



Centro de Investigación Científica de Yucatán, A.C.

Posgrado en Ciencias Biológicas

"ANÁLISIS TRANSCRIPTÓMICO DEL DESARROLLO
FLORAL DE CIRICOTE BLANCO (*CORDIA
SEBESTENA*)"

Tesis que presenta

ORLANDO EMANUEL OSORIO PINELO

En opción al título de

MAESTRO EN CIENCIAS
(Ciencias Biológicas: opción biotecnología)

Mérida, Yucatán, México

2023

CENTRO DE INVESTIGACIÓN CIENTÍFICA DE YUCATÁN, A. C.

POSGRADO EN CIENCIAS BIOLÓGICAS



RECONOCIMIENTO

Por medio de la presente, hago constar que el trabajo de tesis de **Orlando Emanuel Osorio Pinelo**, titulado “**Análisis transcriptómico del desarrollo floral de ciricote blanco (*Cordia sebestena*)**”, fue realizado en el laboratorio de Biotecnología computacional de la Unidad de Biotecnología del Centro de Investigación Científica de Yucatán, A.C. bajo la dirección de la Dra. Elsa Góngora Castillo y la Dra. Miriam Ferrer Ortega, dentro de la opción de Biotecnología, perteneciente al Programa de Posgrado en Ciencias Biológicas de este Centro.

Atentamente



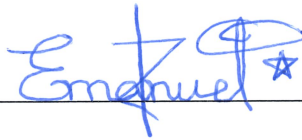
Dra. Celia Hernández Zepeda
Directora de Docencia

Mérida, Yucatán, México, a 02 de marzo de 2023

DECLARACIÓN DE PROPIEDAD

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

Firma: _____



Nombre: Orlando Emanuel Osorio Pinelo

Este trabajo se llevó a cabo en la Unidad de Biotecnología del Centro de Investigación Científica de Yucatán, A.C., y forma parte del proyecto titulado “**Efecto de la domesticación en árboles multipropósito de la Península de Yucatán: Ecología y genómica funcional de *Cordia dodecandra*** convocatoria CB2017-2018 proyecto A1-S-30471 bajo la dirección de la **Dra. Miriam M. Ferrer Ortega**.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca No. 1067627.

Al Centro de Investigación Científica de Yucatán y a la Unidad de Tecnologías de Información y Comunicación, por el asesoramiento para el uso del equipo de súper cómputo “Hobón”, recurso básico para el análisis computacional de este trabajo.

A la unidad de Biotecnología por las instalaciones prestadas para el desarrollo de la parte experimental de este trabajo.

A mis asesoras, Dra. Elsa Góngora Castillo y Dra. Miriam Ferrer Ortega, por la mentoría, paciencia y los conocimientos compartidos a lo largo de este proyecto.

Al comité tutorial: Dra. Ivón Mercedes Ramírez Morillo, Dr. Jorge Humberto Ramírez Prado, Dr. Ramón Pacheco Arjona por los consejos, revisiones y observaciones durante el desarrollo del trabajo.

A la M. en C. Adriana Quiroz Moreno por su dedicación y compromiso en la parte técnica de este trabajo, sus consejos y mentoría fueron clave para el proyecto.

A mis compañeros de laboratorio: M. en C.C. Cynthia Soto, M. en C.B. Daniel Macías, M. en C.B Perla Analuz, M. en C. Jorge Tzec y M. en C.B. Juan Martínez por sus consejos y retroalimentación durante el desarrollo de este trabajo.

Al M. en C. Ivan isidro Cordova Lara por sus sugerencias y consejos técnicos en los protocolos de laboratorio.

Al Dr. en Arq. Manuel Alfonso Barrero por su mentoría, objetividad y respaldo a lo largo de este proyecto.

Productos académicos

2022. Presentación de poster: Transcriptomical analisis of floral development of ciricote (*Cordia sebestena*). En coautoría con: Dra. Elsa Góngora-Castillo, Dra. Miriam Ferrer-Ortega, M. en C. Adriana Quiroz-Moreno, M. C. Norka Fortuny-Fernández. En el XXXIII Congreso Nacional de Bioquímica con sede en Mérida, Yucatán, México, del 16-21 de octubre.

DEDICATORIAS

A mis padres Guadalupe y Orlando, la herencia más grande que me dieron fue el conocimiento, su amor, guía y consejo, hoy, se ve materializada con esta tesis porque siempre me enseñaron a perseguir mis anhelos.

A mi hermana Irma, las palabras de aliento y la comprensión trascienden distancias, que este logro sea reflejo de los tuyos no importa a donde estés.

A Vianey, gracias por darme fuerzas cuando más lo necesitaba, por no dejarme darme por vencido, por siempre estar en mi vida.

A mis padrinos Lucy y Arturo, reitero que el mundo necesita más gente como ustedes, este trabajo es resultado de esa semilla que sembraron.

A la Dra. Silvia Hernández, siempre ha sido una inspiración desde que me embarqué en este camino, si se pudo y usted siempre lo supo.

A la Dra. Lety Peraza, los consejos y la paz que transmite me ayudaron a no perder la esperaza.

A Bartolomé Chi, dejaste tu contribución en este trabajo, pero sobretodo nos recordaste lo importante que es vivir cada día como el último.

ÍNDICE

| | |
|---|----|
| INTRODUCCIÓN..... | 1 |
| CAPÍTULO I | 5 |
| ANTECEDENTES | 5 |
| 1.1 Familia Cordiaceae | 5 |
| 1.2 Fenómeno distílico | 7 |
| 1.3 Desarrollo Floral | 14 |
| 1.4 Transcriptómica..... | 21 |
| JUSTIFICACIÓN..... | 30 |
| HIPÓTESIS..... | 31 |
| OBJETIVOS..... | 31 |
| OBJETIVO GENERAL | 31 |
| Objetivos particulares..... | 31 |
| ESTRATEGIA EXPERIMENTAL | 32 |
| CAPÍTULO II | 33 |
| MATERIALES Y MÉTODOS | 33 |
| 2.1 Caracterización del transcriptoma de <i>Cordia sebestena</i> | 33 |
| CAPÍTULO III | 42 |
| RESULTADOS | 42 |
| 3.1 Caracterización del transcriptoma de <i>Cordia sebestena</i> | 42 |
| 3.1.1 Pureza e integridad de RNA en muestras de <i>C. sebestena</i> | 42 |
| CAPITULO IV..... | 64 |
| DISCUSIÓN | 64 |
| 4.1 Caracterización del transcriptoma floral de <i>Cordia sebestena</i> | 64 |
| CAPITULO V..... | 69 |

| | |
|--|-----------|
| CONCLUSIONES GENERALES Y PERSPECTIVAS | 69 |
| 5.1 Conclusiones | 69 |
| 5.2 Perspectivas..... | 71 |
| REFERENCIAS | 72 |
| ANEXOS. | 90 |
| ANEXO I. RNeasy® Plant Mini Kit QIAGEN..... | 90 |

INDICE DE FIGURAS

Figura 1.1. (A) Hercogamia de aproximación, los estigmas están por encima de las anteras, y son los primeros en contactar a los polinizadores a su ingreso en la flor. (B) Hercogamia reversa, las anteras están por encima de los estigmas (Lloyd y Webb, 1986; Ganders, 1979). 8

Figura 1.2. Diagrama simplificado de la posición del pistilo (verde) y estambres (negro y amarillo) en el sistema de distilia (Schopfer et al., 1999). 8

Figura 1.3. Distribución filogenética de los géneros de angiospermas con especies distílicas. Órdenes a los que pertenecen las familias que contienen especies distílicas: (Boraginaceae no pertenece a ningún orden) incluye Asparagales (Amaryllidaceae e Iridaceae), Fabales (Fabaceae), Oxalidales (Connaraceae y Oxalidaceae), Malpighiales (Erythroxylaceae, Hypericaceae, Linaceae y Passifloraceae), Malvales (Malvaceae), Myrtales (Lythraceae), Saxifragales (Saxifragaceae), Santalales (Santalaceae y Schoepfiaceae), Caryophyllales (Plumbaginaceae y Polygonaceae), Ericales (Polemoniaceae y Primulaceae), Gentianales (Gentianaceae, Gelsemiaceae y Rubiaceae), Lamiales (Acanthaceae, Lamiaceae and Oleaceae), Boraginaceae y Asterales (Menyanthaceae) (tomado de Naiki, 2012). 9

Figura 1.4. Desarrollo floral. Las flores se desarrollan a partir de los meristemas florales, con frecuencia surgen en los laterales de un meristemo de inflorescencia. (A). El meristemo floral es un pequeño grupo de células en forma de domo y está conformado de varias capas celulares (B). Las células en los laterales del meristemo floral proliferan para formar primero el primordio de los sépalos (C), seguidos por el primordio de los pétalos (D). Finalmente, cada una de las estructuras florales se define, consistiendo en sépalos, pétalos, estámenes y carpelos (E) (modificado de Zik y Irish, 2003). 14

| | |
|--|----|
| Figura 1.5. Modelo ABC de desarrollo floral en <i>A. thaliana</i> . A) verticilos florales dispuestos alrededor del meristemo floral y los órganos que dichos verticilos originan. B) distintas clases de genes (A, B y C), verticilos en los que actúan y que estructura generan (tomado de Mendoza y Encina, 2018). | 16 |
| Figura 1.6. El modelo clásico ABCE especifica cuatro órganos florales morfológicamente discretos: los sépalos se producen donde sólo actúa la función A, los pétalos se producen con la expresión combinada de las funciones A y B, los estambres se producen con la expresión combinada de las funciones B y C, y los carpelos se producen donde sólo actúa la función C (tomado de Chanderbali et al., 2016). | 17 |
| Figura 1.7. Posibles patrones de desarrollo para la altura de la antera y el estigma en los morfos de estilo largo y corto de las especies distílicas (Richards y Barrett, 1992)..... | 18 |
| Figura 1.8. Patrones de desarrollo para la altura de la antera y del estigma para (A)(B) <i>Cordia sebestena</i> y (C)(D) <i>Cordia boissieri</i> . Cuadros azules: órganos de morfo largo; cuadros rojos: órganos de morfo corto (C. sebestena Ferrer (comunicación personal); C. boissieri Modificado de Cohen, 2019). | 20 |
| Figura 1.9. Esquema de los pasos a seguir en la técnica de RNA-seq (tomado de Ovalle Rivera, 2018). | 23 |
| Figura 1.10. Haplotipos asociados a la morfología dominante brevistilo (S/s) y longistilo (s/s). El S-locus de 278 kb contiene cinco genes que no están presentes en las plantas de morfo largo y están ausentes en el haplotipo s: CCMT, GLOT, CYPT (CYP374A50), PUMT y KFBT. Las regiones de la izquierda (azul) y derecha (naranja) contienen genes expresados en ambos morfos. Tomado de McClure, 2016. | 27 |
| Figura 1.11. Diagrama de Venn resultante de la comparación de los genes diferencialmente expresados por estadio y tejido en <i>Lithospermum multiflorum</i> . Tomado de Cohen (2016)..... | 28 |

| | |
|--|----|
| Figura 1.12. Diagrama de flujo de trabajo para la obtención del transcriptoma de <i>Cordia sebestena</i> | 32 |
| Figura 2.1. Ubicación del área de estudio..... | 34 |
| Figura 2.2. Diseño experimental asociado con el morfo floral y estadio de desarrollo de <i>Cordia sebestena</i> con número total de muestras..... | 35 |
| Figura 2.3. Diagrama de flujo para la liofilización de muestras pertenecientes a <i>Cordia sebestena</i> desarrolladas en un ámbito silvestre usando un equipo LABCONCO Freezone 4.5 Liter..... | 36 |
| Figura 3.1. Geles de agarosa al 1% teñidos con bromuro de etidio. Muestras de RNA de <i>Cordia sebestena</i> después del tratamiento con DNase I, Amplification grade de la compañía Invitrogen..... | 44 |
| Figura 3.2. Electroferograma de la prueba de integridad RIN (RNA integrity number) Agilent 2100 Bioanalyzer®. A) Muestra B9 (Gineceo de botón pre-antes de morfo Brevistilo) con un RIN de 9.2 B) Muestra B3 con un RIN de 4.3 (Gineceo de antes de morfo Brevistilo). FU: Intensidad de fluorescencia, Nt: tamaño del producto en nucleótidos..... | 49 |
| Figura 3.3. Diseño experimental. En rojo se señala las muestras elegidas para el ensamblado de novo de los transcriptomas de <i>Cordia sebestena</i> | 55 |
| Figura 3.4. Transcritos conservados posterior al agrupamiento a través de CD-Hit v.4.8.1 a través del equipo de súper cómputo Hobón en los transcriptomas de los morfos brevistilo y longistilo de <i>Cordia sebestena</i> | 56 |

Figura 3.5. Porcentaje de secuencias anotadas contra diferentes bases de datos de proteínas de plantas de los proteomas predichos de los morfos brevistilo y longistilo de *Cordia sebestena* a partir de sus transcriptomas ensamblados.....58

Figura 3.6. Términos de GO asignados en la categoría de Función Molecular en los transcriptomas de los A) morfos brevistilo y B) longistilo *Cordia sebestena* a partir de la anotación funcional realizada con el proteoma predicho de *Arabidopsis thaliana* (TAIR11).
..... 59

Figura 3.7. Términos GO asignados en la categoría de Procesos Biológicos para el transcriptoma de los morfos A) brevistilo y B) longistilo de *Cordia sebestena* a partir de la anotación funcional realizada con el proteoma de *Arabidopsis thaliana* (TAIR11).....60

INDICE DE TABLAS

| | |
|--|----|
| Tabla 1.1. Resultados de los análisis filogenéticos de las especies distílicas, incluyendo el número de especies, las regiones y la longitud alineada de las regiones de DNA, en pares de bases (pb). Tipo ancestral de hercogamia señalado, para cada especie, mediante la reconstrucción de caracteres ancestrales en los marcos de referencia de máxima parsimonia (MP) y máxima verosimilitud (ML) tomado de (Cohen, 2019). | 11 |
| Tabla 1.2. Colección de especies distílicas y sus patrones de altura en estigma y anteras. (Modificado de Cohen, 2019)..... | 19 |
| Tabla 1.3. Número de transcritos asociados a tipos de términos de la Ontología Genética (GO) durante las etapas de desarrollo de los órganos florales en <i>Lythospermum multiflorum</i> | 29 |
| Tabla 2.1. Requerimientos de la compañía Novogene para secuenciación de RNA..... | 37 |
| Tabla 2.2. Muestras pertenecientes a árboles de <i>Cordia sebestena</i> que fueron empleadas para el ensamblado del transcriptoma del morfo brevistilo y longistilo. | 39 |
| Tabla 3.1. Muestras y tejidos empleados para la extracción de RNA total de <i>Cordia sebestena</i> | 42 |
| Tabla 3.2. Concentración y valores de absorbancia para las muestras de RNA total de <i>Cordia sebestena</i> así como la concentración y pureza requerida por la empresa Novogene..... | 45 |
| Tabla 3.3. Resultados de pureza y concentración de RNA, así como el número de integridad de RNA obtenido en muestras de <i>Cordia sebestena</i> | 47 |

| | |
|---|----|
| Tabla 3.4 Lecturas y estadísticos resultado de la secuenciación en muestras de estructuras florales de <i>Cordia sebestena</i> colectadas del parque Gran Santa Fe..... | 50 |
| Tabla 3.5. Resultados del análisis de calidad de lecturas con la herramienta FastQC en muestras de estructuras florales de <i>Cordia sebestena</i> | 52 |
| Tabla 3.6. Muestras elegidas para el ensamblado de novo del transcriptoma floral de <i>Cordia sebestena</i> y sus estadísticos..... | 54 |
| Tabla 3.7. Datos estadísticos de los transcriptomas ensamblados de novo de los morfos florales longistilo y brevistilo de <i>Cordia sebestena</i> | 55 |
| Tabla 3.8. Porcentajes de anotación en los transcriptomas del morfo brevistilo y longistilo de <i>Cordia sebestena</i> contra bases de datos de proteínas de plantas a través de BLAST versión 2.11.0..... | 57 |
| Tabla 3.9. Genes únicos identificados en transcriptomas de los morfos brevistilo y longistilo de <i>Cordia sebestena</i> | 62 |

RESUMEN

La anacahuita (*Cordia sebestena*) es un árbol distílico que presenta dos morfos florales, nombrados de acuerdo con la longitud del pistilo: longistilo (largo) y brevistilo (corto). Esta especie es también conocida como ciricote blanco. Este es el primer estudio de transcriptoma empleando tecnologías de secuenciación de nueva generación para obtener información molecular de *Cordia sebestena* en Yucatán e identificar genes implicados en el desarrollo floral de ambos morfos. Se recogieron por triplicado muestras compuestas de estambres y de pistilos para tres estadios de desarrollo: temprano (yemas cerradas), medio (flor pre-antesis) y tardío (flor en antesis) para cada morfo. De cada muestra se aisló el RNA total y se secuenciaron los transcriptomas utilizando la plataforma Illumina en modo paired-end para obtener 719 millones de lecturas de 150 pb de longitud para cada muestra secuenciada. Se realizó un ensamblado *de novo* para cada morfo floral, obteniendo un total de 121,871 y 126,260 transcritos, respectivamente. Los transcriptomas fueron anotados funcionalmente por homología de secuencias utilizando diversas bases de datos como referencia identificando más de 35% de los transcritos ensamblados. Se identificaron las funciones de grupos de genes utilizando la base de datos de Gene Ontology y esto reveló que mecanismos relacionados a regulación de fitohormonas, crecimiento celular, así como, respuesta ante estrés están activos durante el desarrollo floral.

