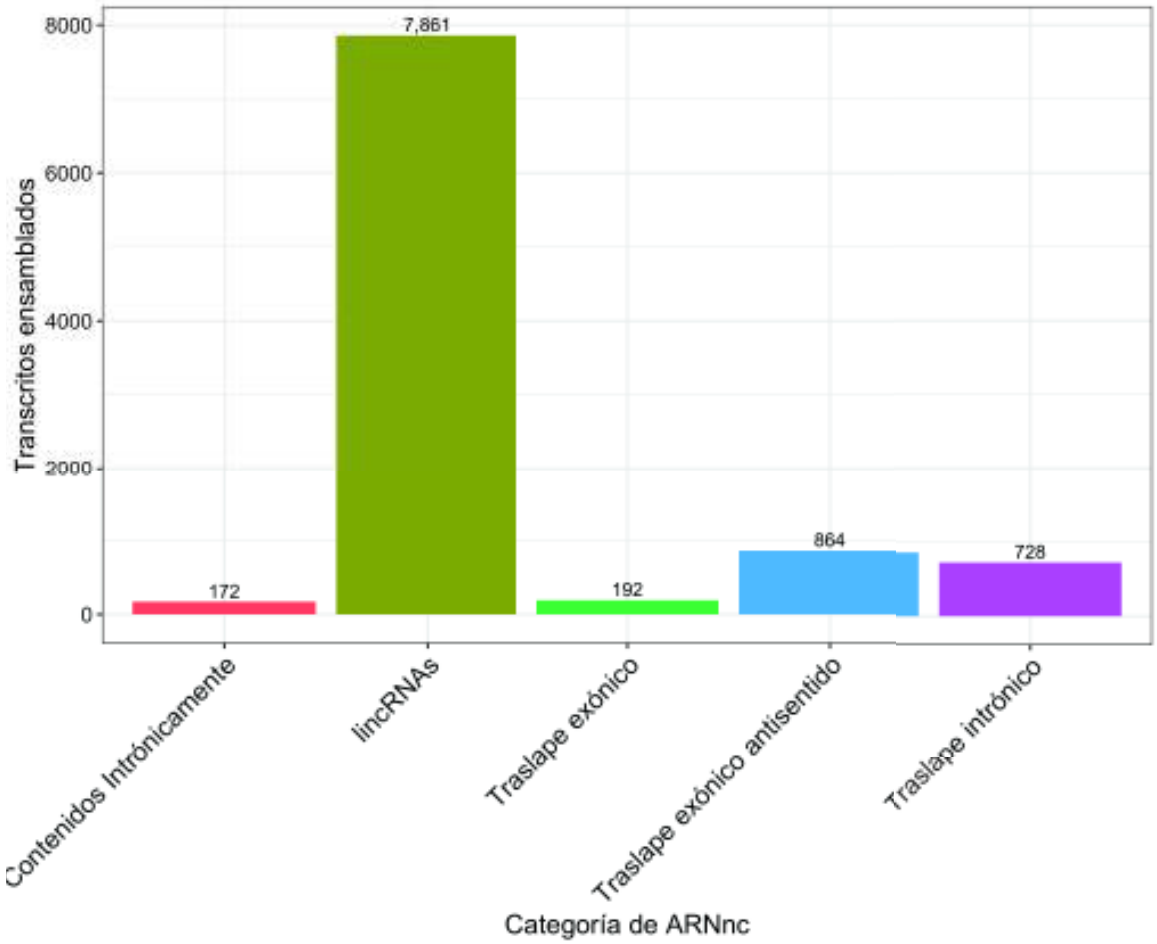


Figura 32. Clasificación de ARNs no codificantes de cadena larga de acuerdo a su posición en el genoma respecto a los genes codificantes de proteínas.

contramos 7 transcritos como precursores de micro ARNs, 6 transcritos como ARNs nucleolares pequeños (snoRNA) relacionados con la modificación química de los ARN ribosomales, incluyendo un ARN pequeño específico al cuerpo cajal, esta clase de snoARN guían la modificación de los ARNs del spliceosoma U1, U2, U4, U5 y U12. Por último, se identificaron 2 transcritos anotados como KCNQ1OT1 (KCNQ1 overlapping transcript 1), y que corresponden a un ARN no codificante largo que regula la transcripción de múltiples genes a través de modificaciones epigenéticas de la cromatina.

V.10.4. Identificación de plantillas para miRNAs

Existen diferentes enfoques *In silico* para la detección de miRNAs, los métodos tradicionales consisten en la identificación de estructuras secundarias típicas como horquillas (*hairpins*) y regiones conservadas para predecir regiones del genoma del organismo a partir de las cuales se pueden producir miRNAs. Otra forma de identificar miRNAs consiste en utilizar algoritmos aprendizaje máquina capaces de extraer características de miRNAs conocidos y adaptarse para descubrir nuevos, algoritmos de clasificación como máquina de soporte vector (SVM), o Hidden Markov Model (HMM) o clasificadores Naïve Bayesianos (NBC) son implementados en herramientas que utilizan el genoma del organismo para predecir miRNAs conocidos y nuevos.

Los miRNAs pueden surgir a partir de ARN transcritos y a partir de intrones eliminados durante el proceso de splicing alternativo, por lo cual es posible identificar transcritos que serán utilizados como plantillas para la producción de miRNAs, para realizar esta tarea se utilizó el método mejor alineamiento bi-direccional (BRH) con la base de datos MiRBase que contiene secuencias de 38,589 micro ARNs con blastn con la opción *-task blastn-short* dado que los miRNAs poseen secuencias de tamaño muy pequeño.

Se identificaron 97 transcritos primarios (pri-miARNs) que pueden ser utilizados como plantillas para la generación de 18 micro ARNs diferentes, la figura 34 muestra la cantidad de

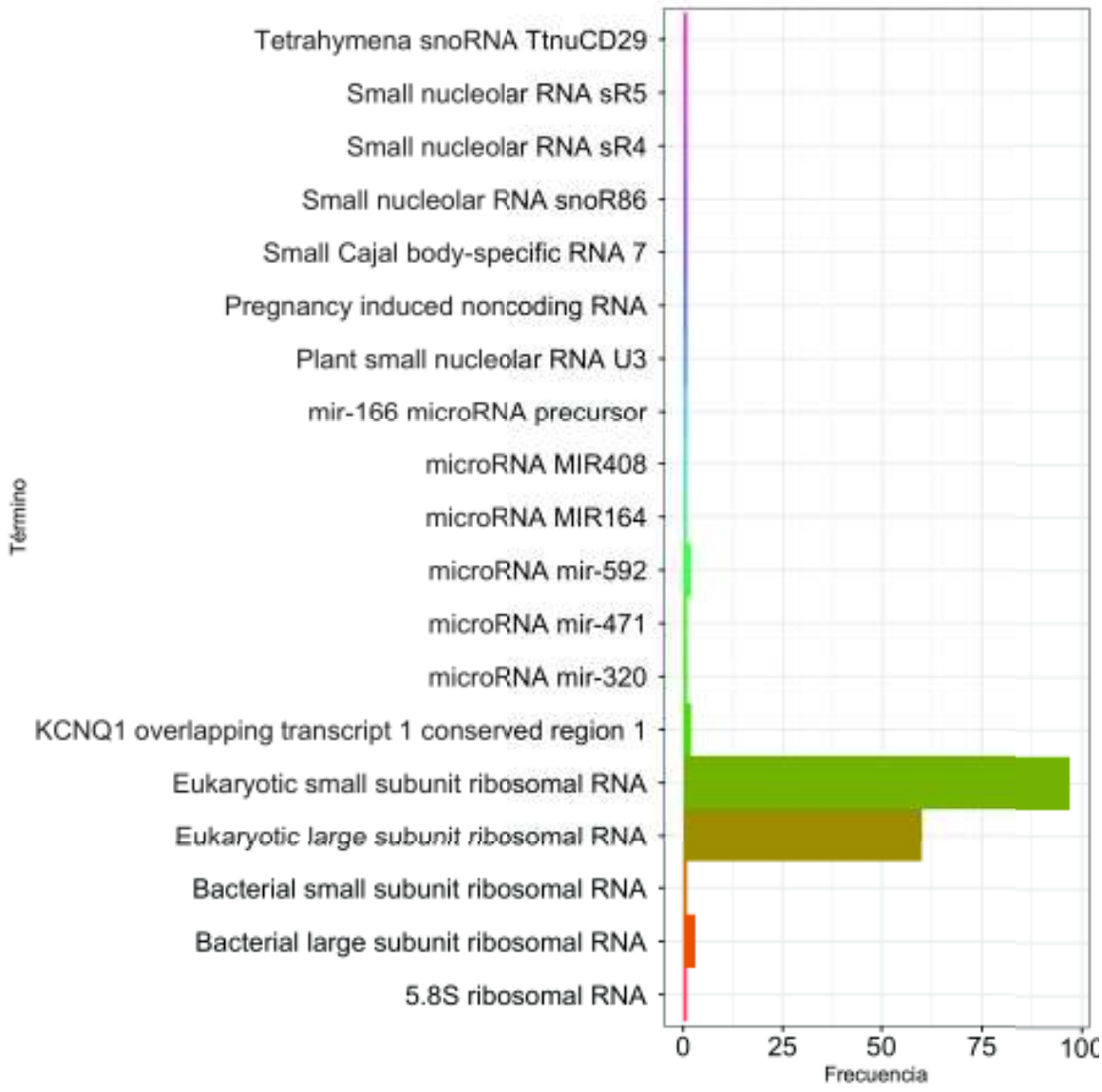


Figura 33. ARN no codificantes de Chiltepín anotados con Infernal utilizando la base de datos Rfam, en el eje Y se encuentran los términos o anotaciones asignadas y en el eje X se muestra la cantidad de transcritos anotados en cada uno de los términos.

transcritos que pueden servir de plantilla para la producción de los 18 miRNAs detectados; la mayor cantidad de transcritos-plantilla corresponden al MiRNA mir2916 (51), seguido de miR5303c con 10 transcritos-plantilla, los MiRNAs miR5303b y miR5303i con 7 transcritos-plantilla identificados, miR7997b y miR5303e con 3 transcritos-plantilla cada uno, miR162a, miR8026, miR5368 y miR5303a con 2 transcritos-plantilla y por último los MiRNAs miR5303f, miR159, miR5523, miR-4426, miR166c, miR408, miR391 y miR5141 con solo un transcrito-plantilla identificado (figura 34).

La anotación de ARNs no codificantes con Infernal utilizó un enfoque distinto al utilizado para la detección de MiRNAs por similitud de secuencia, Infernal utiliza una búsqueda a través de modelos ocultos de Markov creados en función de regiones de ARNnc conservadas en varias especies; a través de la anotación creada por Infernal se identificaron 7 transcritos-plantilla de 6 MiRNAs diferentes: MIR164, mir-166, mir-320, MIR408, mir-471 y mir-592; de los cuales se encontraron 2 plantillas para mir-592 y un transcrito-plantilla para el resto, solamente el MiRNA miR408 fue identificado por ambos enfoques, se obtuvieron un total de 103 transcritos que pueden ser utilizados para la producción de 23 MiRNAs diferentes (figura 35).

V.10.5. Transcritos plantilla de MiRNAs expresados en Chiltepín

Una vez identificados los transcritos que pueden ser utilizados como plantillas para la producción de micro ARNs, se realizó un filtrado de los transcritos que se encuentran expresados en Chiltepín, para esto se tomaron en cuenta solamente los transcritos con una expresión FPKM >= 0.1 en al menos un grupo de muestras. La figura 36 muestra la cantidad de transcritos por MiRNA expresados en las muestras de Chiltepín, se encontraron un total de 89 transcritos plantilla de 20 MiRNAs diferentes; la mayor cantidad de transcritos (49) corresponden al miRNA miR2916 que de acuerdo a estudios previos han mostrado que este miRNA se ha encontrado expresado en diferentes tejidos de un considerable número de especies de plantas y podría estar

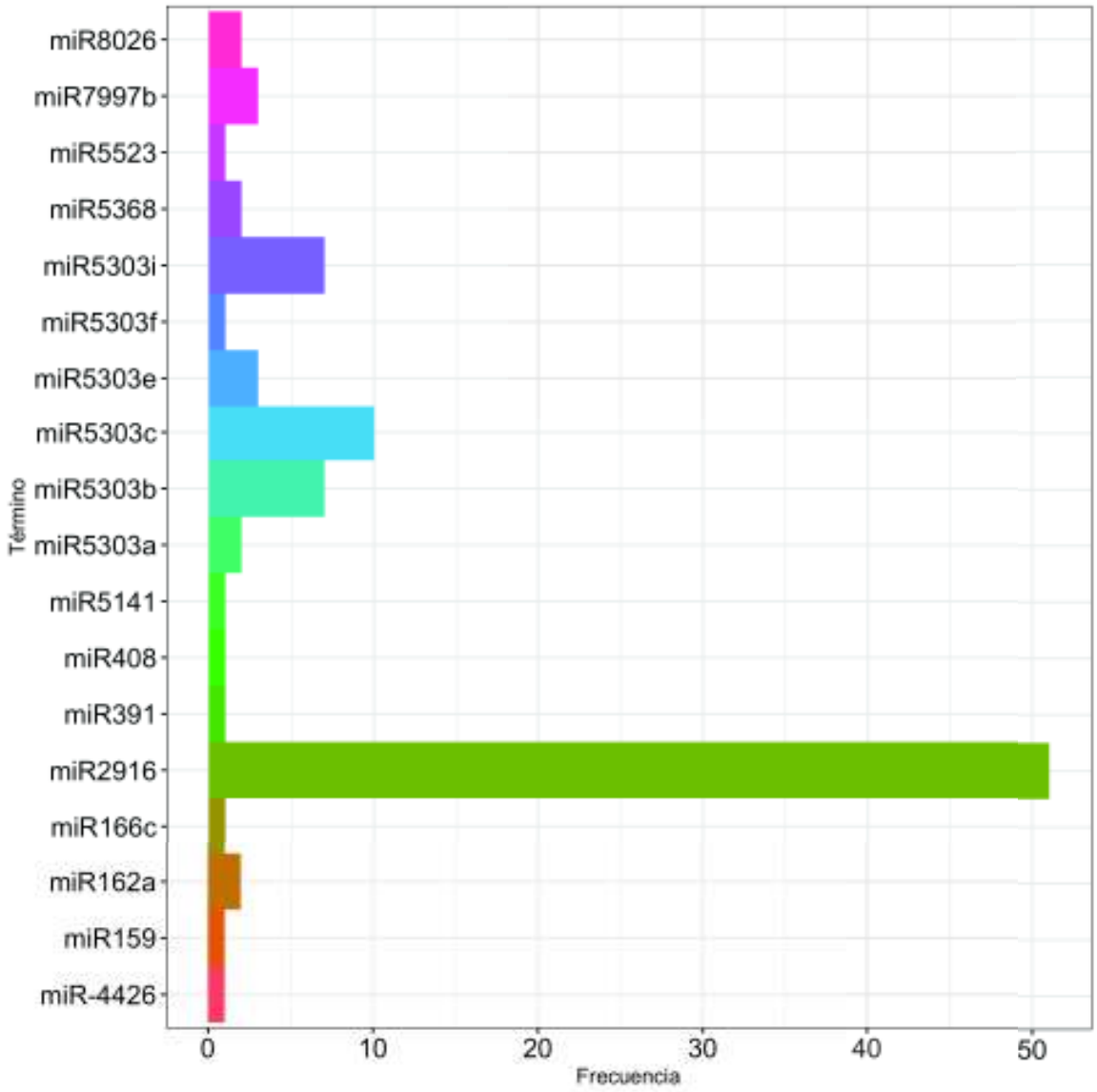


Figura 34. Cantidad de transcritos potenciales plantillas para la producción de los 18 miRNAs identificados a través de MiRBase.