ABSTRACT

Anacahuita (*Cordia sebestena*) is a distyllic tree with two floral morphs, named according to the length of the pistil: long and short. This species is also known as the white ciricote. This is the first transcriptome study using next-generation sequencing technologies to obtain molecular information from *Cordia sebestena* in Yucatan and to identify genes related to the floral development of both morphs. Composite samples of stamens and pistils were collected in triplicate for three developmental stages: early (closed buds), middle (pre-anthesis flower), and late (anthesis flower) in both long and short morphs. For each composite sample, total RNA was isolated and the transcriptomes were sequenced using the Illumina platform in paired-end mode to obtain 719 million reads of 150 bp length for each sequenced sample. A *de novo* assembly transcriptome was performed for each floral morph, obtaining a total of 121,871 and 126,260 transcripts, respectively. The transcriptomes were functionally annotated based on sequence homology using several databases as reference, identifying more than 35% of the assembled transcripts. Gene functions were identified using the Gene Ontology database and revealed that phytohormone regulation, cell growth, and stress response mechanisms are active during the floral development.

INTRODUCCIÓN

La distilia es un sistema reproductivo que se caracteriza por poseer dos morfos florales, los cuales presentan una variación en el posicionamiento de las anteras y los estigmas, sin perder su condición hermafrodita y, por lo general, poseen un mecanismo de incompatibilidad esporofítica. Los morfos reciben distintos nombres relacionados con la altura del pistilo: el morfo floral largo, con un pistilo largo y estambres cortos se denomina también “pin” o “longistilo”, mientras que el morfo corto, con pistilo corto y estambres largos es llamado “thrum” o “brevistilo” (Barrett, 1990, Kohn y Barrett, 1992; Lloyd y Webb, 1986). La altura de las anteras en uno de los morfos es la misma que la del estigma en el morfo opuesto, a esta condición se le conoce como hercogamia recíproca (Cohen, 2019). Este sistema ha surgido de manera independiente en 27 familias de angiospermas (Ganders, 1979; Barrett y Shore, 2008; Naiki, 2012; Cohen, 2019). Cohen (2019) comparó los patrones florales de 15 especies de plantas con distilia que representaban 12 orígenes distílicos distintos y encontró que por lo menos 10 familias distílicas tienen como carácter ancestral la hercogamia de aproximación, es decir, presentan los estigmas por encima de las anteras (*sensu* Lloyd y Webb, 1986), mientras que en cuatro familias el carácter ancestral es la hercogamia reversa, es decir el estigma está situado por debajo de las anteras (*sensu* Lloyd y Webb, 1986). Una de estas cuatro familias es Cordiaceae del orden Boraginales. En varias especies que tienen como carácter ancestral la hercogamia de aproximación se ha estudiado su desarrollo floral. Analizar cómo es el desarrollo floral de los morfos longistilo y brevistilo en las familias con hercogamia reversa puede ayudar a entender la diversidad de mecanismos que distinguen las variantes de los sistemas distílicos.

El desarrollo floral es un proceso que incluye cambios en las tasas y patrones de división celular del meristemo que originan los diferentes tejidos que conforman la flor hasta que se forman las semillas (Alvarez-Buylla *et al.*, 2010). Este proceso consta en términos simples de a) la formación del meristemo floral, b) el establecimiento de las identidades únicas de los órganos y c) la diferenciación y crecimiento de las estructuras florales. En *Arabidopsis thaliana* se caracterizaron 20 etapas, en la que la 13 corresponde a la anthesis floral (Alvarez-Buylla *et al.*, 2010; Zik y Irish, 2003). El desarrollo floral se ha estudiado desde diferentes enfoques. Desde el punto de vista morfológico se analiza la filotaxia floral, que es el modo en el que se disponen las estructuras florales alrededor del eje floral principal, las veces que distintos órganos se presentan en la flor y su inclinación respecto al eje (Endress, 2006). El enfoque molecular estudia los diversos genes que regulan desde la diferenciación del

meristemo floral hasta la formación de semillas, incluyendo el inicio y terminación de la floración, los diferentes patrones florales de acuerdo con la especificidad de los órganos florales y su morfogénesis (Thomson y Wellmer, 2019). Los estudios en mutantes de *Arabidopsis* y *Antirrhinum* han conducido a proponer el modelo ABC molecular de especificación de la identidad de los órganos florales (Coen y Meyerowitz, 1991; Weigel y Meyerowitz, 1994). Este modelo postula tres clases de función génica: A, B y C, que actúan de forma combinada para especificar de manera única cada tipo de órgano floral -sépalos, pétalos, androceo y gineceo (Bowman et. al, 2012). Las flores desarrollan órganos reproductivos que pueden ser masculinos o femeninos. El gineceo (órgano sexual femenino) se ubica al centro de la flor y alberga los óvulos en formación. Está conformado por la fusión de uno o más carpelos y se encuentra cubierto por tejido estigmático (estigma) el cual se compone por una sola capa de células papilares alargadas. La presencia de los genes de tipo C permite el desarrollo del gineceo (Romero-Arias, et. Al, 2010). Rodeando al estigma y por debajo de este se encuentra el androceo (órgano sexual masculino). Está conformado por una estructura alargada y fina llamada filamento que en su interior contiene el tracto transmisor (el cual se ubica en su centro) y en su extremo por una parte ensanchada (fértil) conocida como antera (Simonini y Ostergaard, 2018). La presencia de los genes B+C regulan la expresión del androceo (Fosket, 1994). Además de la identidad sexual de los órganos, diferentes genes pueden estar involucrados en un crecimiento diferencial de los órganos masculinos y femeninos, lo que puede condicionar el tipo de hercogamia y la presencia de distilia, como se discutirá más adelante.

En plantas con flores distílicas se han descrito 12 patrones de desarrollo morfo-específico responsables de las variaciones en las alturas y posicionamiento que presentan los órganos masculinos y femeninos previos a la antesis (Cohen, 2019). Estos patrones de desarrollo floral se describen con base en las tasas de crecimiento del estambre y pistilo desde la formación de las yemas florales hasta la antesis floral. Por lo tanto, el potencial de crecimiento de las distintas estructuras (androceo y gineceo) permite un gran número de combinaciones potenciales de patrones de crecimiento específicos de cada morfo y especie para estos órganos sexuales. En la familia Cordiaceae el patrón de crecimiento de los órganos sexuales se caracteriza por presentar tasas de crecimiento similares del androceo y gineceo en etapas tempranas de desarrollo en ambos morfos. Posteriormente, el androceo detiene su elongación en la etapa media a diferencia del gineceo que continúa proyectándose, quedando por encima de los órganos masculinos en el morfo longistilo. Por otro lado, en el morfo brevistilo las tasas de crecimiento del gineceo se mantienen

constantes en etapas tempranas y medias, en la etapa media de desarrollo el androceo incrementa su tasa de crecimiento quedando por encima del gineceo (Cohen, 2019). En un estudio preliminar en *Cordia sebestena* y *Cordia dodecandra*, Ferrer (comunicación personal) encontró el mismo patrón de desarrollo. Este patrón de desarrollo floral es similar al descrito por Cohen, (2012) de *Lithospermum multiflorum*, una especie de la familia hermana Boraginaceae, quien posteriormente asoció los genes involucrados en el desarrollo floral de la especie con herramientas transcriptómicas (Cohen, 2016).

La transcriptómica es el estudio de todas las moléculas de mRNA en una célula. Es una herramienta capaz de identificar genes relacionados con los procesos biológicos. Permite realizar comparaciones en condiciones contrastantes (ejemplo: diferentes tipos de células o tejidos, estímulos ambientales, enfermedad o salud, estadíos de desarrollo, entre otros). Esta comparación permite comprender la base molecular de la variación fenotípica entre diversos tratamientos para medir la ausencia o presencia de genes específicos o su expresión diferencial (Gironella, 2010). La transcriptómica se apoya en las herramientas de secuenciación masiva NGS (por sus siglas en inglés; Next-Generation Sequencing) así como los análisis bioinformáticos, para evaluar la expresión diferencial de los genes. Así, los datos de los perfiles de expresión aportan información necesaria para identificar genes diferencialmente expresados que podrían estar asociados a respuestas celulares concretas (Franch, 2019).

En el estudio transcriptómico del patrón de desarrollo floral en *Lithospermum multiflorum* realizado por Cohen en el 2016 se analizaron tres etapas de desarrollo floral (yema floral, flor pre-antesis y antesis floral) y los transcritos del gineceo y de la corola y androceo. El transcriptoma floral fue ensamblado incluyendo 97,603 genes putativos y 265,144 isoformas. Se identificaron 12 genes diferencialmente expresados (DE) en el gineceo en todas las etapas de desarrollo con un número reducido de coincidencias entre la fase de desarrollo de yema floral y flor pre-antesis (dos) y la fase pre-antesis y antesis floral (cuatro). Los genes con mayor expresión se encontraban en el gineceo del morfo brevistilo y en la corola y androceo del morfo longistilo. Los genes sobreexpresados en la corola y androceo del morfo brevistilo estaban relacionados en la división celular y dinamismo de la actina. En cambio, los genes sobreexpresados en el desarrollo del gineceo para el morfo longistilo controlan el crecimiento celular a través de la regulación de la auxina. En conjunto, este patrón de expresión sugiere que diferentes genes están implicados en la elongación y el crecimiento de los órganos sexuales en el botón floral en desarrollo.

Considerando que los patrones de desarrollo floral de *C. sebestena* y *L. multiflorum* son similares, esperaríamos que los genes del morfo floral largo y corto se encuentren diferencialmente expresados. El presente trabajo pretende aportar elementos para realizar un análisis de expresión diferencial de los dos morfos florales y la identificación global de los transcritos que cada uno de ellos tiene, para que posteriormente se puedan construir perfiles de expresión para los morfos florales de los estadios de yema floral, flor pre-antesis y flor en antesis en el desarrollo floral de *Cordia sebestena*.

CAPÍTULO I

ANTECEDENTES

1.1 Familia Cordiaceae

Tradicionalmente la familia Boraginaceae se consideró constituida por cuatro subfamilias: *Ehretioideae*, *Cordioideae*, *Heliotropioideae* y *Boraginoideae* (Johnston, 1948). Recientemente, diversos estudios moleculares han permitido considerar que éstas deben ser elevadas al rango de familia (Cordiaceae, Ehretiaceae, Heliotropiaceae y Boraginaceae) e incluirlas junto a Hydrophyllaceae y Lennoaceae en el orden Boraginales (Gottschling *et al.*, 2001).

En este contexto, Cordiaceae es un grupo monofilético sustentado por datos moleculares y apomorfías morfológicas, como la presencia de endocarpos enteros, cuatro lóbulos estigmáticos y cotiledones plegados (Gottschling *et al.*, 2005). Sin embargo, la extrema diversidad morfológica de *Cordia sensu stricto* ha provocado intensa controversia en su delimitación taxonómica. Borhidi y colaboradores (1988) hacen una revisión de este género de distribución pantropical e indican que fue Britton (1925) el primero en dividir a *Cordia L.* en varios géneros, siendo uno de ellos *Varronia P. Browne*.

La heterostilia es una característica bien conocida en las especies de Cordiaceae es una de las familias que contiene el mayor número de especies de plantas con heterostilia, y está presente en al menos nueve géneros de los 14 de esta familia (Ganders, 1979; Opler *et al.*, 1975; Taroda, 1984; Tomlinson, 1974).

El género más grande es *Cordia L.*, con aproximadamente 350 especies distribuidas por las zonas tropicales y subtropicales del mundo, e incluye casi todas las especies de la subfamilia (Miller, 2001). Incluye árboles con hojas de borde entero e inflorescencias de muchas flores con un eje conspicuo. Por otro lado, las especies suelen ser subarbustos, arbustos o arbustos escandentes con hojas dentadas o con márgenes serrados e inflorescencias poco floridas con un eje poco visible, que produce inflorescencias en forma de espiga o globosas (Estrada 1995; Melo y Stapf 2014). Especies, en su mayoría restringidas a regiones cálidas del Neotrópico, en América, desde el sur de Estados Unidos hasta Argentina, pero con centros de diversidad en Brasil, el norte de Sudamérica y México (Miller y Gottschling 2007; Miller 2013).

1.1.1 Ciricote (*Cordia sebestena* L.)

1.1.2 Taxonomía del ciricote

Reino: Plantae

División: Magnoliophyta

Clase: Magnoliopsida

Orden: Boraginales

Familia: Cordiaceae

Género: *Cordia*

Especie: *Cordia sebestena*

1.1.3 Características del ciricote blanco

Árboles pequeños de hasta 8 m de alto, con ramas glabrescentes; plantas hermafroditas. Hojas verde oscuras persistentes, ovadas, 9–20 cm de largo y 6–12 cm de ancho, ápice agudo, base redondeada a obtusa, escabrosas en el haz, casi glabras en el envés; pecíolos 1.5–3.8 cm de largo, pubescentes. Inflorescencias cimosas, subterminales, 6.5–12 cm de ancho; flores distilas, pedicelos 4–6 mm de largo; cáliz tubular-campanulado, 11–24 mm de largo, puberulento a glabro, generalmente 2-lobado pero a veces con hasta 5 lobos irregulares; corola infundibuliforme, 30–58 mm de largo, anaranjado brillante, 5–7-lobada, los lobos ovados a muy ampliamente ovados, 8–10 mm de largo; estambres 5–7, filamentos 22–33 mm de largo, glabros, anteras oblongas, 2.8–3.8 mm de largo; ovario cónico, 1.5–3 mm de largo, estilo 13–35 mm de largo, estigmas claviformes. Fruto drupáceo, presenta coloración verde cuando joven y blanco al madurar, completamente envuelto por el cáliz creciente; hueso ovoide y 1.9–4 mm de largo (Durán, 2000).

En México, se distribuye en el norte de Veracruz, Chiapas, Península de Yucatán y Tabasco. Se utiliza como árbol de ornato, sus frutos son comestibles, pero no tienen buen sabor, sus hojas rasposas se utilizaban para lavar trastos.

Crece en todo el litoral costero hasta unos 500 metros sobre el nivel del mar (m s.n.m.), formando parte importante de la vegetación de duna costera, aunque algunas veces se le puede encontrar tierra adentro, principalmente en poblados donde se cultiva esporádicamente. Es una especie heliófita (demandante de luz para crecer), aunque sobrevive bajo el dosel cerrado. La polinización de las flores se debe principalmente a

colibríes y la dispersión de semillas se les atribuye principalmente a murciélagos (Arellano *et al.*, 2003).

En el estado de Yucatán el ciricote blanco es una especie de gran valor, por ser un elemento ecológicamente importante en la composición florística y por proporcionarle a los habitantes de la zona su madera (de exquisito veteado y dureza), sus hojas ásperas para ser usadas como lija, sus vistosas flores, así como sus abundantes frutos que suelen consumirse por las propiedades medicinales que se les atribuye especialmente como antitusivo (Rico *et al.*, 1991).

El ciricote blanco es considerado en la región un árbol de uso múltiple. Hoy en día es una especie que tiene densidades muy bajas en las poblaciones silvestres, debido principalmente a su sobreexplotación y por la destrucción de su ecosistema. Su población prácticamente ha quedado confinada a zonas urbanas y poblaciones rurales, específicamente como árbol cultivado con fines de ornato. *Cordia sebestena* pertenece a la sección infragenérica de *Cordia* (Miller, 1985), y al clado monofilético *Sebestena* (Gottschling *et al.*, 2001). Es exclusiva de la vertiente Atlántica de América (Miller, 1985; Opler *et al.*, 1975; Percival, 1974).

1.2 Fenómeno distílico

En las plantas con flor (angiospermas) son frecuentes los polimorfismos florales que pueden y no estar necesariamente asociados a diferenciación taxonómica. El término distilia es utilizado para describir aquellas flores hermafroditas entomófilas con dos morfos florales, en las que la autofecundación se ve obstaculizada por las posiciones relativas de los órganos sexuales. Las flores tienen una separación espacial significativa del polen y de los estigmas, debido a que algunos individuos presentan flores con hercogamia de aproximación y otros presentan flores con hercogamia reversa, con las anteras de cada tipo de flor situadas a la misma altura que los estigmas del otro tipo (Herrera, 1996; Lloyd y Webb, 1986; Barret, 2019).

La hercogamia de aproximación se caracteriza por tener el estigma situados por encima de las anteras mientras que la hercogamia reversa posee el estigma situado por debajo de las anteras como se pueda apreciar en la figura 1.1.

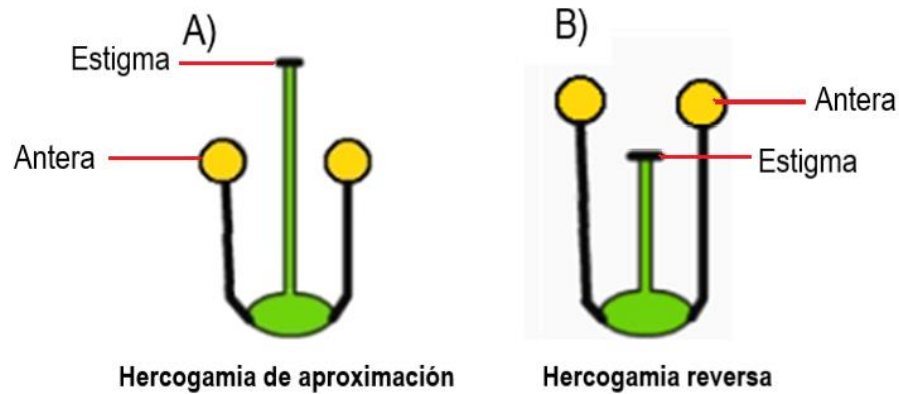


Figura 1.1. (A) Hercogamia de aproximación, los estigmas están por encima de las anteras, y son los primeros en contactar a los polinizadores a su ingreso en la flor. (B) Hercogamia reversa, las anteras están por encima de los estigmas (Lloyd y Webb, 1986; Ganders, 1979).

La distilia afecta las alturas en las que se posicionan los órganos sexuales de la flor (estambres y pistilos) sin que ésta pierda la condición hermafrodita. La principal característica de las plantas distilas es la presencia de dos morfos en las poblaciones, la aparición de estos morfos está ligada a un sistema de incompatibilidad llamado heteromórfico (dos morfos). Dichas plantas son autoincompatibles así como también lo son los mismos morfos, es decir que no puede ser autofecundada con su propio polen; de forma que los únicos cruces legítimos (productores de progenie fértil) son aquellos que ocurren entre morfos distintos (Kohn y Barrett, 1992). Estos morfos han recibido distintos nombres con el tiempo pin y trum; longistilas y brevistilas; morfo largo y corto, etc.

Las flores longistilas se caracterizan por poseer estilos largos y estambres cortos en cambio en las flores brevistilas las anteras están por encima de los estigmas (figura 1.2). La forma de la flor de casi todas las plantas es actinomorfa (que tiene al menos dos planos de simetría), aunque algunas plantas tienen flores zigomorfas (sólo tienen un plano de simetría) (Ganders, 1979; Naiki, 2012).

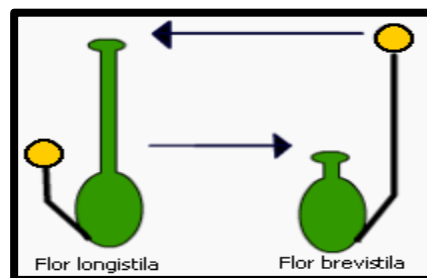


Figura 1.2. Diagrama simplificado de la posición del pistilo (verde) y estambres (negro y amarillo) en el sistema de distilia (Schopfer et al., 1999).

1.2.1 Filogenia de las plantas con flores distílicas

La distilia ha surgido de manera independiente en al menos 27 familias de angiospermas, con múltiples orígenes en algunas familias. Por ejemplo, al menos dos en Linaceae, 12 en Boraginaceae, 20 en Rubiaceae y es probable que otras familias, como Polygonaceae, también incluyen múltiples orígenes del sistema reproductivo (Ganders, 1979; Barret y Shore, 2008; Schuster *et al.*, 2011; Ferrero *et al.*, 2012; Cohen, 2014). La distilia se encuentra en 13 órdenes en tres clados: dos familias en un orden en monocotiledóneas, 10 familias en seis órdenes de rósidas y 14 familias en siete órdenes de astéridas (figura 1.3).

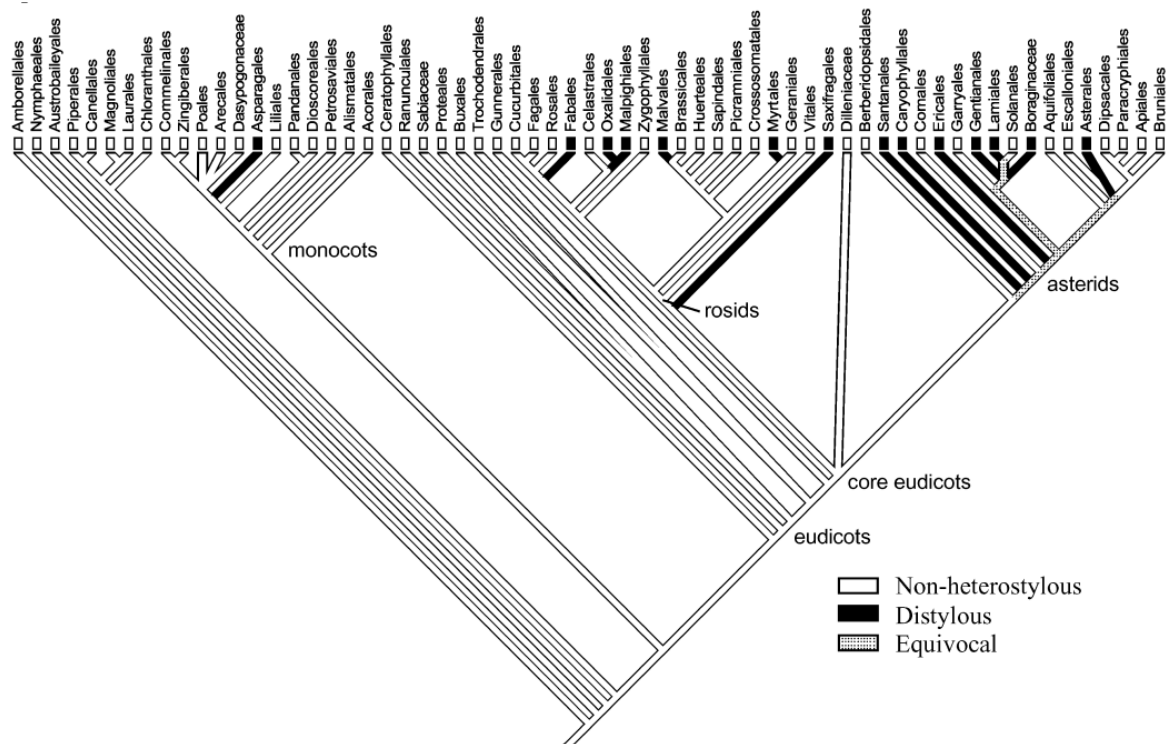


Figura 1.3. Distribución filogenética de los géneros de angiospermas con especies distílicas. Órdenes a los que pertenecen las familias que contienen especies distílicas: (Boraginaceae no pertenece a ningún orden) incluye Asparagales (Amaryllidaceae e Iridaceae), Fabales (Fabaceae), Oxalidales (Connaraceae y Oxalidaceae), Malpighiales (Erythroxylaceae, Hypericaceae, Linaceae y Passifloraceae), Malvales (Malvaceae), Myrtales (Lythraceae), Saxifragales (Saxifragaceae), Santalales (Santalaceae y Schoepfiaceae), Caryophyllales (Plumbaginaceae y Polygonaceae), Ericales (Polemoniaceae y Primulaceae), Gentianales (Gentianaceae, Gelsemiaceae y Rubiaceae), Lamiales (Acanthaceae, Lamiaceae and Oleaceae), Boraginaceae y Asterales (Menyanthaceae) (tomado de Naiki, 2012).

En un estudio de las interacciones del polen y pistilo de especies auto-incompatibles (*Amsinckia* sp. y *Eichhornia paniculata*) Lloyd y Webb (1992) habían propuesto que el ancestro de los grupos distílicos poseía hercogamia de aproximación (o, menos frecuentemente, inversa) y que la hercogamia recíproca se originó antes de la auto e intra incompatibilidad mórfica. Sin embargo, al concluir con su trabajo consideraban que esta hipótesis podría modificarse debido a la falta de estudios de desarrollo en otras familias distílicas.

Naiki (2012) con base en el Sistema de clasificación APGIII (APG III 2009), reportó 199 géneros en 26 familias en 15 órdenes como taxones reconocidos por contener especies distílicas. Solo cuatro familias habían sido agregadas en 30 años. Además de acuerdo con estudios moleculares filogenéticos el autor sugiere que en algunos linajes (*Amsinckia*, Schoen et al. 1997; *Houstonia*, Church 2003; *Turnera*, Truyens et al. 2005; Menyanthaceae, Tippery et al. 2008) la distilia es una de las plesiomorfías; es decir, una característica ancestral.

Cohen (2019), realizó una reconstrucción de las filogenias para 15 familias y parientes que representaban 12 orígenes de la distilia y resolvió los patrones de hercogamia ancestral. Los ancestros de la mayoría de las especies tienen como estado ancestral la hercogamia, siendo la hercogamia de aproximación, el estado ancestral más probable. De este resultado, se deriva que en estas especies distílicas, el morfo brevistilo será la condición derivada. En un pequeño número de especies, como *Cordia boissieri*, la hercogamia inversa, se resuelve como la condición ancestral, por lo que en estas especies el morfo longistilo sería la condición derivada (tabla 1.1). Estas diferencias en los patrones de desarrollo sugieren que los sistemas distílicos pueden tener diferentes orígenes, incluso dentro de una misma familia.

Tabla 1.1. Resultados de los análisis filogenéticos de las especies distílicas, incluyendo el número de especies, las regiones y la longitud alineada de las regiones de DNA, en pares de bases (pb). Tipo ancestral de hercogamia señalado, para cada especie, mediante la reconstrucción de caracteres ancestrales en los marcos de referencia de máxima parsimonia (MP) y máxima verosimilitud (ML) tomado de (Cohen, 2019).

| Especies | Familia | No. de especies | Tipo de hercogamia ancestral | |
|------------------------------|---------------|-----------------|---|---|
| | | | MP | ML |
| <i>Aliciella heterostyla</i> | Polemoniaceae | 336 | Aproximación/Ausencia de hercogamia | Ausencia de hercogamia |
| <i>Averrhoa carambola</i> | Oxalidaceae | 298 | Heterostilia ancestral | |
| <i>Cordia boissieri</i> | Cordiaceae | 1200 | Hercogamia reversa | |
| <i>Fagopyrum esculentum</i> | Polygonaceae | 544 | Ausencia de hercogamia | Heterostilia más probable ancestral; Ausencia de, menos probable |
| <i>Houstonia acerosa</i> | Rubiaceae | 3377 | Heterostilia ancestral | |
| <i>Houstonia nigricans</i> | Rubiaceae | 3377 | Heterostilia ancestral | |
| <i>Houstonia wrightii</i> | Rubiaceae | 3377 | Heterostilia ancestral; Aproximación y Hercogamia reversa resuelta en ancestros y familias cercanos | Heterostilia ancestral; Hercogamia reversa menos probable |
| <i>Linum perenne</i> | Linaceae | 97 | Heterostilia ancestral; Ausencia de hercogamia resuelta en ancestros del clado de las especies herostilas | Heterostilia ancestral; otros tipos de hercogamia también probables |
| <i>Nivenia parviflora</i> | Iridaceae | 1025 | Hercogamia de aproximación | |

| Especies | Familia | No. de especies | Tipo de hercogamia ancestral | |
|-------------------------------|----------------|-----------------|--|--|
| | | | MP | ML |
| <i>Nivenia stenosphon</i> | Iridaceae | 1025 | Hercogamia de aproximación | |
| <i>Oreocaraya paysonii</i> | Boraginaceae | 1200 | Heterostilia ancestral; Ausencia de hercogamia en familias cercanas | |
| <i>Plumbago auriculata</i> | Plumbaginaceae | 308 | Heterostilia ancestral; Hercogamia reversa resuelta en ancestros | Heterostilia ancestral; reversa o Hercogamia de aproximación resuelta en ancestros |
| <i>Pulmonaria mollis</i> | Boraginaceae | 1200 | Heterostilia ancestral; Hercogamia de aproximación resuelta en el ancestro más cercano | |
| <i>Salvia brandegeei</i> | Labiatae | 486 | Hercogamia de aproximación | |
| <i>Turnera diffusa</i> | Passifloraceae | 473 | Heterostilia ancestral | |
| <i>Lithospermum canescens</i> | Boraginaceae | 1200 | Hercogamia de aproximación | |
| <i>Psychotria chiapensis</i> | Rubiaceae | 3377 | No incluida en análisis filogenéticos | |
| <i>Psychotria poeppigiana</i> | Rubiaceae | 3377 | Heterostilia ancestral | |
| <i>Bouvardia ternifolia</i> | Rubiaceae | 3377 | Heterostilia ancestral | |
| <i>Guettarda scabra</i> | Rubiaceae | 3377 | Hercogamia de aproximación | Heterostilia ancestral o Hercogamia de aproximación |
| <i>Amsinckia</i> | Boraginaceae | 1200 | Ausencia de hercogamia | Heterostilia ancestral |

| Especies | Familia | No. de especies | Tipo de hercogamia ancestral | |
|-------------------------------|---------------|-----------------|--|---|
| | | | MP | ML |
| <i>Houstonia caerulea</i> | Rubiaceae | 3377 | Heterostilia ancestral; Hercogamia reversa resuelta en familias cercanas | Heterostilia ancestral; Hercogamia reversa resuelta en familias cercanas y posible en ancestros |
| <i>Quinchamalium chilense</i> | Schoepfiaceae | 159 | Hercogamia de aproximación | |
| <i>Oreocarya crassipes</i> | Boraginaceae | 1200 | Heterostilia ancestral; familias cercanas con faltos de hercogamia | |
| <i>Polygonum jucundum</i> | Polygonaceae | 544 | Ausencia de hercogamia | Heterostilia ancestral más probable; Ausencia de hercogamia menos probable |
| <i>Primula vulgaris</i> | Primulaceae | 734 | Heterostilia ancestral | |

1.3 Desarrollo Floral

El desarrollo de las flores es un proceso que consta de varios pasos que incluyen el desarrollo del meristemo floral, el establecimiento de identidades únicas de órganos, la diferenciación y crecimiento de estructuras florales. Las estructuras como la flor y los órganos que la conforman pueden surgir en diferentes arreglos desde el eje central, la disposición particular en que las estructuras vegetales se encuentran alrededor de un eje o centro se conoce como filotaxia y es variable de acuerdo a la especie (figura 1.4) (Simpson, 1981; Krammer, 2018; Smyth, 2018).

La floración está asociada con un cambio en la filotaxia, mismo que se refleja en variaciones de la tasa de crecimiento y división de las células en el meristemo floral. Además, el meristemo floral produce órganos florales, a diferencia de la formación de hojas y ramas por los brotes vegetativos de los meristemos apicales. Las posiciones de estos órganos alrededor del eje central y sus respectivas identidades se especifican de forma única durante la floración (Carpenter *et al.*, 1995; Zik y Irish, 2003). Una característica de los brotes de meristemo apical floral es que son determinados en el sentido de que producen un conjunto finito de órganos y luego dejan de proliferar. Esta transformación a un destino determinado (no se dividirán más) tiene implicado que todas las células del meristemo floral se comprometen a una vía de diferenciación única y definitiva (Zik y Irish, 2003).

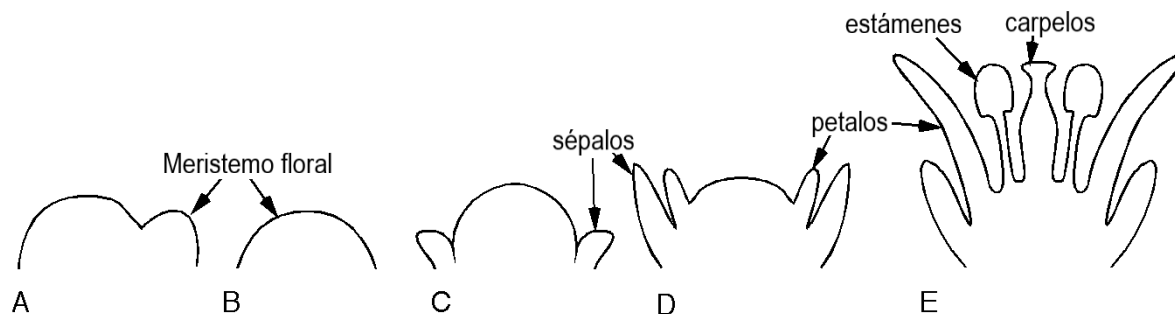


Figura 1.4. Desarrollo floral. (A) Las flores se desarrollan a partir de los meristemos florales, con frecuencia surgen en los laterales de un meristemo de inflorescencia. (B) El meristemo floral es un pequeño grupo de células en forma de domo y está conformado de varias capas celulares. (C) Las células en los laterales del meristemo floral proliferan para formar primero el primordio de los sépalos, (D) seguidos por el primordio de los pétalos. (E) Finalmente, cada una de las estructuras florales se define, consistiendo en sépalos, pétalos, estámenes y carpelos (modificado de Zik y Irish, 2003).