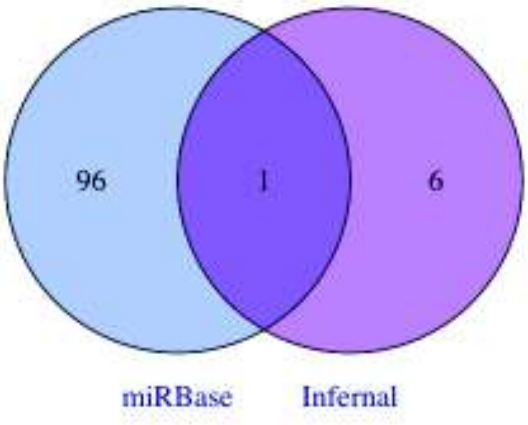


Figura 35

relacionado con procesos biológicos importantes que aún no han sido identificados; sin embargo, se ha mostrado que genes objetivos de este miRNA están relacionados con la respuesta a estrés abiótico (Davoodi Mastakani *et al.*, 2018). Los siguientes 6 MiRNAs con mayor número de transcritos-plantilla (23 en total) pertenecen a la familia miR5303 que en estudios previos se ha mostrado que solo se encuentran en la familia de las *solanaceas*, ésta podría ser una de las familias de MiRNAs con mayor número de miembros y sus objetivos potenciales incluyen enzimas metabólicas (dioxigenasas de ruptura de carotenoides, aciltransferasas, y aldehido oxidasas), proteinas de respuesta a estrés abiótico y proteínas no caracterizadas (Gu *et al.*, 2014). Se encontraron 3 transcritos-plantillas para el MiRNA miR7997b que en tomate se ha mostrado que están relacionados con la ruta metabólica de transducción de señales de hormonas en plantas y sus respectivas proteínas (Candar-Cakir *et al.*, 2016). Se identificaron 2 transcritos-plantilla del MiRNA miR5368 que tiene como objetivo la enzima glicósido hidrolasa que cataliza la hidrólisis de azúcares complejos y está relacionada con la degradación de biomasa y algunas estrategias de defensa contra patógenos (Liu *et al.*, 2018). El MiRNA miR162a con 2 transcritos-plantilla

ha sido identificado como un regulador esencial de la inmunidad contra *Magnaporthe oryzae* , y de la producción de grano en arroz (Li *et al.*, 2020).

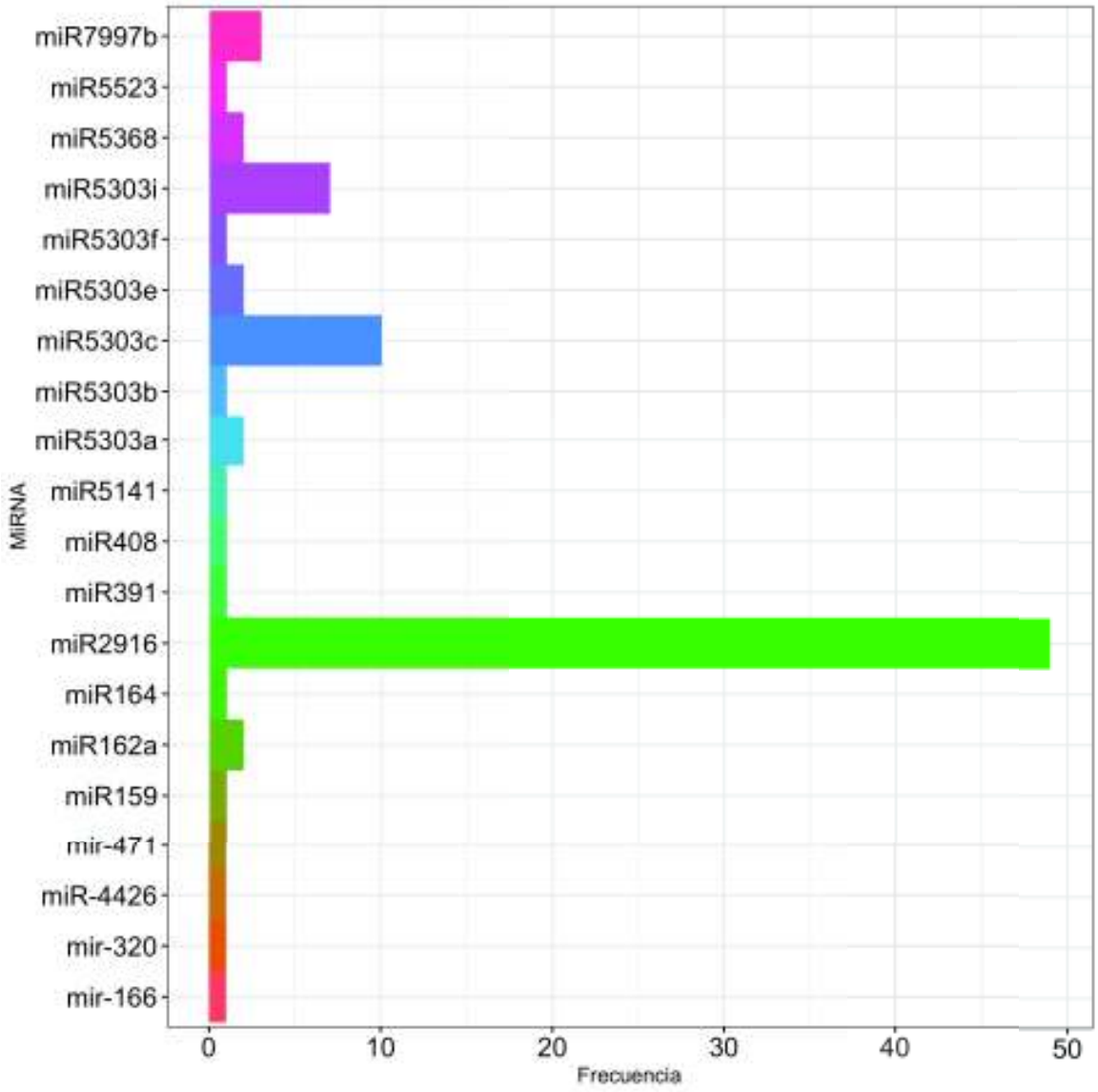


Figura 36. Cantidad de transcritos potenciales plantillas para la producción de los 18 miRNAs identificados a través de MiRBase y que se encuentran expresados en Chiltepín.

V.11. Investigación adicional

Durante el desarrollo de este trabajo de investigación se realizó una estancia en el laboratorio de Biología Computacional de la Unidad de Genómica Avanzada (UGA-Langebio), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav) en la ciudad de Irapuato, México, en la cual se trabajó en dos proyectos adicionales de los cuales se obtuvieron dos artículos científicos y un software para el procesamiento de datos transcriptómicos.

V.11.1. Transcriptome Analyses throughout Chili Pepper Fruit Development Reveal Novel Insights into the Domestication Process

El objetivo de este proyecto consistió en el análisis de datos transcriptómicos y metabolómicos de frutos de 10 accesiones de chiles (4 silvestres y 6 domesticados) en 7 etapas de desarrollo y maduración de fruto (0, 10, 20, 30, 40, 50, and 60 DDA) (Martínez *et al.*, 2021). Durante los 6 meses de estancia se realizaron cruzas entre variedades de chile silvestre y cultivado ('criollo de Morelos 334' y chiltepín de Querétaro y Sonora), esterilización y germinación de semillas, cultivo y cuidado de plantas de chile en invernadero, así como el etiquetado de flores y recolección de frutos para la creación de más de 170 librerías de secuenciación.

Las muestras de ARN de frutos colectados fueron secuenciadas y se realizó el filtrado de las lecturas de secuenciación, alineamiento al genoma de referencia de chile (Criollo de Morelos CM334) y la cuantificación de los genes expresados. Posteriormente, se llevó a cabo un análisis de expresión diferencial, para cada gen en cada accesión se calculó la expresión media estandarizada y el conjunto de estos en las 7 etapas de maduración en cada accesión conformó el perfil de expresión estandarizada.

Se utilizaron los perfiles de expresión en lugar de la expresión de genes individuales para analizar la dinámica compleja del transcriptoma a través de las etapas de maduración muestrea-

das. Se encontró que las accesiones domesticadas y silvestres tienen un perfil de expresión media diferente, y pueden ser agrupadas en un dendograma calculando una matriz de distancias entre los perfiles de cada uno de los genes expresados. Además, las accesiones domesticadas presentaban un pico de expresión a los 10 DDA, mientras que las accesiones silvestres lo presentaban a los 20 DAA, en estas etapas se identificaron 543 genes con diferente patron de expresión entre los dos tipos de accesiones, y se encontró que existe una mayor diversidad de expresión génica en las accesiones silvestres que en las cultivadas.

Adicionalmente, 463 genes (2.06 % de los genes expresados) mostraron diferencias de expresión significativas entre las accesiones silvestres y domesticadas, estos se encontraron relacionados con procesos de regulación celular, localización, movilidad y ensamblado, autofagia y organización de organelos. Además, se encontraron diferencias en el tiempo y los niveles de expresión de genes relacionados con la reproducción celular que podrían explicar las diferencias en tamaño de fruto entre las accesiones silvestres y domesticadas.

Los resultados de este proyecto de investigación coinciden con los obtenidos en este trabajo de tesis, ya que han mostrado que 96.6 % de los genes expresados no muestran diferencias significativas entre las accesiones silvestres y domesticadas, sugiriendo que esta parte del transcriptoma no fue modificado por el proceso de domesticación, y al rededor del 2 % de los genes presenta las principales diferencias entre las accesiones silvestres y domesticadas. En nuestra investigación encontramos que la diferencias de expresión entre chiltepín y serrano durante el proceso de maduración de fruto son ligeras y en su mayoría se encuentran específicamente en genes relacionados con las características deseables en un fruto comercial.

En este artículo además se concluye que las diferencias de tamaño entre los frutos de las accesiones silvestres y domesticadas se deben a una expresión temprana y un mayor nivel de expresión de genes relacionados con la división celular, mientras que en este proyecto de tesis se encontraron genes que limitan la división celular expresados en la etapa inmadura de Chilte-

pín, mientras que en Serrano se encontraron expresados genes relacionados con el crecimiento celular por vía ácida.

V.11.2. Compacta: a fast contig clustering tool for de novo assembled transcriptomes

Inicialmente este proyecto de investigación se realizaría utilizando un transcriptoma ensamblado *de novo* de chiltepín conformado por 383,476 transcritos, de los cuales solo el 48 % habían sido identificados a través del método mejor alineamiento bi-direccional (BRH). El 52 % de los transcritos restantes probablemente eran isoformas de los genes identificados, por lo cual se procedió a realizar un agrupamiento de transcritos utilizando las herramientas: CD-HIT, que se basa en la similitud de las secuencias para agruparlas, Grouper y Corset, que utiliza los alineamientos de las lecturas de secuenciación a los transcritos como prueba de su similitud. Debido al tamaño y complejidad del transcriptoma de chiltepín, ninguna de las herramientas utilizadas produjo resultados adecuados, CD-Hit redujo el tamaño del transcriptoma en menos del 10 % y Corset se ejecutó ininterrumpidamente por 32 días sin llegar a un resultado.

CD-Hit utiliza similitud de secuencias, por lo cual depende de parámetros que deben ser ajustados manualmente, si se utilizan parámetros muy estrictos el agrupamiento se limita a secuencias idénticas, mientras que con parámetros laxos el algoritmo tiende a agrupar transcritos que comparten fragmentos de secuencias. Por su parte, Corset presenta tres desventajas principales: utiliza un número de lecturas fijas para evaluar la cobertura de los contigs, ignorando el tamaño de éste, además, depende de una prueba de razón de similitud para separar contigs que comparten lecturas de secuenciación y podrían ser producto de dos genes diferentes, por último, la implementación de Corset utiliza vectores para almacenar millones de datos ordenados, y estos vectores deben ser reordenados en cada iteración del algoritmo, lo cual representa una complejidad computacional de $\mathcal{O}(n^3)$, por lo cual, este algoritmo tiende a volverse extremadamente lento cuando se utiliza una cantidad de datos de entrada de tamaño considerable,

como suelen ser los de experimentos de RNA-Seq. Grouper por su parte, reduce el tiempo de lectura de datos de entrada al utilizar archivos de equivalencias obtenidos de herramientas de pseudo-alineamiento, sin embargo, utiliza una función de corte mínimo Stoer–Wagner que se ejecuta en cada iteración y tiene una complejidad computacional $\mathcal{O}(mn + n^2 \log n)$, por lo cual, cuando se utilizan conjuntos de datos pequeños tiene un tiempo de ejecución corto pero aumenta considerablemente con el tamaño de los datos de entrada.

Durante los 6 meses de estancia se discutió y diseñó el funcionamiento de una herramienta que permitiera reducir la complejidad del transcriptoma a través del agrupamiento de genes, de esta manera se desarrolló Compacta, una herramienta bioinformática que utiliza los alineamiento de las lecturas de RNA-Seq al transcriptoma en forma de archivos BAM, filtra contigs con una baja cobertura considerando el tamaño de contig. Posteriormente, crea grafos (pre-clusters) de contigs conectados por lecturas de secuenciación compartidas, donde los contigs corresponden a los vértices y las lecturas compartidas entre dos contigs crean aristas entre sus respectivos vértices con una distancia que se calcula en función de la razón entre el número de reads compartidos y la cantidad de reads en cada vértice. A continuación, Compacta utiliza una estructura de árbol heap auto-ordenado para ordenar las aristas por su distancia de forma ascendente, y en cada iteración evalúa una arista y decide si los vértices correspondientes pertenecen a un mismo grupo. Una vez que Compacta finaliza el agrupamiento de los pre-clusters, crea un archivo que relaciona a cada contig con su correspondiente grupo, y un archivo con un conjunto de contigs representativos de los grupos generados que puede ser utilizado en análisis posteriores.

Cualitativamente, Compacta produce resultados con una precisión y *recall* similares a Grouper y Corset; sin embargo, la implementación de una estructura de árbol heap auto-ordenable reduce la complejidad del algoritmo hasta un $\mathcal{O}(n^2 \log n)$, además, Compacta puede ejecutar el agrupamiento de los pre-clusters en paralelo, lo cual reduce significativamente el tiempo de ejecución del algoritmo, de forma conservada podemos decir que Compacta es al menos 10

veces más rápido que Grouper y Corset para agrupar transcriptomas complejos de organismos eucariotas.

Se realizó una publicación donde se comparan los resultados de Compacta y las herramientas de agrupamiento actuales (Razo-Mendivil *et al.*, 2020), el código y ejecutable de Compacta se encuentra disponible como herramienta de software libre bajo licencia GNU 3.0 en http://doi.org/10.5281/zenodo.3469484.

VI. DISCUSIÓN

La maduración de frutos carnosos es un proceso de desarrollo en el cual el fruto atraviesa por varios cambios metabólicos y bioquímicos que producen las propiedades organolépticas que ayudan a la dispersión de semillas, adicionalmente, estos compuestos proveen los componentes nutricionales esenciales para la dieta humana (Aivalakis y Katinakis, 2008). Al contrario del tomate, su pariente cercano, que ha sido utilizado como modelo de desarrollo y maduración de frutos carnosos climatéricos (Dong *et al.*, 2014), el chile es un fruto no-climatérico cuyo proceso de maduración de fruto no ha sido totalmente entendido, al igual que en otros frutos no climatéricos el etileno no tiene efecto desencadenante de la maduración, estudios previos han mostrado que la maduración no climatérica de frutos de chile podría ser regulada por las fitohormonas etileno y ácido abscísico (ABA) (Hou *et al.*, 2018; Paul *et al.*, 2012).

En este estudio se analizaron datos de RNASeq de una variedad cultivada y una silvestre de chile durante el desarrollo y maduración de fruto, con el propósito de identificar diferencias de expresión génica producidas por el proceso de domesticación de chile. Nuestro resultados mostraron que ambas variedades presentan pocas diferencias en los patrones de expresión, sin embargo, estas diferencias se presentaron principalmente en genes clave relacionados con el proceso de desarrollo y maduración de fruto. Con base en los resultados obtenidos se creó un modelo biológico hipotético para la maduración de fruto de chile Serrano y Chiltepín 37.

El análisis de enriquecimiento de términos GO mostró que no se encontraron términos enriquecidos para los contrastes entre las dos variedades comparadas en la misma etapa de maduración (Ch20-St20 y Ch68-St60), mientras que en los contrastes Ch20-Ch68 y St20-St60 la

mayoría de los términos enriquecidos se encontraban relacionados con la respuesta a estrés, patógenos y luz, lo cual tiene sentido dado que es conocido que la maduración de fruto es regulada por las mismas cascadas de hormonas que son utilizadas para la respuesta como defensa y es muy complicado separar ambos procesos. El análisis de rutas metabólicas enriquecidas (KEGG) mostró que no existen rutas metabólicas enriquecidas en los contrastes Ch20-St20 y Ch68-St60; sin embargo, en el contraste St20-St60 la ruta metabólica *transducción de señal de hormonas* se encontró enriquecida y en ella se encontraron genes expresados diferencialmente que funcionan como reguladores de la biosíntesis de etileno (ERF1), ácido abscísico (PP2C, SnRK2) y auxina (IAA, SAUR50). El gen PP2C se encuentra inducido y es un regulador negativo de la producción de ABA y el gen SnRK2 se encuentra reprimido y es un regulador positivo de ABA en la cascada de señalización 'PYR-PP2C-SnRK2', estudios previos han mostrado que en el proceso de maduración la expresión de SnRK2 de disminuye mientras los niveles de expresión de PP2C aumentan (Hou *et al.*, 2018), lo cual es consistente con nuestros resultados.

Comercialmente, el tamaño de fruto es una de las características más importantes en el chile, en este sentido, en Serrano se encontraron expresados diferencialmente los genes PP2C y SAUR50 de la ruta metabólica de la auxina, estudios previos han mostrado que el gen SAUR50 reduce la actividad de las proteínas fosfatasas clado-D tipo 2C (PP2C-Ds), responsables de la fosforilación del segundo aminoácido de la H+-ATPasas (PM H+-ATPasas) en la membrana plasmática, la reducción en la actividad de estas proteínas fosforilasas activan la elongación celular por H+-ATPasas a través de un mecanismo de crecimiento ácido (Spartz *et al.*, 2014; van Mourik *et al.*, 2017; Stortenbeker y Bemer, 2018), esto ha sido probado en varios tipos de órganos pero no en frutos y podría ser responsable por la diferencia en tamaño en los frutos de chile Serrano, técnicamente este conjunto de genes podrían trabajar de forma simultánea desde las etapas de expansión celular, pero, es imposible saberlo dado que los puntos de recolección de muestras de este trabajo se encuentran en los extremos de dicha etapa. El gen SAUR50 es

regulado por el gen FRUITFUL (FUL2), expresado en Serrano pero no en Chiltepín a 68 DDA, estudios previos han mostrado que el gen FUL2 por sí mismo puede producir un incremento en el tamaño de los frutos (Slugina *et al.*, 2018), reduce la deshidratación post-recolección y regula los genes claves en la biosíntesis de etileno ACO1 y ACS2 (Wang *et al.*, 2014), en este trabajo se encontró que una copia del gen ACO3 presenta el mismo patrón de expresión que el gen FUL2 y se encuentra en el mismo grupo en el mapa de calor de la figura 27, el gen FUL2 es uno de los genes reguladores de la cascada de señalización de etileno durante el proceso de maduración de fruto, mientras que el gen ACO3 está relacionado con la respuesta y biosíntesis del etileno (Naing *et al.*, 2019). EL patrón de expresión de los genes ACO3 y FUL2 sugiere que estos podrían estar relacionados con la maduración temprana de los frutos de chile Serrano, ya que la maduración de fruto de chile no es iniciada por etileno pero éste si posee un papel importante en la regulación de los procesos involucrados.

La forma de fruto es otra característica con claras diferencias entre los dos frutos analizados, el Chiltepín de la región de Sonora presenta una forma redondeada y pequeña, mientras que el chile Serrano 'Tampiqueño 74' presenta una forma grande y alargada, en nuestros resultados se encontró un ortólogo de un gen OVATE, relacionado con la forma del fruto, expresado a los 20 DDA en ambas variedades, aunque con menor nivel de expresión (debajo de las 4 veces) en Serrano. En chile esta proteínas del tipo OVATE ha sido asociada con las diferencias en la forma de fruto (Tsaballa *et al.*, 2011), esta proteína es un regulador negativo del crecimiento celular a través de la supresión de la biosíntesis de GA y su sobreexpresión produce frutos enanos (Hackbusch *et al.*, 2005). La represión de la expresión del gen OVATE en chile convierte frutos redondos en formas oblongas o alargada (Tsaballa *et al.*, 2011). Interesantemente, nuestro resultados muestran que el gen OVATE se agrupa con un ortólogo del gen FAS de tomate, este es un homólogo del gen CLAVATA3 (CLV3) que en tomate contiene una mutación que produce frutos más grandes con una forma fasciada; sin embargo, la función original del gen CLV3

consistía en restringir el tamaño de la población de células madres, por lo cual, en Chiltepín el conjunto OVATE-FAS/CLV3 podría ser responsable de la forma redondeada y el tamaño pequeño de los frutos.

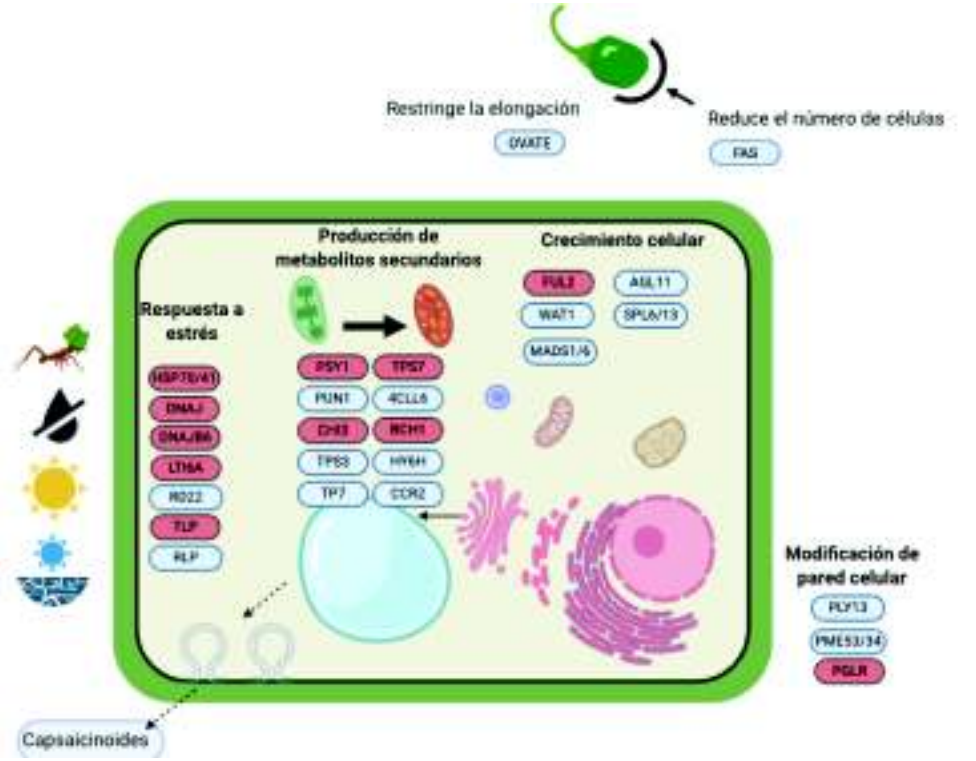
Los metabolitos secundarios son compuestos que no son necesarios para los procesos básicos de las plantas pero son útiles para la interacción con el medio ambiente que les rodea, usualmente como una defensa contra diferentes tipos de estrés. Hay un considerable número de metabolitos secundarios sintetizados a partir de metabolitos primarios (aminoácidos, lípidos y carbohidratos) incluyendo: terpenos, compuestos fenólicos y compuestos a base de nitrógeno y sulfuro; adicionalmente, los metabolitos secundarios contribuyen a las características organolépticas del fruto de chile (olor, sabor, color y pungencia) y son una buena fuente de elementos nutritivos esenciales para la dieta humana (Mazid *et al.*, 2011; Akula y Ravishankar, 2011). En este estudio se encontraron diferencias significativas en los patrones de expresión de genes relacionados con la biosíntesis de metabolismo secundarios en las dos variedades de chile. Usualmente, la mayoría de los genes involucrados en el metabolismo secundario tienen un pico de expresión en las etapas de desarrollo y dejan de expresarse a lo largo de la etapa de maduración de fruto, una vez que los pigmentos, compuestos volátiles, capsaicinoides, etc, han sido almacenados y el fruto ha obtenido las características que le harán atractivo para la dispersión de semillas, encontramos un patrón de expresión muy similar en ambas variedades; sin embargo, 3 genes fueron expresados diferencialmente solo en Chiltepín: Chalcona isomerasa (CHI3) involucrada en la biosíntesis de antocianina (reprimida en el contraste Ch20-St20 y Ch20-Ch68); el gen (E)-β-ocimeno sintasa (reprimido en el contraste Ch68-St60), encargado de la producción de ocimeno, un monoterpeno que está asociado a la respuesta a ataques de herbivorísmo (Tohge *et al.*, 2005) y el gen β-caroteno hidroxilasa-2 (BCH2, inducido en el contraste Ch20-Ch68), una enzima responsable de la biosíntesis de xantófilas (Bouvier *et al.*, 1998b). Estos resultados sugieren que podrían existir diferencias en la concentración de metabolitos secundarios en los

frutos de ambas variedades de chile. De acuerdo con esto, estudios previos han mostrado que un aumento en la expresión del gen Chalcona isomerasa en tomate produce un incremento de hasta 78 veces en la concentración flavonoles en el fruto (Muir *et al.*, 2001).