La floración inicia en respuesta a aspectos ambientales y rutas endógenas, éstas últimas referidas como autónomas (controladas genéticamente) que funcionan independientemente de las señales ambientales y están relacionadas con el estado de desarrollo de la planta (Amasino y Michaels, 2010). Los estímulos exógenos activan la división mitótica del meristemo, para otorgar identidad a los órganos presentes en los verticilos florales se requiere de la interacción de al menos tres tipos de productos génicos con funciones distintas de acuerdo con el modelo ABC molecular (Bowman *et al.*, 1991; Coen y Meyerowitz, 1991).

1.3.1 Modelo ABC molecular de la floración

Estudios en mutantes de *Arabidopsis* y *Antirrhinum* han conducido al modelo ABC de especificación de la identidad de los órganos florales (Coen y Meyerowitz, 1991; Weigel y Meyerowitz, 1994). Este modelo postula tres clases de función genética: A, B y C, que actúan de forma combinada para especificar de manera única cada tipo de órgano. La expresión combinada de los genes ABC produce el desarrollo de los órganos florales necesarios para la reproducción (figura 1.5). Los genes de clase A determinan la formación de sépalos en el primer verticilo, mientras que al coexpresarse con los genes de clase B promueven el desarrollo de pétalos. La combinación de la expresión de los genes de clase B y C desarrollan los estambres, y los genes de clase C por sí solos controlan la diferenciación de los carpelos en la región central de la flor (Bowman *et al.*, 1991; Weigel y Meyerowitz, 1994).

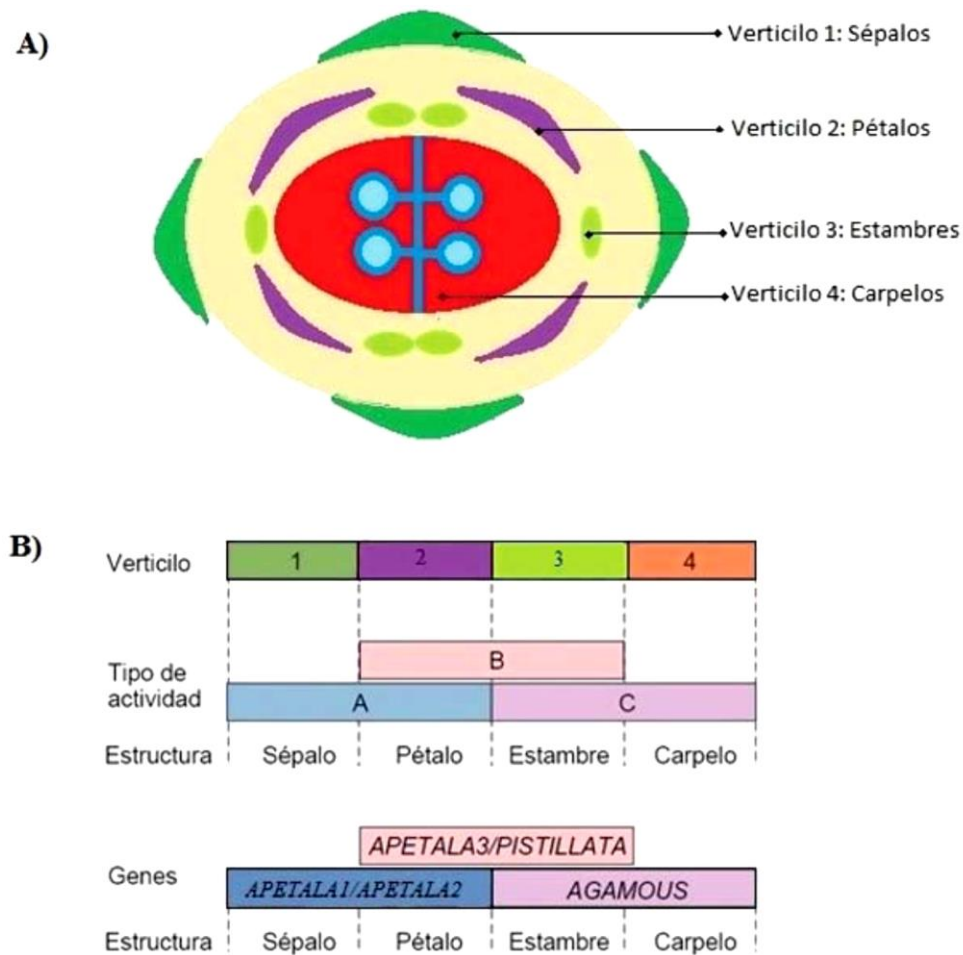


Figura 1.5. Modelo ABC de desarrollo floral en *A. thaliana*. A) verticilos florales dispuestos alrededor del meristemo floral y los órganos que dichos verticilos originan. B) distintas clases de genes (A, B y C), verticilos en los que actúan y que estructura generan (tomado de Mendoza y Encina, 2018).

El modelo también propone que los genes de la clase A y los de la clase C se reprimen mutuamente. De esta manera los genes de tipo A reprimen la actividad de los genes de tipo C en la parte externa de la flor (verticilos correspondientes a pétalos y sépalos) y los genes de clase C reprimen la actividad de los de clase A en la parte interna de la flor (verticilos estaminales y carpelares) (Bowman et al., 1991; Weigel y Meyerowitz, 1994; Kramer, 2018).

En *A. thaliana* los genes de clase A son APETALA1 y APETALA2 (AP1 y AP2 respectivamente), para los de la clase B tenemos a APETALA3 (AP3) y PISTILLATA (PI), y el único gen de la clase C se llama AGAMOUS (AG) (Bowman et al., 1991; Weigel y Meyerowitz, 1994; Smith, 2018). El gen AP1, además de participar en la formación de los sépalos, también participa en la formación del primordio floral, ya que regula enzimas que

participan en el metabolismo de las giberelinas y ajusta factores de transcripción como LEAFY (LFY), que es codificado por un gen de identidad de meristemo (Abe *et al.*, 2005; Wigge *et al.*, 2005; Hames *et al.*, 2008; Abelenda *et al.*, 2014; Sayou *et al.*, 2016).

Se han descubierto otros genes que también participan en la formación de los órganos florales además de los genes de identidad floral ABC (Fig. 6). Los genes de clase E son necesarios para que se desarrollen todos los órganos florales y pertenecen también a la familia de los MADS-box (Theißen y Saedler, 2001). En la clase de genes E se han descrito: SEPALLATA1, SEPALLATA2 y SEPALLATA 3 (SEP1, SEP2 y SEP3). Estos genes se requieren junto con los genes B y C para la formación de los pétalos (A+B+E), estambres (B+C+E) y carpelos (C+E) (Theißen, 2001).

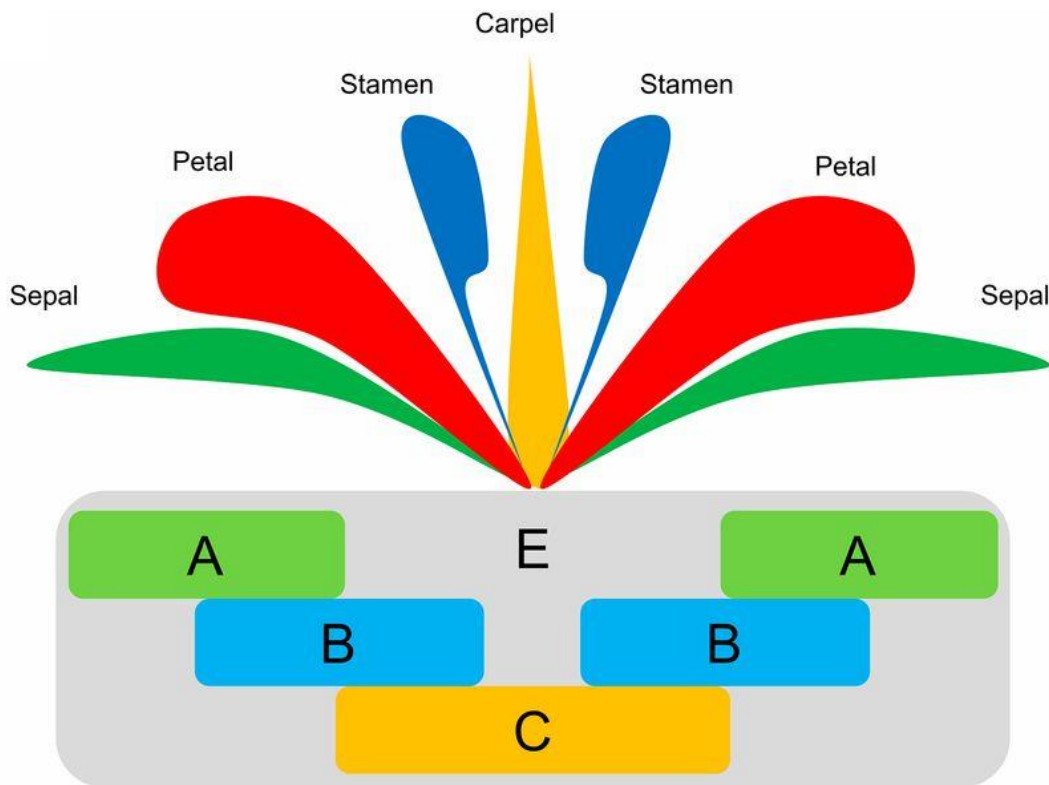


Figura 1.6. El modelo clásico ABCE especifica cuatro órganos florales morfológicamente discretos: los sépalos se producen donde sólo actúa la función A, los pétalos se producen con la expresión combinada de las funciones A y B, los estambres se producen con la expresión combinada de las funciones B y C, y los carpelos se producen donde sólo actúa la función C (tomado de Chanderbali *et al.*, 2016).

Cada especie tiene rasgos característicos en su desarrollo floral, los órganos florales poseen diversos números y acomodos en la flor, incluso, pueden presentar variaciones en las alturas a las que se encuentran debido a mecanismos de desarrollo relacionados con el sistema reproductivo. A estos patrones de desarrollo específicos de las flores se les conoce como patrones de desarrollo floral.

1.3.2 Patrones de desarrollo floral en especies distílicas

Investigaciones a nivel macroscópico (morfológico) aportan una visión en las numerosas maneras en los que los mismos órganos pueden alcanzar diferentes longitudes y alturas en las flores de cada morfo, que incluyen distintas (A) tasas de crecimiento diferentes a lo largo del desarrollo; (B) tasas de crecimiento diferencial durante momentos específicos de desarrollo; (C) crecimiento diferencial de los órganos de los meristemas durante el desarrollo temprano; (D) crecimiento similar en etapas tempranas u cese de éste para uno de los órganos sexuales; (E) crecimiento y cese de uno de los órganos para que el otro crezca; o combinaciones de estos posibles patrones (figura 7) (Bull *et al.*, 2016; Cohen *et al.*, 2012; Richards y Barrett, 1992).

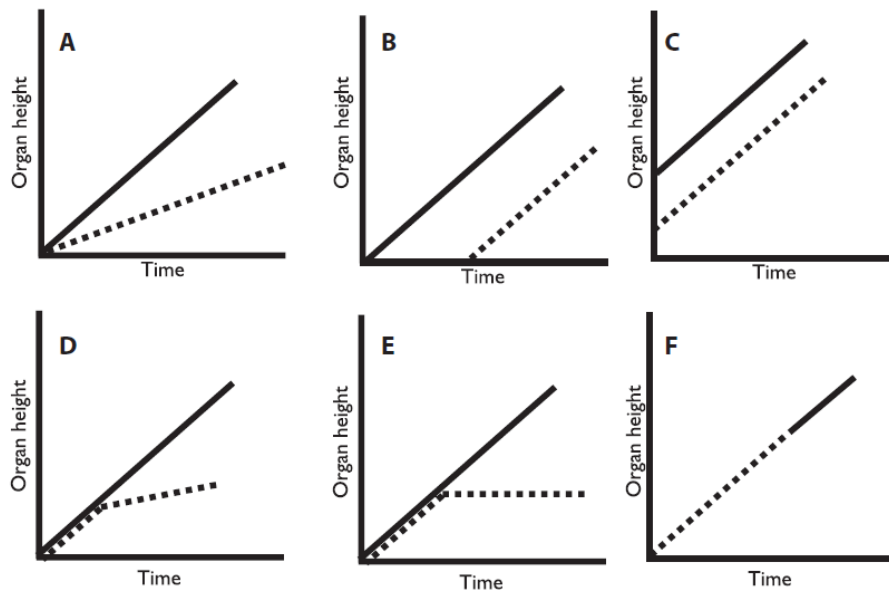


Figura 1.7. Posibles patrones de desarrollo para la altura de la antera y el estigma en los morfos de estilo largo y corto de las especies distílicas (Richards y Barrett, 1992).

En 2019 Cohen comparó patrones florales de 15 especies de plantas distílicas que representan 12 orígenes distílicos distintos. En las especies analizadas, los patrones de desarrollo en la altura de las anteras y estigma diferían significativamente entre morfos, pero no para todas las estructuras y órganos demostrando la existencia de patrones de desarrollo floral específicos por especie como Richards y Barret habían propuesto. (tabla 1.2, figura 1.7).

Tabla 1.2. Colección de especies distílicas y sus patrones de altura en estigma y anteras (Modificado de Cohen, 2019).

| Especies | Patrón de altura en las anteras | Patrón de altura del estigma |
|-------------------------------|--|--|
| <i>Aliciella heterostyla</i> | Diferentes tasas de crecimiento. | Tasa de crecimiento distinta tardía |
| <i>Averrhoa carambola</i> | El crecimiento cesa más tarde | Tasa de crecimiento distinta tardía |
| <i>Cordia boissieri</i> | El crecimiento cesa más tarde | Diferentes tasas de crecimiento. |
| <i>Houstonia acerosa</i> | Diferentes tasas de crecimiento. | Tasa de crecimiento distinta tardía |
| <i>Houstonia nigricans</i> | Diferentes tasas de crecimiento y duración del período de crecimiento. | Tasa de crecimiento distinta tardía |
| <i>Houstonia wrightii</i> | Diferentes tasas de crecimiento y duración del período de crecimiento. | Tasa de crecimiento distinta tardía |
| <i>Linum perenne</i> | El crecimiento cesa más tarde | Cese de crecimiento tardío |
| <i>Turnera diffusa</i> | Tasa de crecimiento distinta tardía | Tasa de crecimiento distinta tardía |
| <i>Lithospermum canescens</i> | Diferentes tasas de crecimiento. | El crecimiento cesa más tarde |
| <i>Primula vulgaris</i> | Diferentes tasas de crecimiento. | Crecimiento distinto tardío, ya sea incremento o disminución |

Los patrones de crecimiento en los órganos sexuales pueden variar, no solo durante todo el desarrollo de la flor sino en momentos específicos y velocidades distintas. Los estambres (y las corolas, si los estambres son adnatos) y los carpelos pueden seguir cada uno patrones de crecimiento distintos y específicos para cada morfo, que elevan las anteras y los estigmas, respectivamente, a sus posiciones particulares. Este patrón de desarrollo incluso tiene momentos específicos de crecimiento (de manera inicial o tardía) que definen la formación del morfo brevistilo o longistilo según sea el caso. Por lo tanto, el potencial de crecimiento de las diversas estructuras masculinas y femeninas (androceo y gineceo) permite un gran número de combinaciones potenciales de patrones de crecimiento específicos de cada morfo y especie para estos órganos sexuales (figura 1.8) (Cohen, 2019; Richards y Koptur, 1993).