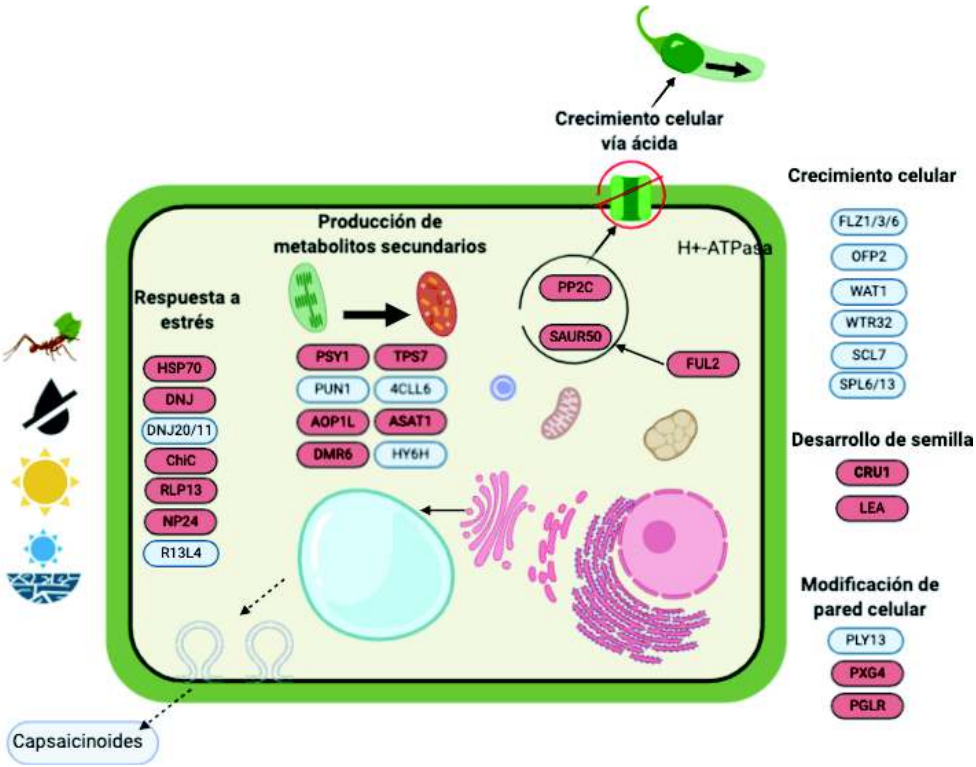
Durante el proceso de domesticación se llevó a cabo una presión selectiva para obtener las características de interés en un fruto comercial, como tamaño, sabor, color, olor, etc., produciendo regiones de barrido selectivo. Estudios previos han identificado 115 de estas regiones en el cultivar de chile 'zunla-1'. En esta investigación hemos encontrado 19 genes localizados en regiones de barrido selectivo expresados diferencialmente en los contrastes de Serrano y Chiltepín, 6 de estos genes no se encuentran anotados, la mayoría de los DEGs en Chiltepín están relacionados con la resistencia a enfermedades, mientras que en Serrano se encontró solamente un gen de respuesta a enfermedades y 3 genes asociados a elongación de transcripción, modificación de la conformación de la cromatina y un factor de transcripción de respuesta a etileno relacionado con el desarrollo de la planta y la respuesta a estrés.

Los ARNs no codificantes son un grupo de ARNs que han mostrado capacidad regulatoria, ya sea a través del silenciamiento, inducción o represión o a través de modificaciones epigenéticas; sin embargo, en la mayoría de los casos los ARN no codificantes se encuentrar poco estudiados, esto se ve reflejado en el resultado de la anotación de ARNnc realizada en este estudio, donde solamente una pequeña porción de los transcritos clasificados como ARNnc fueron anotados y en la mayoría de los casos se trataban de ARNnc de *housekeeping* como componentes de los ribosomas, modificadores de ARNr y elementos de spliceosomas. Por otro lado, en el análisis de micro ARNs se identificaron 103 transcritos que pueden ser utilizados como plantillas para la producción de micro ARNs expresados en Chiltepín, la mayoría de estos corresponden a miRNAs con objetivos relacionados con la respuesta a estrés, pero se encontraron micro ARNs de interés de acuerdo al enfoque de este estudio; por ejemplo, se encontró un grupo de miRNAs que es exclusivo de la familia de las *sonalaceas* y que se encuentran relacionados

con la regulación de genes de rutas metabólicas importantes en el proceso de maduración de fruto, incluyendo la producción de carotenoides, capsaicinoides y de respuesta a estrés abiótico; también se identificaron miARNs reguladores de genes involucrados en la pérdida de biomasa por la degradación de azúcares complejos. Desafortunadamente, con las técnicas y bases de datos actuales no fue posible anotar 9,542 transcritos no codificantes de modo que por el momento no es posible saber cual es su rol en el proceso de maduración de fruto.



(a)



(b)

Figura 37. Modelo biológico hipotético propuesto para la maduración de fruto en A) Chiltepín (Ch20-Ch68) y B) Serrano (St20-St60). Recuadros rojos indican genes inducidos (con expresión al menos 4 veces mayor en la etapa madura), recuadros azules indican genes reprimidos (con expresión al menos 4 veces mayor en la etapa de desarrollo).

VII. CONCLUSIONES

- La diferencia en la profundidad de secuenciación de las librerías de Chiltepín y Serrano impactó significativamente la cantidad de genes expresados en cada una de ellas. Por lo cual, las muestras de Chiltepín presentaron un mayor número de genes expresados y el contraste Ch20-Ch68 presentó la mayor cantidad de genes expresados diferencialmente (DEGs).

- La mayoría de los genes relacionados con el proceso de desarrollo y maduración de fruto presentan una disminución en sus niveles de expresión al acercarse a la etapa de fruto maduro. Por lo cual, la mayor cantidad de genes expresados en ambas variedades de chile se encontraron en las librerías de 20 DDA.

- El proceso de maduración de fruto es guiado por las rutas metabólicas de respuesta a estímulos bióticos y abióticos, la mayoría de los términos GO enriquecidos en ambas variedades corresponden a categorías de respuesta a estímulos bióticos o abióticos.

- Las diferencias en los patrones de expresión de genes de las rutas metabólicas de señalización por auxina y ácido abscísico sugieren que el tamaño y forma elongada de los frutos de chile serrano podría deberse a un crecimiento vía ácida.

- Los genes OVATE y FAS/CLV3 expresados en mayor cantidad en chiltepín a 20 DDA, limitan la producción de celulas madres y el crecimiento celular y se sugiere que son responsables del tamaño pequeño y forma redonda de los frutos de chiltepín.

- La expresión en serrano a 60 DDA de los genes ACO3 y FUL2 involucrados en la ruta metabólica del etileno sugiere que podrían ser responsables de la maduración y senescencia temprana de los frutos de Serrano.

- Los DEGs de la ruta metabólica de metabolismo secundario sugieren que en chiltepín la biosíntesis de algunos metabolitos secundarios continúa durante el estado de fruto maduro.

- Los miRNAs identificados en las bibliotecas de chiltepín están relacionados a genes involucrados con la producción de carotenoides, degradación de azúcares, regulación de aciltransferasas y señalización hormonal, lo cual sugiere que podrían ser reguladores del proceso de maduración de fruto.

- El análisis transcriptómico sugiere que los genes involucrados en defensa mostraron una reducción en chile serrano comparado con los expresados en chiltepín; pudiendo ser una característica interesante en el proceso de domesticación del chile.

- Con base en los resultados obtenidos se corrobora la hipótesis planteada ya que se encontraron cambios cuantitativos y cualitativos en la expresión génica durante el desarrollo y maduración del fruto entre el chile serrano y chiltepín.

- A pesar de las grandes diferencias morfológicas entre los frutos de Chiltepín y Serrano, ambos presentan un patrón de expresión génica muy similar con diferencias en genes relacionados con la forma, tamaño y biosíntesis de etileno y metabolitos secundarios, sugiriendo que los procesos de domesticación de chile podrían ser muy específicos en el patrón de expresión de los genes involucrados en las características de importancia agronómica en frutos comerciales.

VIII. RECOMENDACIONES

- Se sugiere realizar un análisis de expresión diferencial utilizando un mayor número de puntos de muestreo, con lo cual podría ser posible identificar diferencias tanto en el desarrollo de fruto como en el proceso de maduración, incluso identificar los cambios en la fase de transición entre ambas etapas.
- Un análisis metabolómico podría ser realizado para dar soporte a los resultados obtenidos en este trabajo de investigación.
- Se sugiere realizar un análisis de interacción de ARNs no codificantes con las regiones promotoras de genes de interés, especialmente con las regiones de barrido selectivo previamente identificadas.

Referencias

2004. Calidad fisiológica de la semilla de chile piquín (Capsicum annuum var. aviculare) de dos localidades de Querétaro.

Abeles, F. B., P. W. Morgan, y M. E. Saltveit, 1992. Chapter 5 - roles and physiological effects of ethylene in plant physiology: Dormancy, growth, and development. En: Abeles, F. B., P. W. Morgan, y M. E. Saltveit (editores), Ethylene in Plant Biology, 2 ed^{ón}., págs. 120 – 181. Academic Press, New York. doi:10.1016/B978-0-08-091628-6.50011-4.

Aguilar-Meléndez, A., P. L. Morrell, M. L. Roose, y S. C. Kim, 2009a. Genetic diversity and structure in semiwild and domesticated chilies (capsicum annuum; solanaceae) from mexico. Am j Bot. 96:1190–1202.

Aguilar-Meléndez, A., P. L. Morrell, M. L. Roose, y S. C. Kim, 2009b. Genetic diversity and structure in semiwild and domesticated chilies (capsicum annuum; solanaceae) from mexico. Am j Bot. 96:1190–1202.

Aivalakis, G. y P. Katinakis, 2008. Biochemistry and molecular physiology of tomato and pepper fruit ripening. The Fruiting Species of the Solanaceae. The European Journal of Plant Science and Biotechnology 2:145–155.

Akula, R. y G. A. Ravishankar, 2011. Influence of abiotic stress signals on secondary metabolites in plants. Plant Signaling & Behavior 6(11):1720–1731. doi:10.4161/psb.6.11.17613.

Alonso, R. A., C. Moya, A. Cabrera, P. Ponce, R. Quiroga, M. A. Rosales, y J. L. Zuart, 2008. Evaluación in situ de la variabilidad genética de los chiles silvestres (capsicum spp.) en la región frailesca del estado de chiapas, méxico. Cultivos Tropicales 9(2).

Alvarez-Parrilla, E., L. A. de la Rosa, R. Amarowicz, y F. Shahidi, 2011. Antioxidant activity of fresh and processed jalapeño and serrano peppers. Journal of Agricultural and Food Chemistry 59(1):163–173.

Anders, S., P. T. Pyl, y W. Huber, 2014. HTSeq - a Python framework to work with high-throughput sequencing data. Bioinformatics 31(2):166–169. doi:10.1093/bioinformatics/btu638.

Anogianaki, A., N. N Negrev, Y. B Shaik, M. L Castellani, S. Frydas, J. Vecchiet, S. Tete, V. Salini, D. De Amicis, M. A De Lutiis, F. Conti, A. Caraffa, y G. Cerulli, 2007. Capsaicin: An irritant anti-inflammatory compound. Journal of biological regulators and homeostatic agents 21:1–4.

Arimboor, R., R. B. Natarajan, K. R. Menon, L. P. Chandrasekhar, y V. Moorkoth, 2015. Red pepper (capsicum annuum) carotenoids as a source of natural food colors: analysis and stability—a review. Journal of Food Science and Technology 52(3):1258–1271.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, y G. Sherlock, 2000. Gene ontology: tool for the unification of biology. Nature Genetics 25(1):25–29.

Ashrafi, H., T. Hill, K. Stoffel, A. Kozik, J. Yao, S. R. Chin-Wo, y A. Van Deynze, 2012. De novo assembly of the pepper transcriptome (capsicum annuum): a benchmark for in silico discovery of snps, ssrs and candidate genes. BMC Genomics 13(1):571. doi: 10.1186/1471-2164-13-571.

Assis, A. F., E. H. Oliveira, P. B. Donate, S. Giuliatti, C. Nguyen, y G. A. Passos, 2014. What Is the Transcriptome and How it is Evaluated?, págs. 3–48. Springer International Publishing, Cham. doi:10.1007/978-3-319-11985-4_1.

Atta-Aly, M. A., J. K. Brecht, y D. J. Huber, 2000. Ethylene feedback mechanisms in tomato and strawberry fruit tissues in relation to fruit ripening and climacteric patterns. Postharvest Biology and Technology 20(2):151 – 162. doi:10.1016/S0925-5214(00)00124-1.

Barg, R., I. Sobolev, T. Eilon, A. Gur, I. Chmelnitsky, S. Shabtai, E. Grotewold, y Y. Salts, 2005. The tomato early fruit specific gene lefsm1 defines a novel class of plant-specific sant/myb domain proteins. Planta 221(2):197–211. doi:10.1007/s00425-004-1433-0.

Barry, C. S., M. I. Llop-Tous, y D. Grierson, 2000. The regulation of 1-aminocyclopropane-1-carboxylic acid synthase gene expression during the transition from system-1 to system-2 ethylene synthesis in tomato. Plant Physiology 123(3):979–986. doi:10.1104/pp.123.3.979.

Bañuelos, N., P. L. Salido, y A. Gardea, 2008. Etnobotánica del chiltepín: Pequeño gran señor en la cultura de los sonorenses. Estudios sociales (Hermosillo, Son.) 16:177 – 205.

Beas, C., D. Ortuño, y J. S. Armendáriz, 2009. McGraw-Hill Interamericana.

Bhattacharya, A., P. Sood, y V. Citovsky, 2010. The roles of plant phenolics in defence and communication during agrobacterium and rhizobium infection. Molecular plant pathology 11(5):705–719.

Billing, J. y P. W. Sherman, 1998. Antimicrobial functions of spices: Why some like it hot. The Quarterly review of biology 73:3–49.

Bird, C. R., J. A. Ray, J. D. Fletcher, J. M. Boniwell, A. S. Bird, C. Teulieres, I. Blain, P. M. Bramley, y W. Schuch, 1991. Using antisense rna to study gene function: Inhibition of carotenoid biosynthesis in transgenic tomatoes. Bio/Technology 9(7):635–639. doi: 10.1038/nbt0791-635.

Bolger, A. M., M. Lohse, y B. Usadel, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

Bollier, N., A. Sicard, J. Leblond, D. Latrasse, N. Gonzalez, F. Gévaudant, M. Benhamed, C. Raynaud, M. Lenhard, C. Chevalier, M. Hernould, y F. Delmas, 2018. At-mini zinc finger2 and sl-inhibitor of meristem activity, a conserved missing link in the regulation of floral meristem termination in arabidopsis and tomato. The Plant Cell 30(1):83–100. doi: 10.1105/tpc.17.00653.

Bouvier, F., R. Backhaus, y B. Camara, 1998a. Induction and control of chromoplast-specific carotenoid genes by oxidative stress. The Journal of biological chemistry 273:30651–9.

Bouvier, F., Y. Keller, A. d'Harlingue, y B. Camara, 1998b. Xanthophyll biosynthesis: molecular and functional characterization of carotenoid hydroxylases from pepper fruits (capsicum annuum l.). Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism 1391(3):320 – 328. doi:https://doi.org/10.1016/S0005-2760(98)00029-0.

Brady, C. J. y J. Speirs, 1991. Ethylene in fruit ontogeny and abscission. En: Mattoo, A. K. y J. C. Suttle (editores), The plant hormone ethylene, págs. 235–258. CRC Press, Boca Raton, Fla.