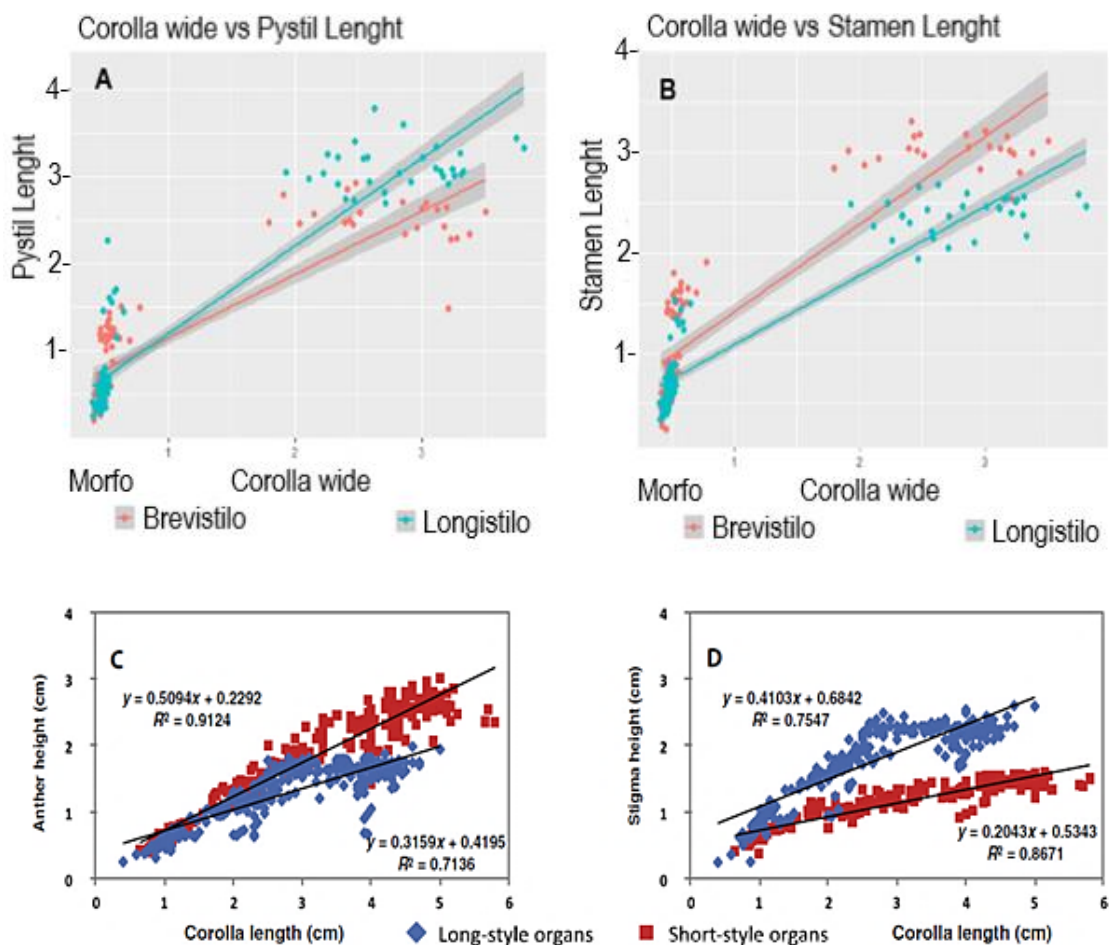


Figura 1.8. Patrones de desarrollo para la altura de la antera y del estigma para (A)(B) *Cordia sebestena* y (C)(D) *Cordia boissieri*. Cuadros azules: órganos de morfo largo; cuadros rojos: órganos de morfo corto (*C. sebestena* Ferrer (comunicación personal); *C. boissieri* Modificado de Cohen, 2019).

Incluso si dos especies tienen patrones de desarrollo similares para los órganos que elevan las anteras y los estigmas (Figura 1.8), diferentes partes de la corola, el androceo y del gineceo pueden participar en la producción de las alturas específicas de cada morfo en cada especie (por ejemplo, en la antesis, las especies con estambres adnatos a la corola pueden tener la altura de las anteras modificadas, ya sea cambiando la longitud de los filamentos o el punto de unión de los filamentos a la corola o una combinación de ambos) (Faivre, 2000; Sampson y Krebs, 2013).

La evidencia demuestra como los patrones de desarrollo floral, mismos que dictan los momentos clave en la conformación de los morfos, varían entre las especies (Cohen, 2010). Estas diferencias pueden apreciarse a distintos niveles cómo: genómico, transcriptómico, micromorfológico y morfológico. Al examinarlas en todos sus niveles aportan una mejor comprensión sobre como los mismos órganos pueden alcanzar diferentes longitudes y alturas en las flores de cada morfo (Armbruster *et al.*, 2006; Faivre, 2000; Richards y Barrett, 1992).

1.4 Transcriptómica

La secuenciación del transcriptoma se utiliza para estudiar el transcriptoma de un organismo, es decir la presencia de todas las moléculas de RNA mensajero (RNAm) bajo cierta condición o tejido específico. Es utilizada para entender cómo la información contenida en el genoma se traduce en funciones celulares que responden, la traducción responde a condiciones del entorno (Fang *et al.*, 2012; Uhlén *et al.*, 2016; Lowe *et al.*, 2017).

Los principales objetivos de la transcriptómica son: 1) catalogar los transcritos que incluyen mRNA, RNA no codificante y RNA pequeños; 2) determinar la estructura transcripcional de los genes, patrones de empalme y modificaciones postranscripcionales, y 3) cuantificar los niveles de expresión de los genes (Wang *et al.*, 2009).

Medir la expresión de los genes bajo condiciones específicas o en tejidos específicos proporciona información sobre cómo se regulan los genes y revela detalles de la biología de un organismo, así como, permite inferir las funciones de genes no anotados previamente. El análisis transcriptómico ha permitido estudiar cómo cambia la expresión de los genes en diferentes organismos. Analizar la expresión génica en su totalidad permite detectar amplias tendencias coordinadas que no se pueden discernir con ensayos más específicos (Lowe, 2017).

De los distintos métodos para evaluar la expresión de los genes, la secuenciación de alto rendimiento del RNA, o RNA-Seq resulta una estrategia muy atractiva ya que no se requiere de la secuenciación del genoma para conocer la expresión de genes bajo determinadas condiciones (Góngora-Castillo y Buell, 2013).

1.4.1 Técnica RNA-seq para obtener perfiles de expresión génica

Es una técnica que puede examinar la cantidad y las secuencias de RNA en una muestra utilizando la secuenciación de nueva generación (NGS, por sus siglas en inglés). Esta tecnología captura el RNA total o RNAm, el cual se fragmenta y convierte en una librería de DNA complementario (cDNA) que posteriormente es amplificada y secuenciada (Chiner *et al.*, 2019). El RNA-seq da una cobertura completa de transcritos, genera información no solo de la secuencia, sino también de la estructura de exones y posibles eventos de splicing alternativo (Lister *et al.*, 2009).

Uno de los pasos previos es la obtención de un RNA de buena calidad que represente todos los transcritos que se producen en la condición y tejido de estudio (Ward *et al.*, 2012). El primer paso de la técnica es la fragmentación del RNA o del cDNA se realiza o bien por nebulización, por digestión con enzimas de restricción o a través del uso de cationes divalentes bajo condiciones de presiones elevadas (Wang *et al.*, 2009). Una vez obtenido el DNAc se ligan adaptadores de tal forma que cada fragmento generado contendrá un adaptador ligado en sus extremos 3' y 5'. Las secuencias de estos adaptadores se conocen y serán necesarias para que cada fragmento pueda ser secuenciado, y en algunos casos pueden emplearse para diferenciar otros grupos de fragmentos obtenidos a partir de muestras de DNAc diferentes (Marguerat y Bähler, 2010).



Figura 1.9. Esquema de los pasos a seguir en la técnica de RNA-seq (tomado de Ovalle Rivera, 2018).

Además de la cantidad de secuencias que se pueden generar durante la secuenciación, se deben tener ciertas consideraciones adicionales como: 1. El método de secuenciación single-end (SE) o paired-end (PE), 2. El largo de los reads (lecturas) y 3. Cobertura (Góngora-Castillo *et al.*, 2018).

En el modo single-end (SE), el secuenciador lee un fragmento sólo de un extremo a otro a diferencia del modo paired-end (PE) donde la lectura comienza en un extremo, termina en esta dirección y luego comienza otra ronda de lectura desde el extremo opuesto del fragmento. El uso de lecturas paired-end en lugar de single-end mejora significativamente el ensamblaje del transcriptoma, así como la longitud de las secuencias (Góngora-Castillo *et al.*, 2018; Goodwin *et al.*, 2016).

Por último, la determinación de la cobertura en la secuenciación del RNA es complicada, debido a que los transcritos se expresan a diferentes niveles, es decir, se secuenciarán más lecturas de genes altamente expresados y se obtendrán menos secuencias de genes poco expresados. Las consideraciones a tener en cuenta para la estimación de la cobertura son continuamente discutidas en artículos y pueden variar de acuerdo al objeto de estudio que emplea el RNA-seq (Góngora-Castillo *et al.*, 2018; Sims *et al.*, 2014).

1.4.2 Análisis bioinformático para RNA-seq

El análisis bioinformático para RNA-seq, empieza con la verificación de la calidad de las secuencias. Las secuencias obtenidas (reads) pueden tener errores de tipo instrumental o fallas técnicas, por lo cual es necesario evaluar la calidad de las mismas. El software más utilizado para estas tareas es FastQC (Andrews, 2010) que permite obtener parámetros de calidad por nucleótido secuenciado. En caso de que la calidad de las secuencias fuera muy baja, existen diversas herramientas que ayudan a eliminar ciertas secuencias o a recortar el extremo de las secuencias que tiene una baja calidad con software como Prinseq ó Trimmomatic (Van Verk *et al.*, 2013).

Mapeo y visualización

El mapeo genómico consiste en alinear las secuencias (reads) obtenidas mediante NGS con el genoma (o transcriptoma) de referencia. Mapeando las secuencias obtenemos información de su localización en el genoma, la cual puede ser utilizada para identificar y cuantificar los genes expresados, conocer las variantes alélicas y además descubrir nuevos genes. En especies poco estudiadas puede ser que el genoma de referencia aún no esté disponible, en estos casos es posible alinear a un transcriptoma ensamblado *de novo*. La primera consideración para elegir un software de alineamiento es el splicing alternativo. En organismos que no tienen intrones es mejor utilizar alineadores contiguos como Bowtie2 ó BWA que fueron diseñados para alinear DNA. Sin embargo, cuando se tienen genomas que poseen intrones es mejor utilizar alineadores como Tophat2, Hisat2, STAR, GSNAP, SOAP2 o Kallisto (Haas *et al.*, 2013; Langmead y Salzberg, 2012; Li y Durbin, 2009; Kim *et al.*, 2013; Dobin *et al.*, 2012; Wu y Nacu, 2010; Li *et al.*, 2009; Bray *et al.*, 2016).

Anotación funcional

Se refiere al proceso de identificación de los genes expresados, y a conocer la información disponible sobre su función en los genomas de referencia y bases de datos públicas. Los transcriptomas ensamblados de novo pueden ser identificados utilizando algoritmos de alineamiento de secuencias como BLAST (Altschul *et al.*, 1990). Las secuencias del transcriptoma se comparan con secuencias de nucleótidos y/o proteínas contenidas en bases de datos públicas para identificar funciones relacionadas a secuencias específicas, así como regiones conservadas con funciones específicas como motivos y dominios. Entre

las bases de datos más utilizadas para este propósito se encuentran RefSeq (O'Leary *et al.*, 2016), Pfam (Mistry *et al.*, 2021), Gene Ontology (Mi *et al.*, 2019), etc.

Análisis de ontología génica (GO)

Una ontología es una representación formal de un conjunto de conocimientos dentro de un dominio determinado. Las ontologías suelen consistir en un conjunto de clases (o términos o conceptos) con relaciones que operan entre ellas. La ontología de genes (GO) describe nuestro conocimiento del dominio biológico con respecto a tres aspectos: proceso biológico, función molecular y componente celular. Los vocabularios están estructurados en una clasificación que admite las relaciones “es un” y “parte de” (Mi *et al.*, 2019; Blake y Harris *et al.*, 2003).

La función molecular (MF) describe actividades, como las catalíticas o de unión, a nivel molecular. Los términos GO de función molecular representan actividades (moléculas o complejos) que realizan las acciones, y no especifican dónde, cuándo o en qué contexto tiene lugar la acción. El proceso biológico (PB) describe los objetivos biológicos que se logran mediante uno o más conjuntos ordenados de funciones moleculares. Los procesos de alto nivel, como la "muerte celular", pueden tener subtipos, como la "apoptosis", y subprocesos, como la "condensación cromosómica apoptótica". El componente celular (CC) describe las localizaciones, a nivel de estructuras subcelulares y complejos macromoleculares (Harris *et al.*, 2004; Mi *et al.*, 2013).

Además, también se puede analizar rutas metabólicas (KEGG) y otras agrupaciones más complejas (PANTHER). Herramientas como AgriGO nos permiten visualizar mediante arboles jerárquicos los términos GO significativamente expresados (Kanehisa, 2004; Mi *et al.*, 2013).

1.4.3 Genes relacionados con la distilia

Estudios sugieren que existe un conjunto de genes controlando los componentes morfológicos, micromorfológicos y fisiológicos de la distilia. En los últimos años diversas investigaciones han propuesto los genes responsables del desarrollo distintivo de los morfos de diferentes especies de plantas distílicas.

El control genético de la distilia se propuso por primera vez por Bateson y Gregory (1905) en su trabajo en la especie *Primula sinensis*, aunque Darwin había realizado ya los cruces necesarios para llegar a esta afirmación en 1877. Se tenía la hipótesis de que el síndrome de los caracteres florales, por el que se diferencian los morfos de estilo largo y corto, se heredaba como si estuviera controlado por un único locus genético con dos alelos, el de estilo largo es homocigótico (ss) mientras que el de estilo corto es heterocigótico (Ss); las plantas de estilo largo son ss y las plantas de estilo corto pueden ser Ss o SS. En todos los casos, el síndrome de caracteres que comprende el polimorfismo estaba controlado por un único locus de un gen con dos alelos (Barrett y Shore, 1985).

En *Turnera* el morfo de estilo corto (morfo S) es controlado por el alelo dominante S y el morfo de estilo largo (morfo L) homocigoto para el alelo recesivo s. Se identificó una poligalacturonasa y una α -dioxigenasa como proteínas diferencialmente abundantes entre los estilos de las flores de morfo S y L; en particular, ambas proteínas sólo se detectan en los estilos del morfo S. El gen que codifica la poligalacturonasa está vinculado al locus S, pero separado de él, mientras que el gen de la α -dioxigenasa no está vinculado, lo que sugiere que la expresión de ambos está controlada secundariamente por los factores distílicos causales (Athanasίου *et al.*, 2003; Ganders, 1979; Khosravi *et al.*, 2004).

Las nuevas tecnologías de secuenciación masiva han supuesto un gran avance para identificar los transcritos restringidos del morfo S, con la nueva hipótesis de que el haplotipo S es hemocigoto. Cocker y colaboradores (2018) ensamblaron el genoma de *Primula vulgaris* y anotaron más de 24, 000 genes encontrando más genes regulados río arriba en el morfo brevistilo en comparación con el morfo longistilo. Análisis posteriores revelaron el control por parte de un grupo de genes, o supergen, (locus S, longitud del estilo). El locus S de *P. vulgaris* está ausente en el morfo largo (pin) y presente en el morfo corto (thrum) además el locus S tiene un elevado contenido de repeticiones (64%) en comparación con el genoma completo (37%). Esta conclusión está asociada con la identificación de cuatro transcritos específicos del haplotipo corto en *Fagopyrum*, denominados SSG1 a SSG4 (Yasui *et al.*, 2012). Este trabajo se amplió con una secuencia borrador combinada con el genotipado por secuenciación de los individuos domesticados de los morfos cortos (S) y largos (L); la búsqueda de más contigs específicos de morfo S identificaron un total de al menos 5,4 Mb de secuencias ausentes en todos los individuos del morfo largo, pero con

presencia en la mayoría o en todos los individuos del morfo corto (Kappel *et al.*, 2017; Yasui *et al.*, 2016). Huu y colaboradores (2016) identificaron en *Primula veris*, el gen CYP734A50 que codifica una enzima putativa degradadora de brassinosteroides (una hormona vegetal que regula la elongación celular). Se encuentra únicamente en el morfo thrum (brevistilo) por lo que su pérdida deriva en el morfo largo (longistilo). Los morfos cortos son heterocigotos para ambos haplotipos mientras que el morfo largo es hemocigoto. Los niveles del brasinoesteroide castasterona son altos en los morfos de estilo largo y son indetectables en los morfos de estilo corto. Este hallazgo sugiere que la ausencia de la función CYP^T (Citocromo P450) permite la expresión del fenotipo de estilo largo y que la pérdida de la función GLO^T (un gen específico de *Globerosa* en el morfo brevistilo) del S-locus se asocia con anteras posicionadas en la base de la corola (McClure, 2016). De la misma manera se observó que las variaciones homostilo largo y corto están asociadas con nulas mutaciones en los genes CYP^T (Citocromo P450) y GLO^T (un gen específico de *Globerosa* en el morfo brevistilo) del S-locus respectivamente, y no a la recombinación (figura 1.10) (nótese que el superíndice T en los genes se refiere a thrum, morfología alternativa para la morfología corta). Este hallazgo sugiere que la ausencia de la función CYP^T permite la expresión fenotipo de estilo largo y que la pérdida de la función GLO^T del S-locus se asocia con anteras posicionadas en la base de la corola (McClure, 2016). Este fenómeno es similar al que tienen los cromosomas sexuales X y Y en mamíferos, aunque en una escala menor. Los cromosomas X tienen varias regiones heterólogas (no presentes) en el cromosoma Y, en las especies que presentan distilia, la región heteróloga está presente en el morfo corto, y ausente en el morfo largo.