Candar-Cakir, B., E. Arican, y B. Zhang, 2016. Small rna and degradome deep sequencing reveals drought-and tissue-specific micrornas and their important roles in drought-sensitive and drought-tolerant tomato genotypes. Plant Biotechnology Journal 14(8):1727–1746. doi: https://doi.org/10.1111/pbi.12533.

Carrari, F. y A. R. Fernie, 2006. Metabolic regulation underlying tomato fruit development. Journal of Experimental Botany 57(9):1883–1897. doi:10.1093/jxb/erj020.

Castro-Concha, L., I. Canche-Chuc, y M. Miranda-Ham, 2012. Determination of antioxidants in fruit tissues from three accessions of habanero pepper (capsicum chinense jacq.). Journal of the Mexican Chemical Society 56:15–18. doi:10.29356/jmcs.v56i1.269.

Chancellor, M. B. y W. C. De Groat, 1999. Intravesical capsaicin and resiniferatoxin therapy: spicing up the ways to treat the overactive bladder. The Journal of Urology 162:3–11.

Chang, Y.-F., J. S. Imam, y M. F. Wilkinson, 2007. The nonsense-mediated decay rna surveillance pathway. Annual Review of Biochemistry 76(1):51–74. doi:10.1146/annurev.biochem.76.050106.093909.

Chen, D. y W. Zanmin, 2009. Study on extraction and purification process of capsicum red pigment. Journal of Agricultural Sciences 1(2). doi:10.5539/jas.v1n2p94.

Chrost, B., F. J., K. B., M. H., y K. K., 1999. Tocopherol Biosynthesis in Senescing Chloroplasts - A Mechanism to Protect Envelope Membranes against Oxidative Stress and a Prerequisite for Lipid Remobilization?, tomo 64. Springer. doi:10.1007/978-94-011-4788-0_27.

Chu, Y.-H., J.-C. Jang, Z. Huang, y E. van der Knaap, 2019. Tomato locule number and fruit size controlled by natural alleles of lc and fas. Plant Direct 3(7):e00142. doi:10.1002/pld3.142.

Chung, M.-Y., J. Vrebalov, R. Alba, J. Lee, R. McQuinn, J.-D. Chung, P. Klein, y J. Giovannoni, 2010. A tomato (solanum lycopersicum) apetala2/erf gene, slap2a, is a negative regulator of fruit ripening. The Plant Journal 64(6):936–947. doi:10.1111/j.1365-313X.2010.04384.x.

Conesa, A. y S. Götz, 2008. Blast2go: A comprehensive suite for functional analysis in plant genomics. International journal of plant genomics 2008:619832–619832. doi:10.1155/2008/619832.

Conesa, A., S. Götz, J. García-Gómez, J. Terol, M. Talon, y M. Robles, 2005. Blast2go: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics (Oxford, England) 21:3674–6. doi:10.1093/bioinformatics/bti610.

Consortium, T. U., 2018. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research 47(D1):D506–D515. doi:10.1093/nar/gky1049.

Costa-Silva, J., D. Domingues, y F. M. Lopes, 2017. Rna-seq differential expression analysis: An extended review and a software tool. PLOS ONE 12(12):1–18. doi:10.1371/journal.pone.0190152.

Cruz, F., 2004. Mechanisms involved in new therapies for overactive bladder. Urology 63(3, Supplement 1):65 – 73. doi:10.1016/j.urology.2003.11.001.

Cuevas-Glory, L. F., O. Sosa-Moguel, J. Pino, y E. Sauri-Duch, 2015. Gc–ms characterization of volatile compounds in habanero pepper (capsicum chinense jacq.) by optimization of headspace solid-phase microextraction conditions. Food Analytical Methods 8(4):1005–1013.

Davoodi Mastakani, F., G. Pagheh, S. Rashidi Monfared, y M. Shams-Bakhsh, 2018. Identification and expression analysis of a microrna cluster derived from pre-ribosomal rna in papaver somniferum l. and papaver bracteatum l. PLOS ONE 13(8):1–20. doi:10.1371/journal.pone.0199673.

De Pace, C., P. Vaccino, P. G. Cionini, M. Pasquini, M. Bizzarri, y C. O. Qualset, 2011. Dasypyrum, págs. 185–292. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-14228-4_4.

del Rosario Abraham-Juárez, M., C. Rocha Granados, M. G. López, R. Rivera-Bustamante, y N. Ochoa-Alejo, 2008. Virus-induced silencing of comt, pamt and kas genes results in a reduction of capsaicinoid accumulation in chili pepper fruits. Planta 227:681–95.

DellaPenna, D. y B. J. Pogson, 2006. Vitamin synthesis in plants: Tocopherols and carotenoids. Annual Review of Plant Biology 57(1):711–738. doi:10.1146/annurev.arplant.56.032604.144301.

DeWitt, D. y P. W. Bosland, 2009. Complete Chile Pepper Book: A Gardenerś Guide to Choosing, Growing, Preserving, and Cooking. Timber Press.

Doka, O., G. Ficzek, S. Luterotti, D. Bicanic, R. Spruijt, J. Buijnsters, y G. Vegvari, 2013. Simple and rapid quantification of total carotenoids in lyophilized apricots (prunus armeniaca l.) by means of reflectance colorimetry and photoacoustic spectroscopy. Food Technology and Biotechnology 51(4):453–459.

Dong, T., G. Chen, S. Tian, Q. Xie, W. Yin, Y. Zhang, y Z. Hu, 2014. A non-climacteric fruit gene camads-rin regulates fruit ripening and ethylene biosynthesis in climacteric fruit. PLOS ONE 9(4):1–12. doi:10.1371/journal.pone.0095559.

Drago-Serrano, M. E., M. López-López, y T. d. R. Sainz-Espuñes, 2006. Componentes bioactivos de alimentosfuncionales de origen vegetal. Revista Mexicana de Ciencias Farmacéuticas 37(4).

Dubey, M., V. Jaiswal, A. Rawoof, A. Kumar, M. Nitin, S. S. Chhapekar, N. Kumar, I. Ahmad, K. Islam, V. Brahma, y N. Ramchiary, 2019. Identification of genes involved in fruit development/ripening in capsicum and development of functional markers. Genomics 111(6):1913 – 1922. doi:https://doi.org/10.1016/j.ygeno.2019.01.002.

Evans, T. G., 2015. Considerations for the use of transcriptomics in identifying the 'genes that matter' for environmental adaptation. Journal of Experimental Biology 218(12):1925–1935. doi:10.1242/jeb.114306.

FAOSTAT, 2020. Production/yield quantities of chillies and peppers in 2018. report update september 14, 2020. (http://www.fao.org).

Felger, R. S. y M. B. Moser, 1995. Rio Mayo plants. A study of the flora and vegetation of the valley of the Rio Mayo, Sonora. The University of Arizona Press.

Forero, M. D., C. E. Quijano, y J. A. Pino, 2009. Volatile compounds of chile pepper (capsicum annuum l. var. glabriusculum) at two ripening stages. Flavour and Fragrance Journal 24(1):25–30. doi:10.1002/ffj.1913.

Geniza, M. y P. Jaiswal, 2017. Tools for building de novo transcriptome assembly. Current Plant Biology 11-12:41 – 45. doi:10.1016/j.cpb.2017.12.004. Genomic resources and databases.

Gentry, H. S., 1942. Rio Mayo plants. A study of the flora and vegetation of the valley of the Rio Mayo, Sonora.

Giovannoni, J., 2001. Molecular biology of fruit maturation and ripening. Annual Review of Plant Physiology and Plant Molecular Biology 52(1):725–749. doi:10.1146/annurev.arplant.52.1.725.

Góngora-Castillo, E., R. Fajardo-Jaime, A. Fernández-Cortes, A. E. Jofre-Garfias, E. Lozoya-Gloria, O. Martínez, N. Ochoa-Alejo, y R. Rivera-Bustamante, 2012. The capsicum transcriptome db: a "hot"tool for genomic research. Bioinformation 8(1):43–47.

González-Jara, P., A. Moreno-Letelier, A. Fraile, D. Piñero, y F. García-Arenal, 2011a. Impact of human management on the genetic variation of wild pepper capsicum annuum var glabriusculum. PLoS ONE. 6(12).

González-Jara, P., A. Moreno-Letelier, A. Fraile, D. Piñero, y F. García-Arenal, 2011b. Impact of human management on the genetic variation of wild pepper capsicum annuum var glabriusculum. PLoS ONE. 6(12).

Govindarajan, V. S. y M. N. Sathyanarayana, 1991. Capsicum — production, technology, chemistry, and quality. part v. impact on physiology, pharmacology, nutrition, and metabolism; structure, pungency, pain, and desensitization sequences. Critical Reviews in Food Science and Nutrition 29(6):435–474. doi:10.1080/10408399109527536.

Greenlaf, W., 1986. Pepper breeding. Breeding Vegetable Crops. págs. 67–134.

Greer, E. L. y Y. Shi, 2012. Histone methylation: a dynamic mark in health, disease and inheritance. Nature Reviews Genetics 13(5):343–357. doi:10.1038/nrg3173.

Grubben, G. y I. El Tahir, 2004. Capsicum annuum l. PROTA 2: Vegetables / Legumes. .

Gu, M., W. Liu, Q. Meng, W. Zhang, A. Chen, S. Sun, y G. Xu, 2014. Identification of micrornas in six solanaceous plants and their potential link with phosphate and mycorrhizal signaling. Journal of Integrative Plant Biology 56(12):1164–1178. doi:https://doi.org/10.1111/jipb.12233.

Guil-Guerrero, J. y M. Rebolloso-Fuentes, 2009. Nutrient composition and antioxidant activity of eight tomato (lycopersicon esculentum) varieties. Journal of Food Composition and Analysis 22(2):123 – 129. doi:10.1016/j.jfca.2008.10.012.

Gómez-García, M. y N. Ochoa-Alejo, 2013. Biochemistry and molecular biology of carotenoid biosynthesis in chili peppers (capsicum spp.). International Journal of Molecular Sciences 14(9):19025–19053. doi:10.3390/ijms140919025.

Haas, B., A. Papanicolaou, M. Yassour, M. Grabherr, P. D Blood, J. Bowden, M. Brian Couger, D. Eccles, B. Li, M. Lieber, M. D MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. Dewey, y A. Regev, 2013. De novo transcript sequence reconstruction from rna-seq using the trinity platform for reference generation and analysis doi:10.1038/nprot.2013.084.

Hackbusch, J., K. Richter, J. Müller, F. Salamini, y J. F. Uhrig, 2005. A central role of arabidopsis thaliana ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins. Proceedings of the National Academy of Sciences 102(13):4908–4912. doi:10.1073/pnas.0501181102.

Hashimshony, T., F. Wagner, N. Sher, y I. Yanai, 2012. Cel-seq: Single-cell rna-seq by multiplexed linear amplification. Cell Reports 2(3):666 – 673. doi:10.1016/j.celrep.2012.08.003.

Hayano-Kanashiro, A. C., N. Gámez-Meza, y L. A. Medina-Juárez, 2016. Capsicum annuum l. var. glabriusculum: Taxonomy, plant morphology, distribution, genetic diversity, genome sequencing and phytochemical compounds. Crop Science 56(1):1–11. doi:10.2135/cropsci2014.11.0789.

Hernández-Verdugo, S., R. Guevara-González, R. Rivera-Bustamante, C. Vázquez-yanes, y K. Oyama, 1998a. Los parientes silvestres del chile (capsicum spp.) como recursos genéticos. Bol. Soc. Bot. Méx. 62:171–181.

Hernández-Verdugo, S., R. Guevara-González, R. Rivera-Bustamante, C. Vázquez-yanes, y K. Oyama, 1998b. Los parientes silvestres del chile (capsicum spp.) como recursos genéticos. Bol. Soc. Bot. Méx. 62:171–181.

Hernández-Verdugo, S., F. Porras, A. Pacheco-Olvera, R. G. López-España, M. Villarreal-Romero, S. Parra-Terraza, y T. Osuna-Enciso, 2012a. Caracterización y variación ecogeográfica de poblaciones de chile (capsicum annuum var. glabriusculum) silvestre del noroeste de méxico. Polibotánica. págs. 175 – 191.

Hernández-Verdugo, S., F. Porras, A. Pacheco-Olvera, R. G. López-España, M. Villarreal-Romero, S. Parra-Terraza, y T. Osuna-Enciso, 2012b. Caracterización y variación ecogeográfica de poblaciones de chile (capsicum annuum var. glabriusculum) silvestre del noroeste de méxico. Polibotánica. págs. 175 – 191.

Hoopes, L., 2008. Introduction to the gene expression and regulation topic room. Nature Education 1(1):160. doi:10.1104/pp.113.226142.

Hou, B.-Z., C.-L. Li, Y.-Y. Han, y Y.-Y. Shen, 2018. Characterization of the hot pepper (capsicum frutescens) fruit ripening regulated by ethylene and aba. BMC Plant Biology 18(1):162. doi:10.1186/s12870-018-1377-3.

Janick, J. y R. E. Paull, 2008. The Encyclopedia of Fruit and Nuts. CAB International, Wallingford, UK.

Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi, y M. Tanabe, 2015. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research 44(D1):D457–D462. doi: 10.1093/nar/gkv1070.

Kehie, M., S. Kumaria, P. Tandon, y N. Ramchiary, 2015. Biotechnological advances on in vitro capsaicinoids biosynthesis in capsicum: a review. Phytochemistry Reviews 14(2):189–201.

Kim, D., B. Langmead, y S. Salzberg, 2015. Hisat: A fast spliced aligner with low memory requirements. Nature methods 12. doi:10.1038/nmeth.3317.

Kim, H.-J., K.-H. Baek, S.-W. Lee, J. Kim, B.-W. Lee, H.-S. Cho, W. T. Kim, D. Choi, y C.-G. Hur, 2008. Pepper est database: comprehensive in silico tool for analyzing the chili pepper (capsicum annuum) transcriptome. BMC Plant Biology 8(1):101.

Kim, S., M. Park, S.-I. Yeom, Y.-M. Kim, J. M. Lee, H.-A. Lee, E. Seo, J. Choi, K. Cheong, K.-T. Kim, K. Jung, G.-W. Lee, S.-K. Oh, C. Bae, S.-B. Kim, H.-Y. Lee, S.-Y. Kim, M.-S. Kim, B.-C. Kang, Y. D. Jo, H.-B. Yang, H.-J. Jeong, W.-H. Kang, J.-K. Kwon, C. Shin, J. Y. Lim, J. H. Park, J. H. Huh, J.-S. Kim, B.-D. Kim, O. Cohen, I. Paran, M. C. Suh, S. B. Lee, Y.-K. Kim, Y. Shin, S.-J. Noh, J. Park, Y. S. Seo, S.-Y. Kwon, H. A. Kim, J. M. Park, H.-J.