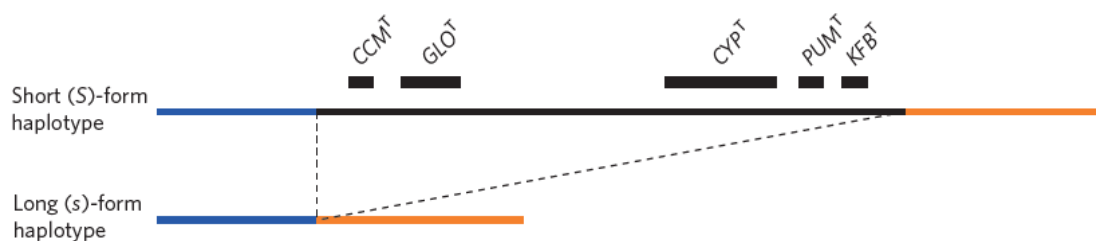


Figura 1.10. Haplotipos asociados a la morfología dominante brevistilo (S/s) y longistilo (s/s). El S-locus de 278 kb contiene cinco genes que no están presentes en las plantas de morfo largo y están ausentes en el haplotipo s: CCMT, GLOT, CYP^T (CYP374A50), PUM^T y KFB^T. Las regiones de la izquierda (azul) y derecha (naranja) contienen genes expresados en ambos morfos. Tomado de McClure, 2016.

En el año 2016 James Cohen, analizó 30 transcriptomas relacionados con tres etapas de desarrollo de los órganos florales (gineceo vs. Corola y androceo) en *Lithospermum multiflorum* Torr. ex A. Gray (especie distilica). Identificó 12 genes diferencialmente expresados (DE) entre morfos, durante el desarrollo (figura 1.11). Pocos de estos genes se encontraban diferencialmente expresados en la etapa temprana de desarrollo en comparación con la tardía. Los genes con mayor expresión se encontraban en el gineceo del morfo brevistilo y en la corola y androceo del morfo longistilo. La expresión diferencial de los genes en las etapas tardías es congruente con un crecimiento interrumpido del gineceo en el morfo corto y del andróceo en el morfo largo.

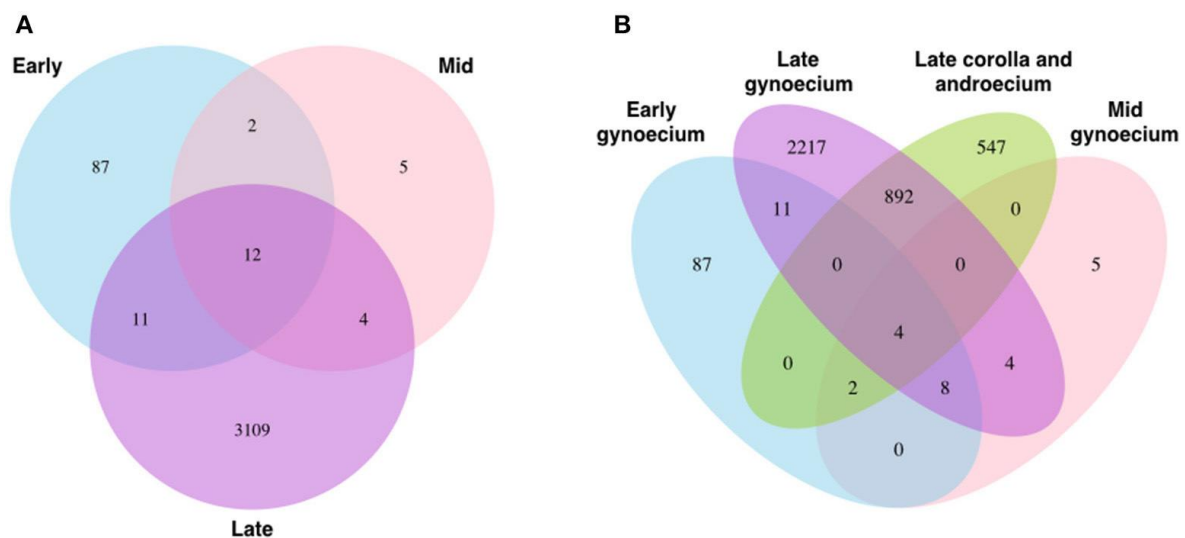


Figura 1.11. Diagrama de Venn resultante de la comparación de los genes diferencialmente expresados por estadio y tejido en *Lithospermum multiflorum*. Tomado de Cohen (2016).

Además, realizó un análisis ontológico el cual arrojó que durante la etapa temprana los genes diferencialmente expresados están relacionados en el crecimiento y desarrollo floral y posteriormente en la etapa tardía se relacionan con funciones fisiológicas además para ambos morfos florales se encontraron genes relacionados con respuesta a estrés (tabla 1.3).

Tabla 1.3. Número de transcritos asociados a tipos de términos de la Ontología Genética (GO) durante las etapas de desarrollo de los órganos florales en *Lythospermum multiflorum*. Tomado de Cohen (2016).

| GO term type | Early corolla and androecium | Early gynoecium | Late corolla and androecium | Late gynoecium | Total |
|--|------------------------------|-----------------|-----------------------------|----------------|-------|
| Biological processes overrepresented | 165 | 38 | 30 | 41 | 210 |
| Biological processes under-represented | 0 | 23 | 199 | 281 | 223 |
| Cellular components overrepresented | 21 | 5 | 8 | 11 | 36 |
| Cellular components under-represented | 0 | 1 | 45 | 49 | 57 |
| Molecular function overrepresented | 45 | 10 | 8 | 15 | 61 |
| Molecular function under-represented | 0 | 12 | 107 | 150 | 159 |

La literatura sugiere en numerosos casos que es el morfo brevistilo en el que se encuentra la mayor cantidad de genes expresados diferencialmente, por lo que los genes expresados en el estilo podrían estar estrechamente relacionados con la regulación o subexpresión de los genes relacionados con la identidad masculina de la planta (androceo) así como la pérdida de la función de un set de genes que suele estar relacionado con factores de transcripción, señalización y hormonas de crecimiento vegetal.

JUSTIFICACIÓN

En la mayor parte de las familias que presentan el polimorfismo floral de distilia la hercogamia de aproximación es el estado ancestral, sin embargo, en la familia *Cordiace* la hercogamia reversa es ancestral. Debido a que en varias familias la heterostilia está determinada por hemicigosis, con la presencia del locus S en el morfo corto y (cinco en *Primula* CCMT, GLOT, CYPT (CYP374A50), PUMT y KFBT) cuya ausencia resulta en el desarrollo del morfo largo. Los estudios transcriptómicos en *Cordia sebestena* permitirán el entendimiento del fenómeno distílico en una especie que pertenece a una familia con un patrón de desarrollo floral único. Por lo que en este estudio se pretende realizar el análisis transcriptómico para poder identificar los genes diferencialmente expresados durante el desarrollo floral de ambos morfos (brevistilo y longistilo) aportando información para un mejor entendimiento del fenómeno distílico.

HIPÓTESIS

“La diferenciación de los morfos florales longistilo y brevistilo de *Cordia sebestena* está relacionada a la expresión de genes involucrados en el desarrollo floral y producción de fitohormonas”.

OBJETIVOS

OBJETIVO GENERAL

Identificar las diferencias de genes expresados que conduzcan a la formación de los morfos brevistilo y longistilo de *Cordia sebestena* a través de un análisis del transcriptoma floral.

Objetivos particulares

1. Obtener y analizar el transcriptoma para los morfos longistilo y brevistilo de *Cordia sebestena* utilizando estrategias de RNA-seq.
2. Identificar las diferencias en los transcriptomas de los morfos longistilo y brevistilo de *Cordia sebestena*, a través de anotación funcional y análisis de enriquecimiento o sobrerrepresentación de genes ontológicos.

ESTRATEGIA EXPERIMENTAL

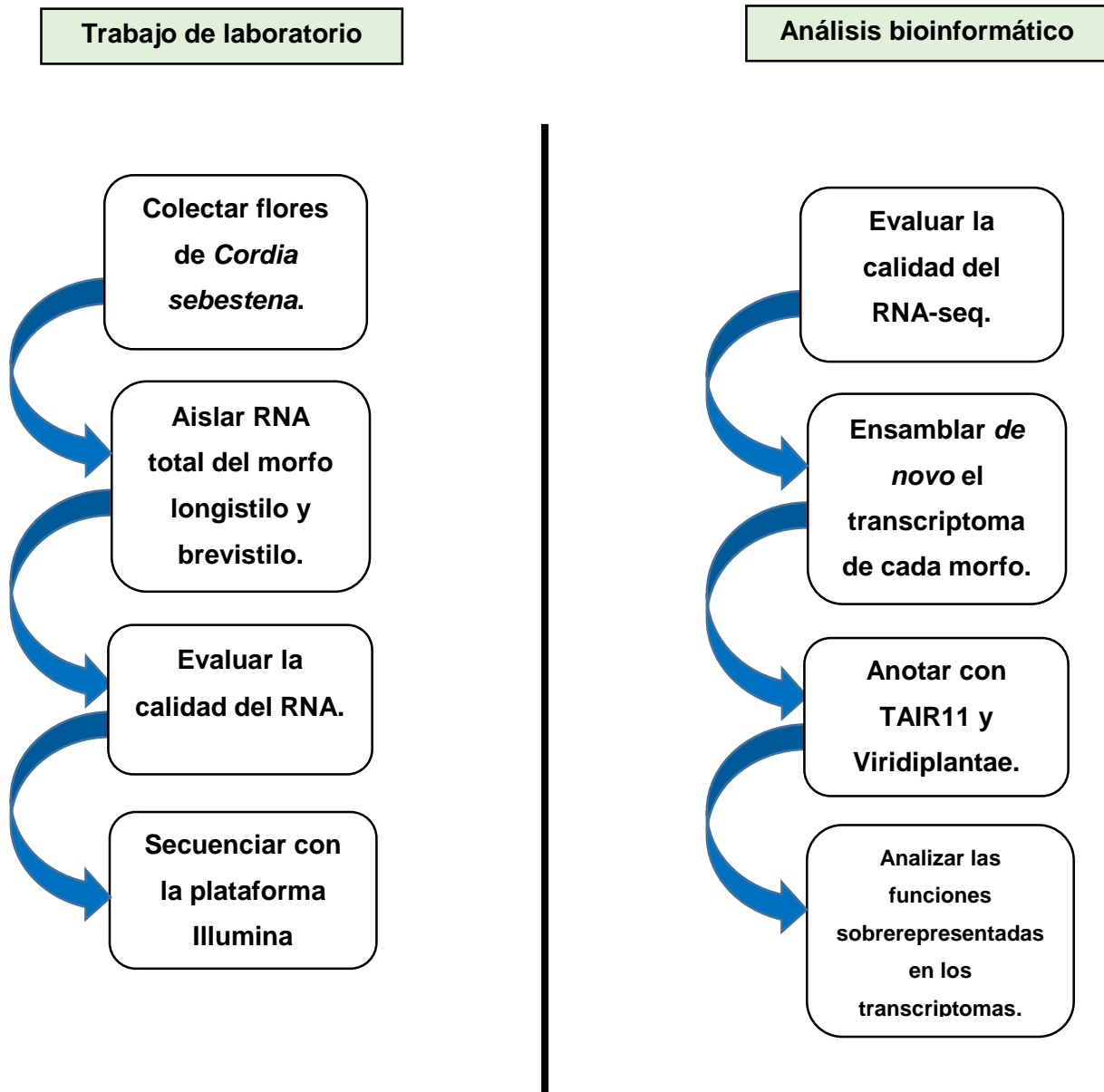


Figura 1.12. Diagrama de flujo de trabajo para la obtención del transcriptoma de *Cordia sebestena*.

CAPÍTULO II

MATERIALES Y MÉTODOS

2.1 Caracterización del transcriptoma de *Cordia sebestena*.

2.1.1 Área de estudio

El estudio se realizó en el Parque Gran Santa Fe 2da etapa, ubicado en la localidad de Cauce al este de la ciudad de Mérida, en el fraccionamiento Gran Santa Fe, Mérida, Yucatán (fig. 6). Ubicado en el Km. 1.6 de la carretera Xmatkuil- Dzununcan 21° 00' 29" latitud norte, y 89° 41' 15" longitud oeste; delimita con la Av. Gran Santa Fe y con casas particulares en los demás puntos colindantes.

El tipo de suelo en la región es cambisol (rojo, o kankab), poco pedregoso, que retiene con mayor facilidad el agua y leptosol (tsekel), poco profundo, con afloramientos de rocas y piedras que dificultan la retención del agua (Estrada, 2000).

El clima es tipo Awo'(w) (i') g cálido subhúmedo, con una temperatura media anual de 26°C y precipitación pluvial de 984.4 mm al año, distribuida de junio a noviembre (Orellana *et al.*, 2003). Predomina la vegetación secundaria proveniente de plantaciones de henequén abandonadas hace varias décadas. De acuerdo con Silberman y colaboradores (2000), la vegetación potencial es de selva baja caducifolia.

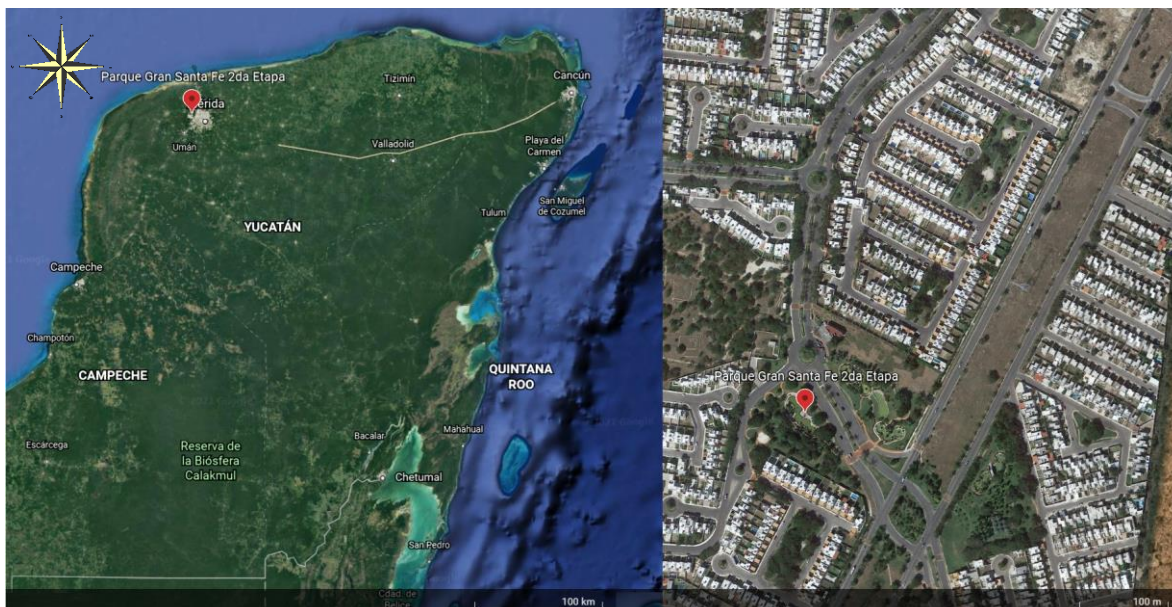


Figura 2.1. Ubicación del área de estudio

2.1.2 Colecta de material biológico

Se seleccionaron individuos con flores brevistilas y longistilas, se identificaron las ramas con presencia de flores para los estadios yema floral, flor pre-antésis y antesis floral para cada morfo floral (longistilo y brevistilo); se tomaron muestras de tres réplicas biológicas por estadio, con un total de 18 muestras por morfo. Las flores colectadas fueron disectadas en campo, colocadas en un tubo eppendorf de 1.5 ml identificados con: 1. el morfo floral, 2. El estadio y 3. Tipo de tejido (androceo o gineceo). Los tubos eppendorf fueron congelados de manera inmediata en nitrógeno líquido para su preservación y posteriormente trasladados al Centro de Investigación Científica de Yucatán, A. C. (CICY) para ser almacenados a -80° C en el ultracongelador del instituto.

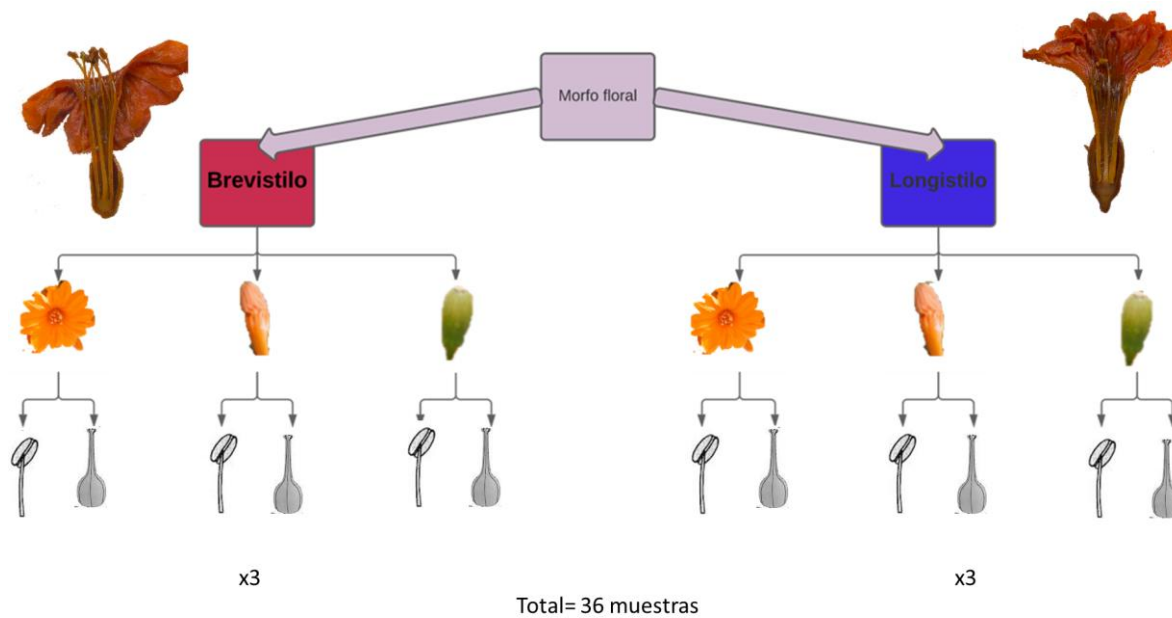


Figura 2.2. Diseño experimental asociado con el morfo floral y estadio de desarrollo de *Cordia sebestana* con número total de muestras.

2.1.3 Extracción de RNA

Para la extracción de RNA se pesaron 100 mg de tejido masculino (antera y filamento) y de tejido femenino (estigma, estilo y ovario). Las muestras congeladas se trituraron hasta obtener un polvo fino en nitrógeno líquido. El RNA total de 100 mg de tejidos congelados de órganos sexuales se aisló mediante el protocolo de RNA RNeasy® Plant Mini Kit de QIAGEN (Anexo I). Se realizó un tratamiento con DNase usando el kit DNase I amplification grade de Invitrogen con el fin de eliminar DNA contaminante. El RNA se visualizó en geles de agarosa al 1% por medio de electroforesis. El RNA total se cuantificó en términos de pureza por medio del dispositivo Qubit™ 2.0 fluorometer (fluorimetría) y Nanodrop™ One (espectrofotometría) equipos que proporcionan la concentración de RNA total de cada muestra, así como sus valores 260/280 y 260/230 de absorbancia, respectivamente.

2.1.4 Tratamiento con DNAsa

Para eliminar la presencia de DNA de las muestras, se agregó en un tubo libre de RNAsa (no incluido): 1 μ l (1 U) de DNase I Amplification grade, 1 μ l del buffer 10X, 7 μ l con agua tratada con DEPC y se mezcló por inversión suave. Por último, la muestra se incubó a 65°C durante 10 min.

2.1.5 Electroforesis en gel de agarosa al 1%

Una vez obtenido el RNA y aplicado el tratamiento con DNasa se corrió una electroforesis en gel de agarosa al 1 %. Brevemente, se agregó 800 mg de agarosa en un volumen de 50 ml de buffer TAE 1X y se clarificó, posteriormente se agregó 0.5 μ l de Bromuro de etidio y se agitó vigorosamente y se vertió en la cubeta de electroforesis colocando un peine de 10 carriles hasta que solidificó, luego se retiró el peine. La cámara de electroforesis se cargó con el buffer TAE 1X. Después se colocó 1.5 μ l del ladder (1 kb) en el carril inicial. Para la preparación de las muestras se agregaron 3 μ l de RNA y 3 μ l del buffer de carga y se cargó en los carriles correspondientes. La electroforesis se realizó a 90 V. durante 30 min.

2.1.6 Liofilización de muestras de RNA de *Cordia sebestena*

Las muestras fueron liofilizadas antes de ser enviadas en tubos eppendorf de 1.5 ml con 20 μ l de RNA por cada muestra. El equipo empleado para esto fue LABCONCO Freezone 4.5 Liter a una temperatura de -51°C y a una presión de 0.024 mBar (figura 2.3) durante 5 horas, con la finalidad de preservarlas y reducir la probabilidad de degradación del RNA.

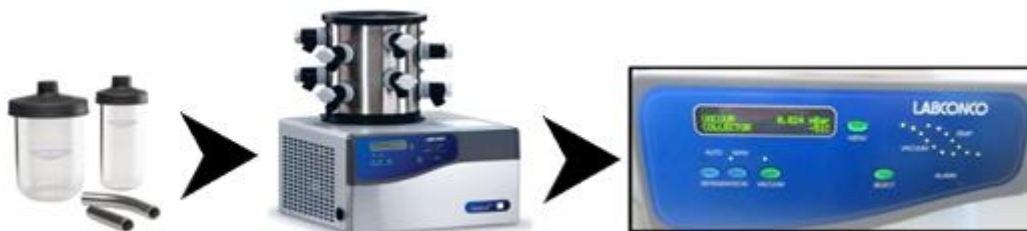


Figura 2.3. Diagrama de flujo para la liofilización de muestras pertenecientes a *Cordia sebestena* desarrolladas en un ámbito silvestre usando un equipo LABCONCO Freezone 4.5 Liter.

2.1.7 Secuenciación de RNA y control de calidad de las secuencias.

Las muestras se enviaron a la empresa Novogene en Sacramento California para su secuenciación utilizando la plataforma Illumina NovaSeq 6000. Se obtuvieron alrededor de 30 millones de reads por cada muestra de *C. sebestena*, en modo Paired end y con una longitud de 150pb.

Tabla 2.1. Requerimientos de la compañía Novogene para secuenciación de RNA

| Tipo de muestra | Recomendaciones | Cantidad de RNA | RIN | Concentración | Pureza |
|-----------------|-------------------------|----------------------|---------------------|----------------------------------|--|
| RNA | Fuertemente recomendado | $\geq 2 \mu\text{g}$ | $\geq 6.8/\geq 6.3$ | $\geq 20 \text{ ng}/\mu\text{l}$ | OD260/280 > 2.0 sin degradación ni contaminación por DNA |
| | Requerido | $\geq 1 \mu\text{g}$ | $\geq 6.8/\geq 6.3$ | | |

Se obtuvieron los datos en forma de archivos en formato fastq y fueron analizados por calidad utilizando el programa FastQC versión 0.11.9 (Andrews, 2010). Este programa otorgó un puntaje Phred a cada nucleótido, y se utilizó un punto de corte de Phred ≥ 20 (Blankenberg *et al.*, 2010; Verk *et al.*, 2013).

Línea de comando para ejecutar FastQC:

```
$ fastqc -f <formato de archivo fastq> <dirección que contiene el archivo de las lecturas a analizar>
```

2.1.8 Ensamblado *de novo* del transcriptoma

Para el ensamblado *de novo* se utilizó la herramienta Trinity versión 2.11.0 (Grabherr *et al.*, 2011), a través del equipo de cómputo en paralelo (HPC) “Hobon” del Centro de Investigación Científica de Yucatán. Trinity combina tres módulos de software independientes aplicados secuencialmente para procesar grandes volúmenes de lecturas de RNA-seq: (i) Inchworm: reúne los datos de secuencia de RNA en secuencias únicas de transcripciones, a menudo generando transcripciones completas para una isoforma dominante, pero luego informa solo las porciones únicas de las transcripciones empalmadas alternativamente; (ii) Chrysalis: agrupa los contigs de Inchworm en grupos y construye gráficos completos de Bruijn para cada grupo. Cada grupo representa la complejidad transcripcional completa para un gen determinado (o conjuntos de genes que comparten secuencias en común), luego divide el conjunto de lectura completa entre estos gráficos; (iii) Butterfly: procesa los gráficos individuales en paralelo, rastreando los caminos que las lecturas y los pares de lecturas toman dentro del gráfico, en última instancia, reporta transcripciones completas para isoformas empalmadas alternativamente y separa las transcripciones que corresponden a genes parálogos (Grabherr *et al.*, 2011). Se realizó el ensamblado de dos transcriptomas uno perteneciente al morfo de brevistilo y el otro al morfo longistilo (tabla 2.5). El tamaño mínimo de contig a ensamblar se estableció en 500 pb de acuerdo a lo sugerido en la literatura, esto debido a que es más probable que los transcritos más pequeños sean menos útiles para el proceso estándar de análisis de RNAseq.

Línea de comando para la ejecución de Trinity:

```
/Trinity --seqType fq <Formato del archivo fastq de lecturas> --max_memory <Memoria RAM usada> --left reads_1.fq <lectura forward> --right reads_2.fq <lectura reverse> --CPU <número de cores a usar> --output < archivo de salida, con los estadísticos de ensamblado>
```


Tabla 2.2. Muestras pertenecientes a árboles de *Cordia sebestena* que serán empleadas para el ensamblado del transcriptoma del morfo brevistilo y longistilo.

| Morfo | Muestras a ensamblar | | Tamaño mínimo contig |
|------------|----------------------|---------------------------------------|----------------------|
| Brevistilo | B2 | Brevistilo Flor en antesis androceo | 500 pb |
| | B6 | Brevistilo Flor en antesis gineceo | |
| | B7 | Brevistilo botón pre-antesis androceo | |
| | B11 | Brevistilo botón pre-antesis gineceo | |
| | BB16 | Brevistilo yema floral androceo | |
| | B17 | Brevistilo yema floral gineceo | |
| Longistilo | LFA | Longistilo Flor en antesis androceo | |
| | LFG | Longistilo Flor en antesis gineceo | |
| | L26 | Longistilo botón pre-antesis androceo | |
| | L34 | Longistilo botón pre-antesis gineceo | |
| | L32 | Longistilo yema floral androceo | |
| | L33 | Longistilo yema floral gineceo | |

2.1.9 Evaluación de la integridad de los transcritos ensamblados

Para evaluar la calidad del transcriptoma ensamblado *de novo* se calculó el valor de N50, el tamaño del contig promedio, y el número de transcritos obtenidos. Adicionalmente, se mapearon las lecturas crudas contra el transcriptoma ensamblado para poder estimar el porcentaje de mapeo y evaluar la integridad de nuestro transcriptoma ensamblado *de novo*, se recomienda un porcentaje >50%.

Línea de comando para la obtención de estadísticos Nx de Trinity:

```
/opt/software/trinityrnaseq/2.4.0/util/TrinityStats.pl /path_file/ Trinity_out.Trinity.fasta
```

2.1.10 Anotación funcional del transcriptoma de *C. sebestana*

Para realizar la anotación funcional en los transcriptomas de morfo brevisitlo y longistilo de *Cordia sebestena* primero se empleó el software Transdecoder versión 5.5.0 (Haas y Papanicolaou, 2016) a través de la plataforma web Galaxy v 22.05 (Afgan *et al.*, 2018) para la traducción de los transcritos a secuencias de aminoácidos. Se estableció una longitud mínima de ORF de 100 aminoácidos (-m 100) debido a que la longitud mínima de ORF reportada es de 96 pb (32 aminoácidos). Se empleó el código genético universal (-G universal) y se generaron los siguientes archivos de salida: (i) archivo con las secuencias de péptidos para los ORF candidatos finales (.fasta.transdecoder.pep) el cuál fue empleado para realizar los análisis de anotación en diferentes bases de datos de plantas; (ii) posiciones dentro de los transcritos provistos por los ORF finales seleccionados (.fasta.transdecoder.gff3); (iii) archivo de secuencia de nucleótidos para todos los ORF candidatos finales (.fasta.transdecoder.cds).

Línea de comando a empleada para ejecutar Transdecoder:

```
TransDecoder.LongOrfs -t 'transcripts.fasta' -G 'Universal' -m 100 TransDecoder.Predict -t 'transcripts.fasta' -G 'Universal' -O 'output'
```

El programa BLAST versión 2.11.0 (Altschul, *et al.*, 1990) fue utilizado para identificar las secuencias traducidas a aminoácidos utilizando el proteoma predicho de *Arabidopsis thaliana* (TAIR11) (Cheng *et al.*, 2017) y la base de datos NR Viridiplantae Refseq del NCBI (Schoch *et al.*, 2020). Se estableció un E-value mínimo de 1E-05, para identificar genes ortólogos.

Línea de comando a empleada para ejecutar Blast:

```
blastp db <Base_datos> -evalue 0.00001 -outfmt '6 qseqid salltitles pident length mismatch gapopen qstart qend sstart send evalue bitscore' -max_target_seqs 5 -query Transdecoder.pep.fa -out
```

2.1.11 Análisis de enriquecimiento funcional

Se realizó un análisis de enriquecimiento para encontrar qué términos GO se encontraban sobrerrepresentados. Se empleó la página PANTHER (www.pantherdb.org) en su versión 17.0. Pantherdb utiliza los términos GO-slim (versiones reducidas de las ontologías GO que contienen un subconjunto de términos del recurso GO) (Huaiyu M. y Paul T., 2009; Huaiyu *et al.*, 2019). Utilizando la herramienta de análisis de lista de genes se cargó la lista de genes con los ids obtenidos al anotar los transcriptomas de los morfos longistilo y brevistilo contra el proteoma predicho de *A. thaliana*. Se seleccionó al organismo de referencia *Arabidopsis thaliana*, por último, se seleccionó la opción “Clasificación funcional vista en gráficos”.

CAPÍTULO III

RESULTADOS

3.1 Caracterización del transcriptoma de *Cordia sebestena*.3.1.1 Pureza e integridad de RNA en muestras de *C. sebestena*.

Se aisló el RNA total de 30 muestras pertenecientes a androceo (antera y filamento) y gineceo (estigma, estilo y ovario), para ambos morfos florales (brevistilo y longistilo) y para cada uno de sus estadios de desarrollo (yema floral, botón pre-antesis y antesis floral) (tabla 3.1). Las muestras **L19** y **L24** se juntaron en una sola (LFA, Longistilo Flor Androceo) y se procedió de la misma manera con las muestras **L22** y **L23** (LFG, Longistilo Flor Gineceo) para poder cumplir con los requisitos para su secuenciación ya que la concentración del RNA se encontraba por debajo del requerido por la empresa Novogene.

Tabla 3.1. Muestras y tejidos empleados para la extracción de RNA total de *Cordia sebestena*.

| ID Muestra | Estadio | Morfo | Tejido | No. réplica |
|-------------|-------------------|------------|----------|-------------|
| BB1 | Flor en antesis | Brevistilo | Androceo | R1 |
| B2 | Flor en antesis | Brevistilo | Androceo | R2 |
| B5 | Flor en antesis | Brevistilo | Androceo | R3 |
| B3 | Flor en antesis | Brevistilo | Gineceo | R1 |
| B4 | Flor en antesis | Brevistilo | Gineceo | R2 |
| B6 | Flor en antesis | Brevistilo | Gineceo | R3 |
| B7 | Botón pre-antesis | Brevistilo | Androceo | R1 |
| B8 | Botón pre-antesis | Brevistilo | Androceo | R2 |
| BB10 | Botón pre-antesis | Brevistilo | Androceo | R3 |
| B9 | Botón pre-antesis | Brevistilo | Gineceo | R1 |
| B11 | Botón pre-antesis | Brevistilo | Gineceo | R2 |
| B12 | Botón pre-antesis | Brevistilo | Gineceo | R3 |
| B14 | Yema floral | Brevistilo | Androceo | R1 |
| BB16 | Yema floral | Brevistilo | Androceo | R3 |

| | | | | |
|-------------|-------------------|------------|----------|-------|
| BB13 | Yema floral | Brevistilo | Gineceo | R1 |
| B17 | Yema floral | Brevistilo | Gineceo | R2 |
| LFA | Flor en antesis | Longistilo | Androceo | R1+R3 |
| L20 | Flor en antesis | Longistilo | Androceo | R2 |
| LFG | Flor en antesis | Longistilo | Gineceo | R2+R3 |
| L26 | Botón pre-antesis | Longistilo | Androceo | R1 |
| L28 | Botón pre-antesis | Longistilo | Androceo | R3 |
| L25 | Botón pre-antesis | Longistilo | Gineceo | R1 |
| L34 | Botón pre-antesis | Longistilo | Gineceo | R2 |
| L35 | Botón pre-antesis | Longistilo | Gineceo | R3 |
| LL29 | Yema floral | Longistilo | Androceo | R1 |
| L30 | Yema floral | Longistilo | Androceo | R2 |
| L32 | Yema floral | Longistilo | Androceo | R3 |
| L31 | Yema floral | Longistilo | Gineceo | R1 |
| L33 | Yema floral | Longistilo | Gineceo | R2 |
| L36 | Yema floral | Longistilo | Gineceo | R3 |

Después del tratamiento con DNase I, la integridad del RNA se verificó tras la electroforesis, en geles de agarosa, se observaron bandas pertenecientes a las unidades ribosomales 28S y 18S para todas las muestras (figura 3.1). L (Ladder 1kb); para el morfo Brevistilo: **BB1**, **B2** y **B5** (Androceo de antesis floral); **B3**, **B4** y **B6** (Gineceo de antesis floral). **B7**, **B8** y **BB10** (Androceo de botón pre-antesis); **BB13** y **B17** (Gineceo de botón pre-antesis). **B14** y **BB16** (Androceo de yema floral) **B9**, **B11**, **B12** (Gineceo de yema floral).