Kim, S.-B. Choi, P. W. Bosland, G. Reeves, S.-H. Jo, B.-W. Lee, H.-T. Cho, H.-S. Choi, M.-S. Lee, Y. Yu, Y. Do Choi, B.-S. Park, A. van Deynze, H. Ashrafi, T. Hill, W. T. Kim, H.-S. Pai, H. K. Ahn, I. Yeam, J. J. Giovannoni, J. K. C. Rose, I. Sørensen, S.-J. Lee, R. W. Kim, I.-Y. Choi, B.-S. Choi, J.-S. Lim, Y.-H. Lee, y D. Choi, 2014. Genome sequence of the hot pepper provides insights into the evolution of pungency in capsicum species. Nature Genetics 46(3):270–278. doi:10.1038/ng.2877.

Klie, S., S. Osorio, T. Tohge, M. F. Drincovich, A. Fait, J. J. Giovannoni, A. R. Fernie, y Z. Nikoloski, 2014. Conserved changes in the dynamics of metabolic processes during fruit development and ripening across species. Plant Physiology 164(1):55–68. doi:10.1104/pp.113.226142.

Koch, M., Y. Arango, H.-P. Mock, y K.-P. Heise, 2002. Factors influencing α-tocopherol synthesis in pepper fruits. Journal of Plant Physiology 159(9):1015 – 1019. doi:10.1078/0176-1617-00746.

Kolodziejczyk, A., J. K. Kim, V. Svensson, J. Marioni, y S. Teichmann, 2015. The technology and biology of single-cell rna sequencing. Molecular Cell 58(4):610 – 620. doi:10.1016/j.molcel.2015.04.005.

Korkmaz, A., özlem Değer, y Y. Cuci, 2014. Profiling the melatonin content in organs of the pepper plant during different growth stages. Scientia Horticulturae 172:242–247. doi:https://doi.org/10.1016/j.scienta.2014.04.018.

Krajayklang, M., A. Klieber, y P. R. Dry, 2000. Colour at harvest and post-harvest behaviour influence paprika and chilli spice quality. Postharvest Biology and Technology 20(3):269 – 278. doi:10.1016/S0925-5214(00)00141-1.

L Deal, C., T. J Schnitzer, E. Lipstein, J. Seibold, R. M Stevens, M. D Levy, D. Albert, y F. Renold, 1991. Treatment of arthritis with topical capsaicin: A double-blind trial. Clinical therapeutics 13:383–95.

Lattanzio, V., 2013. Phenolic compounds: Introduction. En: Ramawat, K. G. y J.-M. Mérillon (editores), Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes, págs. 1543–1580. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-642-22144-6_57.

Lee, J. M., J.-G. Joung, R. McQuinn, M.-Y. Chung, Z. Fei, D. Tieman, H. Klee, y J. Giovannoni, 2012. Combined transcriptome, genetic diversity and metabolite profiling in tomato fruit reveals that the ethylene response factor slerf6 plays an important role in ripening and carotenoid accumulation. The Plant Journal 70(2):191–204. doi:10.1111/j.1365-313X.2011.04863.x.

Lee, S. y D. Choi, 2013. Comparative transcriptome analysis of pepper (capsicum annuum) revealed common regulons in multiple stress conditions and hormone treatments. Plant Cell Reports 32(9):1351–1359. doi:10.1007/s00299-013-1447-9.

Li, M., Z. Pang, W. Xiao, X. Liu, Y. Zhang, D. Yu, M. Yang, Y. Yang, J. Hu, y K. Luo, 2014. A transcriptome analysis suggests apoptosis-related signaling pathways in hemocytes of spodoptera litura after parasitization by microplitis bicoloratus. PLOS ONE 9(10):1–18. doi: 10.1371/journal.pone.0110967.

Li, X.-P., X.-C. Ma, H. Wang, Y. Zhu, X.-X. Liu, T.-T. Li, Y.-P. Zheng, J.-Q. Zhao, J.-W. Zhang, Y.-Y. Huang, M. Pu, H. Feng, J. Fan, Y. Li, y W.-M. Wang, 2020. Osamir162a fine-tunes rice resistance to magnaporthe oryzae and yield. Rice 13(1):38. doi: 10.1186/s12284-020-00396-2.

Lincoln, J. E. y R. L. Fischer, 1988. Regulation of gene expression by ethylene in wild-type and rin tomato (lycopersicon esculentum) fruit. Plant Physiology 88(2):370–374. doi:10.1104/pp.88.2.370.

Liscum, E. y J. W. Reed, 2002. Genetics of aux/iaa and arf action in plant growth and development. Plant Molecular Biology 49(3):387–400. doi:10.1023/A:1015255030047.

Liu, S., W. Li, Y. Wu, C. Chen, y J. Lei, 2013. De novo transcriptome assembly in chili pepper (capsicum frutescens) to identify genes involved in the biosynthesis of capsaicinoids. PLOS ONE 8(1):1–8. doi:10.1371/journal.pone.0048156.

Liu, W., C. Cheng, F. Chen, S. Ni, Y. Lin, y Z. Lai, 2018. High-throughput sequencing of small rnas revealed the diversified cold-responsive pathways during cold stress in the wild banana (musa itinerans). BMC Plant Biology 18(1):308.

López-Aguilar, R., D. Medina-Hernández, F. Ascencio-Valle, E. Troyo-Dieguez, A. Nieto-Garibay, M. Arce-Montoya, J. A. Larrinaga-Mayoral, y G. A. Gómez-Anduro, 2012a. Differential responses of chiltepin (capsicum annuum var. glabriusculum) and poblano (capsicum annuum var. annuum) hot peppers to salinity at the plantlet stage. African Journal of Biotechnology 11(11):2642–2653.

López-Aguilar, R., D. Medina-Hernández, F. Ascencio-Valle, E. Troyo-Dieguez, A. Nieto-Garibay, M. Arce-Montoya, J. A. Larrinaga-Mayoral, y G. A. Gómez-Anduro, 2012b. Differential responses of chiltepin (capsicum annuum var. glabriusculum) and poblano (capsicum annuum var. annuum) hot peppers to salinity at the plantlet stage. African Journal of Biotechnology 11(11):2642–2653.

Lowe, R., N. Shirley, M. Bleackley, S. Dolan, y T. Shafee, 2017. Transcriptomics technologies. PLOS Computational Biology 13(5):1–23. doi:10.1371/journal.pcbi.1005457.

Luo, X.-J., J. Peng, y Y.-J. Li, 2011. Recent advances in the study on capsaicinoids and capsinoids. European Journal of Pharmacology 650(1):1 – 7. doi:10.1016/j.ejphar.2010.09.074.

Martí, M. C., D. Camejo, E. Olmos, L. M. Sandalio, N. Fernández-García, A. Jiménez, y F. Sevilla, 2009. Characterisation and changes in the antioxidant system of chloroplasts and chromoplasts isolated from green and mature pepper fruits. Plant Biology 11(4):613–624. doi: 10.1111/j.1438-8677.2008.00149.x.

Martínez-López, L. A., N. Ochoa-Alejo, y O. Martínez, 2014. Dynamics of the chili pepper transcriptome during fruit development. BMC Genomics 15(1):143. doi:10.1186/1471-2164-15-143.

Martínez, O., M. L. Arce-Rodríguez, F. Hernández-Godínez, C. Escoto-Sandoval, F. Cervantes-Hernández, C. Hayano-Kanashiro, J. J. Ordaz-Ortiz, M. H. Reyes-Valdés, F. G. Razo-Mendivil, A. Garcés-Claver, y N. Ochoa-Alejo, 2021. Transcriptome analyses throughout chili pepper fruit development reveal novel insights into the domestication process. Plants 10(3). doi:10.3390/plants10030585.

Mata, J., S. Marguerat, y J. Bähler, 2005. Post-transcriptional control of gene expression: a genome-wide perspective. Trends in Biochemical Sciences 30(9):506–514.

Matas, A. J., T. H. Yeats, G. J. Buda, Y. Zheng, S. Chatterjee, T. Tohge, L. Ponnala, A. Adato, A. Aharoni, R. Stark, A. R. Fernie, Z. Fei, J. J. Giovannoni, y J. K. Rose, 2011. Tissue- and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. The Plant Cell 23(11):3893–3910. doi:10.1105/tpc.111.091173.

Mazid, M., T. A. Khan, y F. Mohammad, 2011. Role of secondary metabolites in defense mechanisms of plants. Biology and Medicine 3(2):232–249.

Mehrtens, F., H. Kranz, P. Bednarek, y B. Weisshaar, 2005. The arabidopsis transcription factor myb12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. Plant Physiology 138(2):1083–1096. doi:10.1104/pp.104.058032.

Melé, M., P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, T. J. Sullivan, R. Johnson, A. V. Segrè, S. Djebali, A. Niarchou, T. G. Consortium, F. A. Wright, T. Lappalainen, M. Calvo, G. Getz, E. T. Dermitzakis, K. G. Ardlie, y R. Guigó, 2015. The human transcriptome across tissues and individuals. Science 348(6235):660–665. doi:10.1126/science.aaa0355.

Milward, E., A. Shahandeh, M. Heidari, D. Johnstone, N. Daneshi, y H. Hondermarck, 2016. Transcriptomics. En: Bradshaw, R. A. y P. D. Stahl (editores), Encyclopedia of Cell Biology, págs. 160 – 165. Academic Press, Waltham. doi:10.1016/B978-0-12-394447-4.40029-5.

Miranda-Zarazúa, H., L. Villarruel-Sahagun, F. Ibarra-Flores, L. Gastelum-Peralta, y A. Morales-Coen, 2010a. Distribution and environmental factors associated with wild chiltepin in sonora (in spanish). 7th Int. Symp. on the Wild Flora in the Dry Areas págs. 503–513.

Miranda-Zarazúa, H., L. Villarruel-Sahagun, F. Ibarra-Flores, L. Gastelum-Peralta, y A. Morales-Coen, 2010b. Distribution and environmental factors associated with wild chiltepin in sonora (in spanish). 7th Int. Symp. on the Wild Flora in the Dry Areas págs. 503–513.

Moriya, Y., M. Itoh, S. Okuda, A. C. Yoshizawa, y M. Kanehisa, 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Research 35(2):W182–W185. doi:10.1093/nar/gkm321.

Muir, S. R., G. J. Collins, S. Robinson, S. Hughes, A. Bovy, C. H. Ric De Vos, A. J. van Tunen, y M. E. Verhoeyen, 2001. Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. Nature Biotechnology 19(5):470–474. doi: 10.1038/88150.

Nabhan, G., M. Slater, y L. Yarger, 1990a. New crops small farmers in marginal lands? wild chiles as a case study. Agroecology and small farm development. págs. 19–34.

Nabhan, G., M. Slater, y L. Yarger, 1990b. New crops small farmers in marginal lands? wild chiles as a case study. Agroecology and small farm development. págs. 19–34.

Naing, A. H., S. Y. Kyu, P. P. W. Pe, K. I. Park, J. M. Lee, K. B. Lim, y C. K. Kim, 2019. Silencing of the phytoene desaturase (pds) gene affects the expression of fruit-ripening genes in tomatoes. Plant Methods 15(1):110.

Nuñez-Palenius, H. G. y N. Ochoa-Alejo, 2005. Effect of phenylalanine and phenylpropanoids on the accumulation of capsaicinoids and lignin in cell cultures of chili pepper (capsicum annuum l.). In Vitro Cellular & Developmental Biology - Plant 41(6):801–805.

Osorio, S., R. Alba, Z. Nikoloski, A. Kochevenko, A. R. Fernie, y J. J. Giovannoni, 2012. Integrative comparative analyses of transcript and metabolite profiles from pepper and tomato ripening and development stages uncovers species-specific patterns of network regulatory behavior. Plant Physiology 159(4):1713–1729. doi:10.1104/pp.112.199711.

Pacheco-Olivera, A., S. Hernández-Verdugo, V. Rocha-Ramírez, A. González-Rodríguez, y K. Oyama, 2012a. Genetic diversity and structure of pepper (capsicum annuum l.) from northwestern mexico analyzed by microsatellite markers. Crop Science, 52:231– 241.

Pacheco-Olivera, A., S. Hernández-Verdugo, V. Rocha-Ramírez, A. González-Rodríguez, y K. Oyama, 2012b. Genetic diversity and structure of pepper (capsicum annuum l.) from northwestern mexico analyzed by microsatellite markers. Crop Science, 52:231– 241.

Paul, V., R. Pandey, y G. C. Srivastava, 2012. The fading distinctions between classical patterns of ripening in climacteric and non-climacteric fruit and the ubiquity of ethylene—an overview. Journal of Food Science and Technology 49(1):1–21. doi:10.1007/s13197-011-0293-4.

Perkins-Veazie, P., 1995. Growth and Ripening of Strawberry Fruit, cap. 8, págs. 267–297. John Wiley & Sons, Ltd. doi:10.1002/9780470650585.ch8.

Perry, L., 2012a. Ethnobotany., págs. 1–13. CABI. doi:10.1079/9781845937676.0000.

Perry, L., 2012b. Ethnobotany., págs. 1–13. CABI.

Perry, L., R. Dickau, S. Zarrillo, I. Holst, D. Pearsall, D. Piperno, M. Berman, R. Cooke, K. Rademaker, A. Ranere, J. Raymon, D. Sandweiss, F. Scaramelli, k. Tarble, y J. Zeidler, 2007. Starch fossils and the domestication and dispersal of chili peppers (capsicum spp. l.) in the americas. Science 315:986–988.

Perry, L. y K. Flannery, 2007a. Precolumbian use of chili peppers in the valley of oaxaca, mexico. Proceedings of National Academy of Science of the USA 104:11905–11909.

Perry, L. y K. Flannery, 2007b. Precolumbian use of chili peppers in the valley of oaxaca, mexico. Proceedings of National Academy of Science of the USA 104:11905–11909.

Pickersgill, B., 1984. Migrations of chili peppers, capsicum spp., in the americas. in pre-columbian plant migration. Papers of the Peabody Museum of Archaeology and Ethnology. 76:105–124.

Pickersgill, B., 2007. Domestication of Plants in the Americas: Insights from Mendelian and Molecular Genetics. Annals of Botany 100(5):925–940. doi:10.1093/aob/mcm193.

Pomaznoy, M., B. Ha, y B. Peters, 2018. Gonet: a tool for interactive gene ontology analysis. BMC Bioinformatics 19(1):470.

Pretel, M., M. Serrano, A. Amoros, F. Riquelme, y F. Romojaro, 1995. Non-involvement of acc and acc oxidase activity in pepper fruit ripening. Postharvest Biology and Technology 5(4):295 – 302. doi:10.1016/0925-5214(94)00025-N.

Qian, X., Y. Ba, Q. Zhuang, y G. Zhong, 2014. Rna-seq technology and its application in fish transcriptomics. OMICS: A Journal of Integrative Biology 18(2):98–110. doi:10.1089/omi.2013.0110.

Qin, C., C. Yu, Y. Shen, X. Fang, L. Chen, J. Min, J. Cheng, S. Zhao, M. Xu, Y. Luo, Y. Yang, Z. Wu, L. Mao, H. Wu, C. Ling-Hu, H. Zhou, H. Lin, S. González-Morales, D. L. Trejo-Saavedra, H. Tian, X. Tang, M. Zhao, Z. Huang, A. Zhou, X. Yao, J. Cui, W. Li, Z. Chen, Y. Feng, Y. Niu, S. Bi, X. Yang, W. Li, H. Cai, X. Luo, S. Montes-Hernández, M. A. Leyva-González, Z. Xiong, X. He, L. Bai, S. Tan, X. Tang, D. Liu, J. Liu, S. Zhang, M. Chen, L. Zhang, L. Zhang, Y. Zhang, W. Liao, Y. Zhang, M. Wang, X. Lv, B. Wen, H. Liu, H. Luan, Y. Zhang, S. Yang, X. Wang, J. Xu, X. Li, S. Li, J. Wang, A. Palloix, P. W. Bosland, Y. Li, A. Krogh, R. F. Rivera-Bustamante, L. Herrera-Estrella, Y. Yin, J. Yu, K. Hu, y Z. Zhang, 2014. Whole-genome sequencing of cultivated and wild peppers provides insights into capsicum domestication and specialization. Proceedings of the National Academy of Sciences 111(14):5135–5140. doi:10.1073/pnas.1400975111.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rangan, P., A. Furtado, y R. Henry, 2017. The transcriptome of the developing grain: A resource for understanding seed development and the molecular control of the functional and nutritional properties of wheat. BMC Genomics 18. doi:10.1186/s12864-017-4154-z.

Ransohoff, J. D., Y. Wei, y P. A. Khavari, 2018. The functions and unique features of long intergenic non-coding rna. Nature Reviews Molecular Cell Biology 19(3):143–157.

Razo-Mendivil, F. G., O. Martínez, y C. Hayano-Kanashiro, 2020. Compacta: a fast contig clustering tool for de novo assembled transcriptomes. BMC Genomics 21(1):148.

Remm, M., C. Storm, y E. Sonnhammer, 2002. Automatic clustering of orthologs and inparalogs from pairwise species comparisons. Journal of molecular biology 314:1041–52. doi:10.1006/jmbi.2000.5197.

Robinson, M. D., D. J. McCarthy, y G. K. Smyth, 2009. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140. doi:10.1093/bioinformatics/btp616.

Roundtree, I. A., M. E. Evans, T. Pan, y C. He, 2017. Dynamic rna modifications in gene expression regulation. Cell 169(7):1187–1200.

Russo, V., 2012. Peppers, botany production and uses. CABI.

Sandberg, R., 2014. Entering the era of single-cell transcriptomics in biology and medicine. Nature methods 11:22–4. doi:10.1038/nmeth.2764.

Sepúlveda-Jiménez, G., H. Porta-Ducoing, y M. Rocha-Sosa, 2003. La participación de los metabolitos secundarios en la defensa de las plantas. Revista Mexicana de Fitopatología 21(3):355–363.

Si, W., Y. Liang, K. Y. Ma, H. Y. Chung, y Z.-Y. Chen, 2012. Antioxidant activity of capsaicinoid in canola oil. Journal of Agricultural and Food Chemistry 60(24):6230–6234. doi:10.1021/jf301744q.

Silva, L., J. Azevedo, M. J. Valle Pereira, P. Valentão, y P. Andrade, 2012. Chemical assessment and antioxidant capacity of pepper (capsicum annuum l.) seeds. Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association 53. doi:10.1016/j.fct.2012.11.036.

Singh, R. J., 2006a. pág. 203. CRC press.

Singh, R. J., 2006b. Genetic Resources, Chromosome Engineering, and Crop Improvement: Vegetable crops., pág. 203. CRC press.

Singh, S., R. Jarret, V. Russo, G. Majetich, J. Shimkus, R. Bushway, y B. Perkins, 2009. Determination of capsinoids by hplc-dad in capsicum species. Journal of Agricultural and Food Chemistry 57(9):3452–3457. doi:10.1021/jf8040287.

Slugina, M. A., A. V. Shchennikova, O. N. Pishnaya, y E. Z. Kochieva, 2018. Assessment of the fruit-ripening-related ful2 gene diversity in morphophysiologically contrasted cultivated and wild tomato species. Molecular Breeding 38(7):82. doi:10.1007/s11032-018-0842-x.

Smith, C., 1967. The prehistory of the tehuacan valley. Plant Remains. 1:220–255.

Soneson, C. y M. Delorenzi, 2013. A comparison of methods for differential expression analysis of rna-seq data. BMC Bioinformatics 14(1):91.

Spartz, A. K., H. Ren, M. Y. Park, K. N. Grandt, S. H. Lee, A. S. Murphy, M. R. Sussman, P. J. Overvoorde, y W. M. Gray, 2014. Saur inhibition of pp2c-d phosphatases activates plasma membrane h+-atpases to promote cell expansion in arabidopsis. The Plant Cell 26(5):2129–2142. doi:10.1105/tpc.114.126037.

Srinivasan, K., 2005a. Spices as influencers of body metabolism: an overview of three decades of research. Food Res. Int. 38:77–86.

Srinivasan, K., 2005b. Spices as influencers of body metabolism: an overview of three decades of research. Food Research International 38(1):77 – 86. doi:10.1016/j.foodres.2004.09.001.

Stephan, W., 2019. Selective Sweeps. Genetics 211(1):5–13. doi:10.1534/genetics.118.301319.

Stortenbeker, N. y M. Bemer, 2018. The SAUR gene family: the plant's toolbox for adaptation of growth and development. Journal of Experimental Botany 70(1):17–27. doi:10.1093/jxb/ery332.

Struhl, K., 1998. Histone acetylation and transcriptional regulatory mechanisms. Genes and Development 12(5):599 – 606.

Su Han, S., Y.-S. Keum, H.-J. Seo, K.-S. Chun, S. Sup Lee, y Y.-J. Surh, 2001. Capsaicin suppresses phorbol ester-induced activation of nf-??b/rel and ap1 transcription factors in mouse epidermis. Cancer Letters - CANCER LETT 164:119–126.

Surh, Y., 2002. More than spice: Capsaicin in hot chili peppers makes tumor cells commit suicide. JNCI Journal of the National Cancer Institute 94:1263–1265.

Tanksley, S. D., 2004. The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. The Plant Cell 16(suppl 1):S181–S189. doi:10.1105/tpc.018119.

Tian, M. S., S. Prakash, H. J. Elgar, H. Young, D. M. Burmeister, y G. S. Ross, 2000. Responses of strawberry fruit to 1-methylcyclopropene (1-mcp) and ethylene. Plant Growth Regulation 32(1):83–90.

Tohge, T., N. Y, M. Hirai, Y. M, N. J, A. M, I. E, T. H, D. Goodenowe, K. M, M. Noji, M. Yamazaki, y K. Saito, 2005. Functional genomics by integrated analysis of metabolome and transcriptome of arabidopsis plants over-expressing an myb transcription factor. The Plant Journal 42:218–35.

Tsaballa, A., K. Pasentsis, N. Darzentas, y A. S. Tsaftaris, 2011. Multiple evidence for the role of an ovate-like gene in determining fruit shape in pepper. BMC Plant Biology 11(1):46. doi:10.1186/1471-2229-11-46.

van Mourik, H., A. D. J. van Dijk, N. Stortenbeker, G. C. Angenent, y M. Bemer, 2017. Divergent regulation of arabidopsis saur genes: a focus on the saur10-clade. BMC Plant Biology 17(1):245.

Verweij, W., C. E. Spelt, M. Bliek, M. de Vries, N. Wit, M. Faraco, R. Koes, y F. M. Quattrocchio, 2016. Functionally similar wrky proteins regulate vacuolar acidification in petunia and hair development in arabidopsis. The Plant Cell 28(3):786–803. doi:10.1105/tpc.15.00608.

Vrebalov, J., I. L. Pan, A. J. M. Arroyo, R. McQuinn, M. Chung, M. Poole, J. Rose, G. Seymour, S. Grandillo, J. Giovannoni, y V. F. Irish, 2009. Fleshy fruit expansion and ripening are regulated by the tomato shatterproof gene tagl1. The Plant Cell 21(10):3041–3062. doi:10.1105/tpc.109.066936.

Wahyuni, Y., A.-R. Ballester, E. Sudarmonowati, R. J. Bino, y A. G. Bovy, 2013. Secondary metabolites of capsicum species and their importance in the human diet. Journal of Natural Products 76(4):783–793. doi:10.1021/np300898z.

Wang, S., G. Lu, Z. Hou, Z. Luo, T. Wang, H. Li, J. Zhang, y Z. Ye, 2014. Members of the tomato FRUITFULL MADS-box family regulate style abscission and fruit ripening. Journal of Experimental Botany 65(12):3005–3014. doi:10.1093/jxb/eru137.

Wang, Z., M. Gerstein, y M. Snyder, 2009. Rna-seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10(1):57–63.

Wu, S., J. P. Clevenger, L. Sun, S. Visa, Y. Kamiya, Y. Jikumaru, J. Blakeslee, y E. van der Knaap, 2015. The control of tomato fruit elongation orchestrated by sun, ovate and fs8.1 in a wild relative of tomato. Plant Science 238:95 – 104. doi:https://doi.org/10.1016/j.plantsci.2015.05.019.

Young, M. D., M. J. Wakefield, G. K. Smyth, y A. Oshlack, 2010. Gene ontology analysis for rna-seq: accounting for selection bias. Genome Biology 11:R14.

Yu, G., 2018. clusterprofiler: An universal enrichment tool for functional and comparative study. bioRxiv doi:10.1101/256784.

Zeder, M. A., 2015. Core questions in domestication research. Proceedings of the National Academy of Sciences 112(11):3191–3198. doi:10.1073/pnas.1501711112.

Zhang, Z., A. Lu, y W. Dárcy, 2002a. Capsicum annuum linnaeus, special plant. Flora of China. 17:313? 313.

Zhang, Z., A. Lu, y W. Dárcy, 2002b. Capsicum annuum linnaeus, special plant. Flora of China. 17:313– 313.

Zhang, Z.-H., N. R. Wray, y Q. Zhao, 2016. Dear-o: Differential expression analysis based on rna-seq data - online. bioRxiv doi:10.1101/069807.

Zhigila, D. A., A. A. AbdulRahaman, O. S. Kolawole, y F. A. Oladele, 2014. Fruit morphology as taxonomic features in five varieties of capsicum annuum l. solanaceae. Journal of Botany. 2014:6.

APÊNDICES

A. Metodología

A.1. Análisis y control de calidad de las lecturas de secuenciación

Linea de comando

```
FILES=`ls ${INPUTDIR}/*_1.fastq | sed 's/_1.fastq//g'|sed "s,${INPUTDIR}/,,g"`
for filename in $FILES ; do

fastqc -o $FASTQCOUTRAW ${INPUTDIR}/${filename}_1.fastq ${INPUTDIR}/${filename}_2.fastq

java -jar /opt/Trimmomatic/trimmomatic-0.36.jar PE -threads 8 -trimlog trimmlog.txt -
    phred33 ${INPUTDIR}/${filename}_1.fastq ${INPUTDIR}/${filename}_2.fastq -baseout ${
    OUTDIR}/${filename}_TM.fastq  ILLUMINACLIP:/opt/Trimmomatic/adapters/TruSeq3-PE.fa
    :2:30:10 SLIDINGWINDOW:4:15 MINLEN:50

fastqc -o $FASTQCOUT ${OUTDIR}/${filename}_TM_1P.fastq ${OUTDIR}/${filename}_TM_2P.
    fastq

done
```

A.2. Alineamiento al genoma de referencia y estimación de abundancias

Linea de comando

```
FILES=`ls ${INPUTDIR}/*_1P.fastq | sed 's/_1P.fastq//g'|sed "s,${INPUTDIR}/,,g"`
```

```
for filename in $FILES ; do
  ~/Descargas/hisat2-2.1.0/hisat2 -x $INDEX --known-splicesite-infile $SPLICE_SITES  -p
     8 -q --dta -1 ${INPUTDIR}/${filename}_1P.fastq -2 ${INPUTDIR}/${filename}_2P.fastq
     -U ${INPUTDIR}/${filename}_1U.fastq,${INPUTDIR}/${filename}_2U.fastq -S ${OUTDIR}/
    ${filename}.sam
  if [ "$?" = "0"  ]
  then
    samtools view -Su ${OUTDIR}/${filename}.sam | samtools sort -o ${OUTDIR}/${filename
    }.bam
  else
    echo 'ERROR!!,no se encontro el archivo '${OUTDIR}/${filename}.sam', pasare al
    siguiente archivo !\n'
  fi
done
```

Linea de comando

```
  FILES=`ls ${OUTDIR}/*.bam | sed 's/.bam//g'|sed "s,${OUTDIR}/,,g"`
for filename in $FILES ; do
samtools view -bf 1 ${OUTDIR}/${filename}.bam | htseq-count --format bam --order pos --
    mode union --stranded no - ${GTFFILE}  > ${HTSEQOUT}/${filename}_htseq.tsv
done
```

A.3. Análisis estadístico

A.3.1. Análisis de expresión diferencial

Código R

```
library("edgeR")
library("vidger")
```

```
library("VennDiagram")
library("gplots")
library(ggplot2)
library("org.At.tair.db")
library(readr)
library("pheatmap")

setwd("/datos/")

fcVar<-4
filtType<-3

if(!dir.exists("DE")){
  dir.create("DE/")
  dir.create("DE/Plots")
  dir.create("DE/Plots/heatmaps/")
  dir.create("DE/Plots/vennDiagrams")
}
if(!dir.exists("Data_graphs")){
  dir.create("Data_graphs/")
}
if(!dir.exists("GO_enrichment")){
  dir.create("GO_enrichment/")
  dir.create("GO_enrichment/plots")
}
if(!dir.exists("Kegg_enrichment")){
  dir.create("Kegg_enrichment/")
  dir.create("Kegg_enrichment/plots")
}

```

```
print("loading counts and annotation files.")
annot=read.delim("../Cm334_uniprot_wholeTable.txt")
tableCounts=read.table("../All_counts_ChSr.tsv",header=TRUE,row.names =
    1,sep="\t")
tableCounts<-tableCounts[,c(5,6,13,14,15,16,17,18)]
tableCounts = as.matrix(round(tableCounts))
groups<-c("ST20","ST20","ST60","ST60","CH20","CH20","CH68","CH68")
groups<-factor(groups)
annot<-annot[match(rownames(tableCounts),annot$SeqName),]
dgeObject = DGEList(counts=tableCounts, genes=annot$keys,group=groups,
    remove.zeros = TRUE)

temp<-cpm(dgeObject)>1
keep<-(temp[,1]&temp[,2])|(temp[,3]&temp[,4])|(temp[,5]&temp[,6])|(temp
    [,7]&temp[,8])
dgeObject <- dgeObject[keep,]

dgeObject = calcNormFactors(dgeObject)

write.table(rownames(dgeObject$genes),"GO_enrichment/Background_genes.
    csv",quote = FALSE,row.names = FALSE,col.names=FALSE)

temp<-annot[match(rownames(dgeObject$genes),annot$SeqName),c(1,which(
    colnames(annot)=="zunla1_ids" ))]
temp<-unique(temp[!is.na(temp$zunla1_ids),2])
write.table(temp,"Kegg_enrichment/Universe_genes.csv",quote = FALSE,row.
    names = FALSE,col.names=FALSE)

print("plot MDS of all samples")
pch<- c(0,1,2,3,4,15,16,17,13)
```

```
colors<- rep(c("darkgreen","red","blue","darkorchid","darkorange"),2)
png(filename = "Data_graphs/plotMDS_serr_chilt.png",
pointsize = 12, res = NA)
plotMDS(dgeObject,col=colors[groups],pch=pch[groups])
legend("bottomright",legend=levels(groups),pch=pch,col=colors,ncol=2)
abline(0,0,col="black", lty=2, lwd=2)
abline(v=0,col="black", lty=2, lwd=2)
dev.off()

groupDef<-data.frame(c(rep("serrano",4),rep("chiltepin",4)),c(rep("20dda
    ",2),rep("60dda",2),rep("20dda",2),rep("60dda",2)))
rownames(groupDef)<-rownames(dgeObject$samples)
names(groupDef)<-c("species","time")
groupDef$group <- factor(paste(groupDef$species,groupDef$time,sep="."))

design <- model.matrix(~0+groupDef$group)
colnames(design) <- levels(groupDef$group)

dgeObject <- estimateDisp(dgeObject, design, robust=TRUE)
print("plotting BCV")

png(filename = "Data_graphs/plotBCV_Serr_chil.png",
pointsize = 12, res = NA)
plotBCV(dgeObject)
dev.off()

fit <- glmQLFit(dgeObject, design, robust=TRUE)
head(fit$coefficients)
summary(fit$df.prior)

```

```
86 png(filename = "Data_graphs/plotQLDisp_serr_chil.png", pointsize = 12,
       res = NA)
87 plotQLDisp(fit)
88 dev.off()
89
90 png(filename = "Data_graphs/plotviolin_serr_chilt.png",
91 pointsize = 12, res = NA)
92 vsBoxPlot(dgeObject,d.factor=NULL,type="edger",title=T,grid=T,aes = "
       violin")
93 dev.off()
94
95 myContrast<- makeContrasts(
96 chiltepin.20vs68dda=chiltepin.60dda-chiltepin.20dda,
97 serrano.20vs60dda=serrano.60dda-serrano.20dda,
98 both.20vs20dda=serrano.20dda-chiltepin.20dda,
99 both.60vs68dda=serrano.60dda-chiltepin.60dda,
100 chiltepinvsSerrano.20dda=(serrano.60dda-serrano.20dda)-(chiltepin.60dda-
       chiltepin.20dda),
101 levels=design)
102 myContrast
103
104 ContrastsElements<-data.frame(X=c("CH20","ST20","CH20","CH68"),Y=c("CH68
       ","ST60","ST20","ST60"),stringsAsFactors=F)
105 GenesDistribution<-data.frame(Constitutivos=rep(0,5),Inducidos=rep(0,5),
       Reprimidos=rep(0,5))
106 rownames(GenesDistribution)<-colnames(myContrast)
107 all_DETables<-list()
108
109 #
       ##############################################################################
```

```
for (i in 1:length(colnames(myContrast))){
  res <- glmQLFTest(fit, contrast=myContrast[,i])

  trTable <- glmTreat(fit, contrast=myContrast[,i], lfc=log2(fcVar))
  DE_genes <- decideTestsDGE(trTable,p.value = 0.01)
  summary(DE_genes)

  print("plotting MD per contrast")
  png(filename = paste("DE/Plots/PlotMD_",colnames(myContrast)[i],".png"
      ,sep=''),
  pointsize = 12, res = NA)
  plotMD(trTable, status=DE_genes, values=c(1,-1), col=c("red","blue"),
      legend="topright")
  dev.off()

  detags <- rownames(dgeObject)[as.logical(DE_genes)]
  deTable <- cbind(
  trTable$genes,
  sprintf('%0.3f',trTable$table$logFC),
  sprintf('%0.3f',2^abs(trTable$table$logFC)*(trTable$table$logFC/abs(
      trTable$table$logFC))),
  sprintf('%0.3f',trTable$table$logCPM),
  trTable$table$PValue,
  p.adjust(trTable$table$PValue, method="BH")
  )[as.logical(DE_genes),]
  colnames(deTable) <- c("Gene","LFC","FC","Log10_Pvalue","P.value","FDR
      " )
  all_DETables[[colnames(myContrast)[i]]]<-deTable
  #next()
```

```
135  deTable<-merge.data.frame(deTable,annot,by.x = 0,by.y = which(
         colnames(annot)=="SeqName" ),all.x = TRUE)
136  deTable<-deTable[order(deTable$P.value,decreasing=FALSE),]
137  write.table(deTable,file=paste("DE/DEGenes_",colnames(myContrast)[i],"
         .txt",sep=''),quote=FALSE,row.names=T,sep="\t")
138
139  temp<-data.frame(Ids=deTable$zunla1_ids,LFC=deTable$LFC)
140  temp$color<-apply(temp, 1, function(x) if(x[2]>0) paste("#f79797")
         else if(x[2]<0) paste("#97b1f7") else paste("#ffffff"))
141  #temp$LFC<-NULL
142  write.table(na.omit(temp[,c(1,3)]),file=paste("Kegg_enrichment/DEGs_",
         colnames(myContrast)[i],"_kg.txt",sep=''),quote=FALSE,row.names=F,
         sep="\t",col.names = F)
143  write.table(na.omit(temp),file=paste("Kegg_enrichment/DEGs_",colnames(
         myContrast)[i],"_lf_kg.txt",sep=''),quote=FALSE,row.names=F,sep="\t
         ",col.names = F)
144
145  DEGenes<-data.frame(DE_genes[,1])
146  colnames(DEGenes)<-c("isDE")
147  DEGenes<-subset(DEGenes,abs(isDE)==1)
148  DEGenes<-data.frame(rep(colnames(myContrast)[i],length(DEGenes$isDE)),
          rownames(DEGenes))
149  if(!dir.exists(paste("GO_enrichment/",colnames(myContrast)[i],sep=""))
         )
150  dir.create(paste("GO_enrichment/",colnames(myContrast)[i],sep=""))
151  write.table(DEGenes,paste("GO_enrichment/",colnames(myContrast)[i],"/"
         ,colnames(myContrast)[i],"_AllDegs.txt",sep=''),quote = FALSE,sep =
          "\t",col.names = FALSE,row.names = FALSE)
152
153  DEGenes<-data.frame(DE_genes[,1])
```

```
154    colnames(DEGenes)<-c("isDE")
155    DEGenes<-subset(DEGenes,isDE==1)
156    DEGenes<-data.frame(rep(colnames(myContrast)[i],length(DEGenes$isDE)),
           rownames(DEGenes))
157    write.table(DEGenes,paste("GO_enrichment/",colnames(myContrast)[i],"/"
       ,colnames(myContrast)[i],"_UPR.txt",sep=''),quote = FALSE,sep = "\t
       ",col.names = FALSE,row.names = FALSE)
158
159    DEGenes<-data.frame(DE_genes[,1])
160    colnames(DEGenes)<-c("isDE")
161    DEGenes<-subset(DEGenes,isDE==-1)
162    DEGenes<-data.frame(rep(colnames(myContrast)[i],length(DEGenes$isDE)),
           rownames(DEGenes))
163    write.table(DEGenes,paste("GO_enrichment/",colnames(myContrast)[i],"/"
       ,colnames(myContrast)[i],"_DOWNR.txt",sep=''),quote = FALSE,sep = "
       \t",col.names = FALSE,row.names = FALSE)
164    }
165
166 }
```

B. Resultados

B.1. Anotación génica

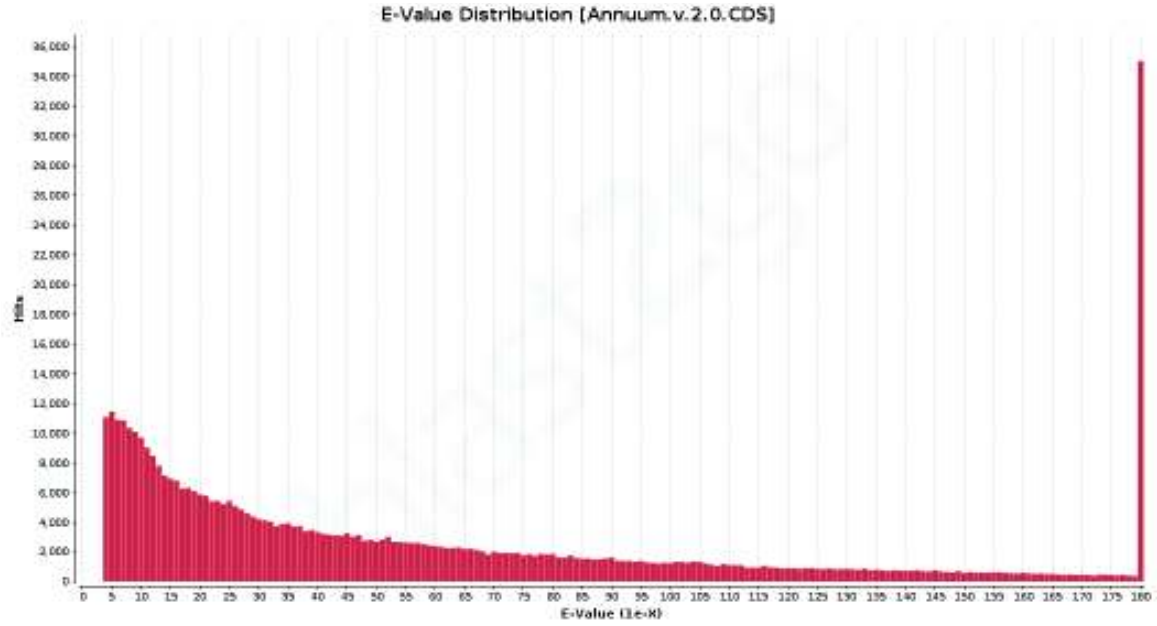


Figura B.38. Cantidad de genes expresados diferencialmente en los cinco contrastes diseñados para las muestras de Chiltepín y Serrano a 20 y 60(68) DDA.

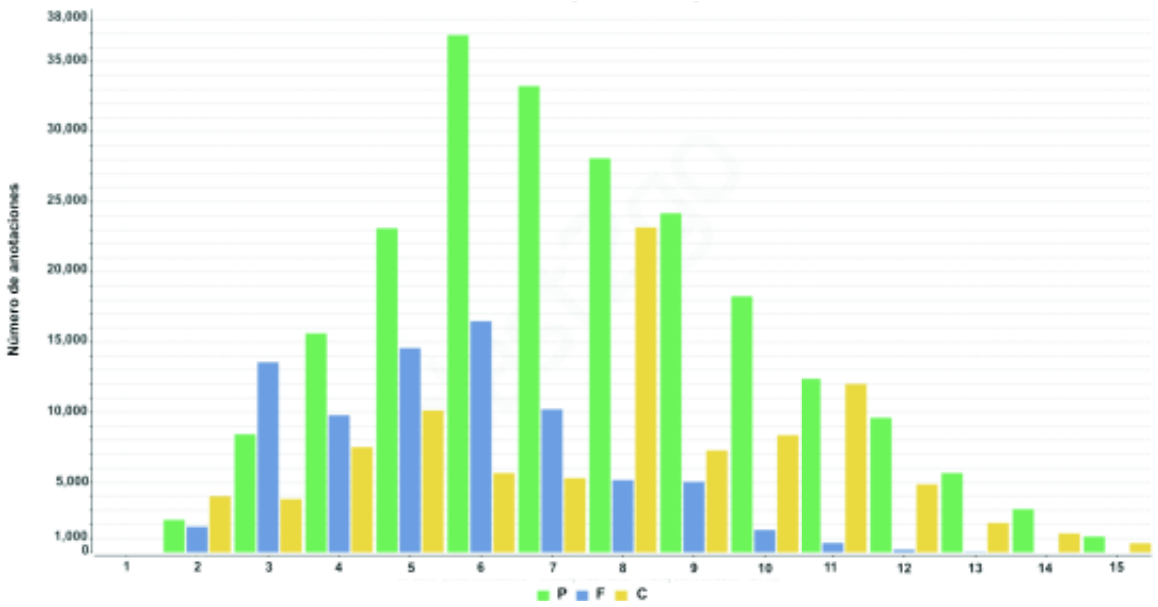


Figura B.39. Cantidad de genes expresados diferencialmente en los cinco contrastes diseñados para las muestras de Chiltepín y Serrano a 20 y 60(68) DDA.