Respecto al morfo Longistilo: **19**, **24** y **L20** (Androceo de flor en antesis); **22** y **23** (Gineceo de flor en antesis). **L26** y **L28** (Androceo de botón pre-antesis). **L25**, **L34** y **L35** (Gineceo de botón pre-antesis). **LL29**, **L30** y **L32** (Gineceo de yema floral); **L31**, **L33** y **L36** (Androceo de yema floral).

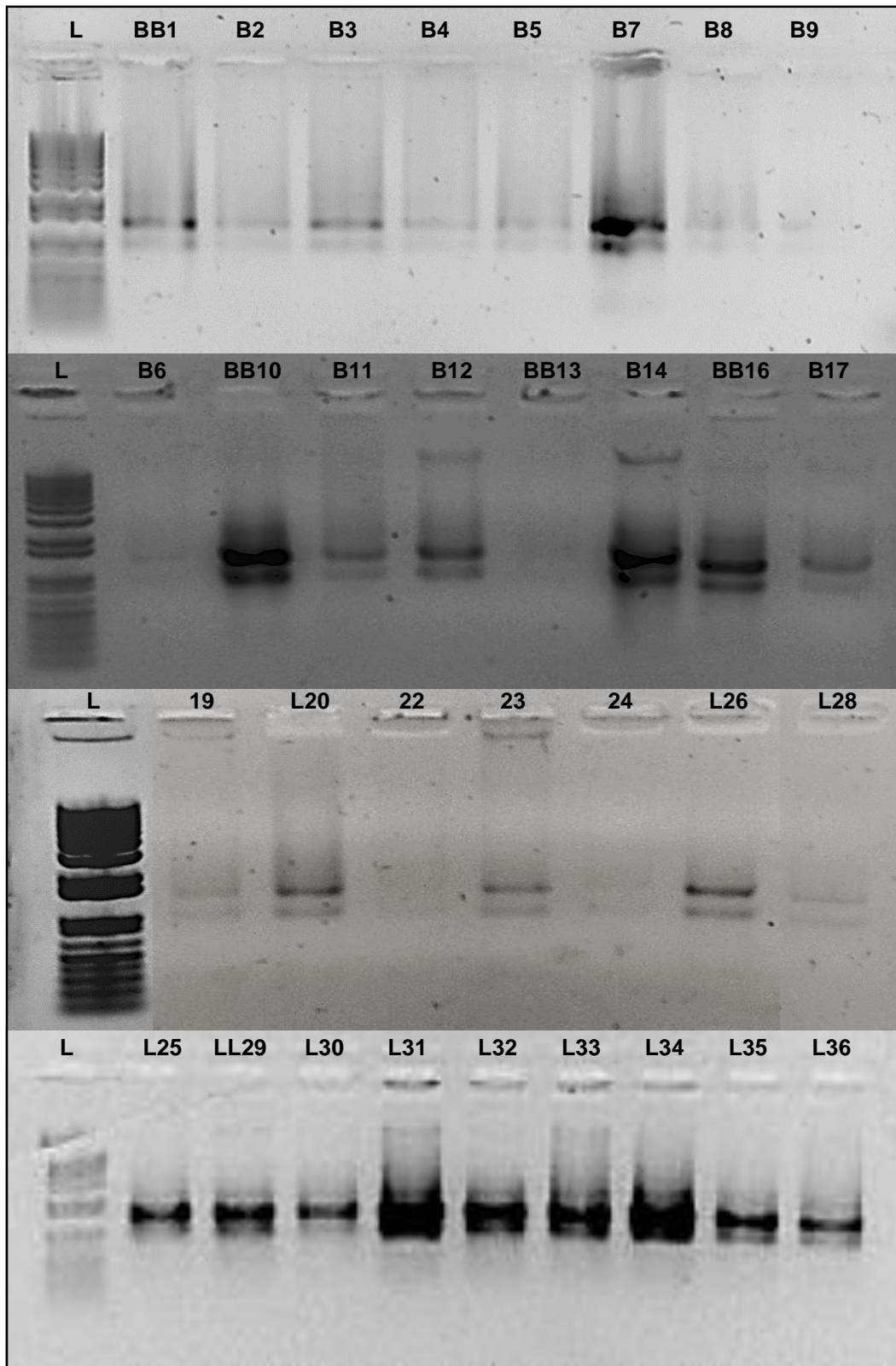


Figura 3.1. Geles de agarosa al 1% teñidos con bromuro de etidio. Muestras de RNA de *Cordia sebestena* después del tratamiento con DNase I, Amplification grade de la compañía Invitrogen.

Posteriormente utilizando los equipos Nanodrop one™ y Qubit ® 2.0 se evaluó la pureza de RNA por los métodos de absorbancia y fluorometría, respectivamente (tabla 3.2).

Tabla 3.2. Concentración y valores de absorbancia para las muestras de RNA total de *Cordia sebestena* así como la concentración y pureza requerida por la empresa Novogene.

| Muestra | Concentración de RNA (ng/μl) | 260/280 | 260/230 | Concentración requerida | Pureza de RNA requerida |
|---------|------------------------------|---------|---------|-------------------------|--|
| BB1 | 64 | 2.1 | 1.68 | 20 ng/μl | OD260/280 > 2.0 sin degradación ni contaminación por DNA |
| B2 | 26.04 | 2.17 | 1.26 | | |
| B5 | 249.91 | 1.99 | 0.87 | | |
| B3 | 190.32 | 1.86 | 0.98 | | |
| B4 | 138.64 | 2 | 0.49 | | |
| B6 | 80.93 | 1.95 | 0.7 | | |
| B7 | 297.53 | 1.78 | 0.86 | | |
| B8 | 161.25 | 1.93 | 1 | | |
| BB10 | 77 | 2.1 | 1.68 | | |
| B9 | 93.06 | 2 | 0.24 | | |
| B11 | 176.95 | 2.05 | 1.26 | | |
| B12 | 155.67 | 1.79 | 0.73 | | |
| B14 | 1105.1 | 2.11 | 1.49 | | |
| BB16 | 120 | 2.13 | 2 | | |
| BB13 | 146 | 2.09 | 1.97 | | |
| B17 | 60.96 | 1.54 | 0.57 | | |
| LFA | 48.62 | 2.09 | 0.13 | | |
| L20 | 107.1 | 2.12 | 0.5 | | |

| | | | | | |
|-------------|---------|------|------|--|--|
| LFG | 55.79 | 1.94 | 0.41 | | |
| L26 | 142.5 | 2.07 | 0.99 | | |
| L28 | 23.94 | 1.95 | 0.47 | | |
| L25 | 69.88 | 2.01 | 0.64 | | |
| L34 | 230.74 | 1.79 | 0.68 | | |
| L35 | 483.01 | 2.01 | 1.16 | | |
| LL29 | 164.08 | 2 | 1.34 | | |
| L30 | 275 | 2.07 | 2.32 | | |
| L32 | 236.55 | 1.91 | 1.01 | | |
| L31 | 1029.26 | 2 | 1.22 | | |
| L33 | 224.12 | 2 | 1.75 | | |
| L36 | 249.06 | 1.93 | 1.56 | | |

De acuerdo con las especificaciones del número de integridad recibidas por la empresa Novogene para la secuenciación del RNA, se obtuvieron números de integridad (RIN) con valores entre 4.1 y 9.1 en las muestras de plantas de *Cordia sebestena* (tabla 3.3; figura 3.5). La escala de valores de RIN va del 1 al 10 (donde 10 representa un RNA integro y 1 un RNA degradado) (Dumur *et al.*, 2004), por lo que los resultados se interpretan como un RNA integro (pureza, concentración de RNA, número de integridad). Debido a esto, se procedió a secuenciar el transcriptoma de estructuras florales de *Cordia sebestena*.

Tabla 3.3. Resultados de pureza y concentración de RNA, así como el número de integridad de RNA obtenido en muestras de *Cordia sebestena*.

| Muestra | Concentración de RNA (ng/μl) | 260/280 | 260/230 | RIN |
|----------------|---|----------------|----------------|------------|
| BB1 | 64 | 2.1 | 1.68 | 6.1 |
| B2 | 26.04 | 2.17 | 1.26 | 7.7 |
| B5 | 249.91 | 1.99 | 0.87 | 8.1 |
| B3 | 190.32 | 1.86 | 0.98 | 4.3 |
| B4 | 138.64 | 2 | 0.49 | 8.6 |
| B6 | 80.93 | 1.95 | 0.7 | 9.1 |
| B7 | 297.53 | 1.78 | 0.86 | 7.2 |
| B8 | 161.25 | 1.93 | 1 | 7.2 |
| BB10 | 77 | 2.1 | 1.68 | 4.1 |
| B9 | 93.06 | 2 | 0.24 | 9.1 |
| B11 | 176.95 | 2.05 | 126 | 8.4 |
| B12 | 155.67 | 1.79 | 0.73 | 7.5 |
| B14 | 1105.1 | 2.11 | 1.49 | 4.5 |
| BB16 | 120 | 2.13 | 2 | 4.7 |
| BB13 | 146 | 2.09 | 1.97 | 8.7 |
| B17 | 60.96 | 1.54 | 0.57 | 8.2 |
| LFA | 48.62 | 2.09 | 0.13 | 8.4 |
| L20 | 107.1 | 2.12 | 0.5 | 5.5 |
| LFG | 55.79 | 1.94 | 0.41 | 7.5 |
| L26 | 142.5 | 2.07 | 0.99 | 6.9 |
| L28 | 23.94 | 1.95 | 0.47 | 8.4 |

| | | | | |
|-------------|---------|------|------|-----|
| L25 | 69.88 | 2.01 | 0.64 | 8.2 |
| L34 | 230.74 | 1.79 | 0.68 | 8.8 |
| L35 | 483.01 | 2.01 | 1.16 | 8.2 |
| LL29 | 164.08 | 2 | 1.34 | 7.4 |
| L30 | 275 | 2.07 | 2.32 | 8.1 |
| L32 | 236.55 | 1.91 | 1.01 | 8.2 |
| L31 | 1029.26 | 2 | 1.22 | 8.4 |
| L33 | 224.12 | 2 | 1.75 | 8.1 |
| L36 | 249.06 | 1.93 | 1.56 | 8.1 |

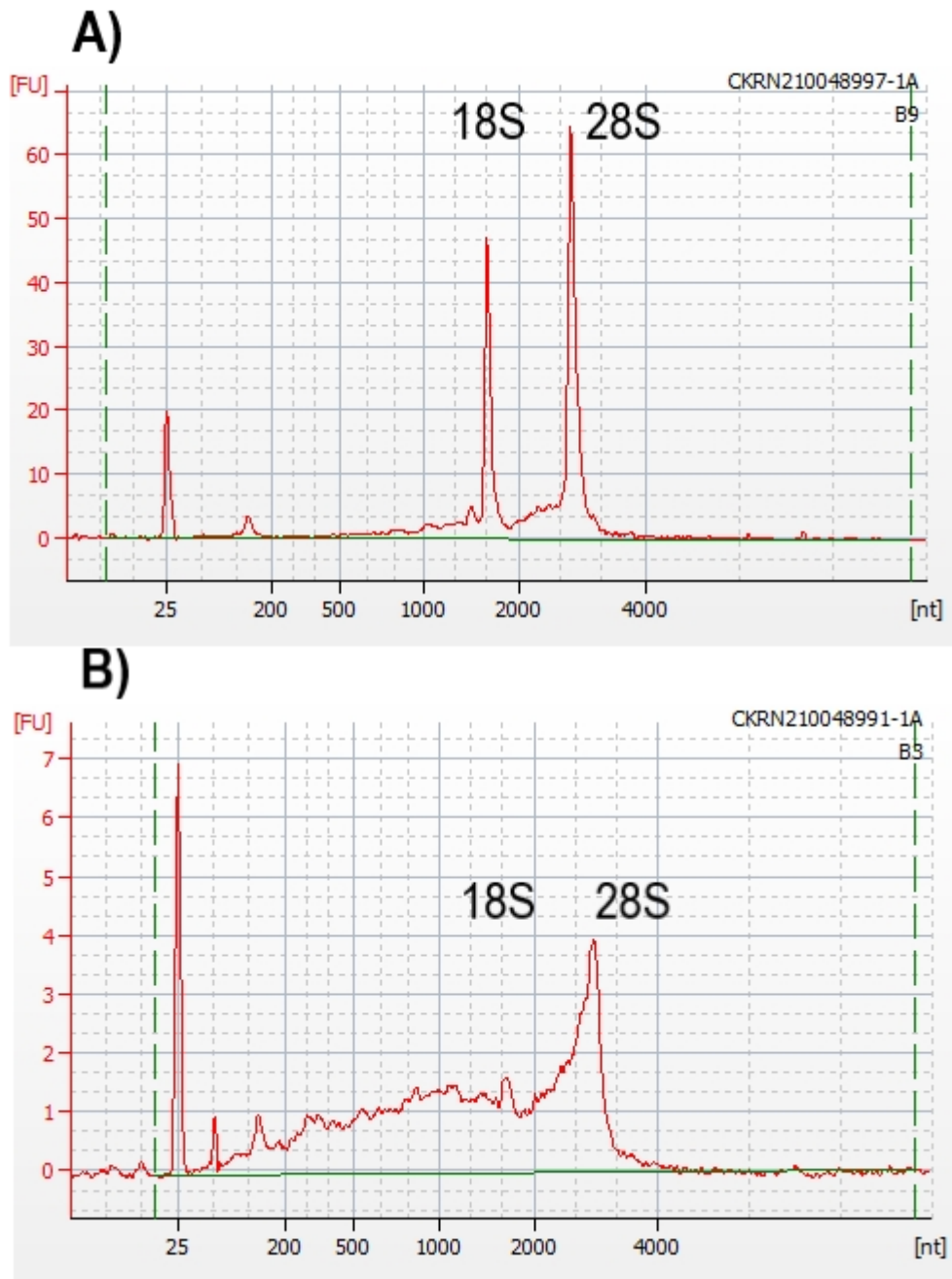


Figura 3.2. Electroferograma de la prueba de integridad RIN (RNA integrity number) Agilent 2100 Bioanalyzer®. **A)** Muestra B9 (Gineceo de botón pre-antesis del morfo Brevistilo) con un RIN de 9.2 **B)** Muestra B3 con un RIN de 4.3 (Gineceo de antesis floral del morfo Brevistilo). FU: Intensidad de fluorescencia, Nt: tamaño del producto en nucleótidos.

3.1.2 Secuenciación del transcriptoma de *Cordia sebestena*

Conforme al reporte de control de calidad de secuenciación enviado por la compañía Novogene, el tamaño total de salida de datos sin procesar de las 30 muestras pertenecientes a estructuras florales secuenciadas fue de 110 Gb. Se generaron más de 23 millones de lecturas crudas por muestra (tabla 3.4). La calidad de Phred promedio de las lecturas generadas fue de 36. Con base en los puntajes de calidad obtenidos en muestras de estructuras florales para cada estadio y morfo de *Cordia sebestena* se considera que el proceso de secuenciación del transcriptoma se realizó de manera eficiente obteniendo lecturas de alta calidad. El porcentaje de GC obtenido fue uniforme entre las muestras siendo en promedio de 44% (tabla 3.4). De igual manera se mantuvo dentro de la distribución teórica de acuerdo con las gráficas de calidad del programa FastQC en su versión 0.11.9.

Tabla 3.4 Lecturas y estadísticos resultado de la secuenciación en muestras de estructuras florales de *Cordia sebestena* colectadas del parque Gran Santa Fe.

| Muestra | # de Lecturas Crudas (Millones) | Tamaño del archivo (Gb) | GC (%) | Puntaje de calidad Phred |
|-------------|---------------------------------|-------------------------|--------|--------------------------|
| BB1 | 43,061,078 | 2.71 | 44.54 | 36 |
| B2 | 45,757,482 | 3.34 | 44.97 | 36 |
| B5 | 42,603,886 | 2.97 | 45.11 | 36 |
| B3 | 45,362,810 | 3.31 | 44.31 | 36 |
| B4 | 45,813,448 | 3.30 | 44.48 | 36 |
| B6 | 47,363,066 | 3.46 | 44.66 | 36 |
| B7 | 64,836,148 | 4.71 | 45.25 | 36 |
| B8 | 50,272,670 | 3.66 | 45.38 | 36 |
| BB10 | 44,916,280 | 2.94 | 43.81 | 36 |
| B9 | 41,102,930 | 3 | 44.36 | 36 |
| B11 | 48,003,268 | 3.49 | 43.98 | 36 |
| B12 | 46,594,680 | 3.37 | 44.42 | 36 |
| B14 | 44,779,648 | 3.27 | 45.12 | 36 |
| BB16 | 45,009,404 | 2.92 | 43.64 | 36 |

| | | | | |
|-------------|------------|------|-------|----|
| BB13 | 46,594,680 | 2.86 | 44.42 | 36 |
| B17 | 47,381,854 | 3.47 | 44.53 | 36 |
| LFA | 45,317,054 | 3.31 | 44.44 | 36 |
| L20 | 40,189,404 | 2.92 | 45.45 | 36 |
| LFG | 42,666,872 | 3.11 | 45.1 | 36 |
| L26 | 46,678,784 | 3.39 | 45.06 | 36 |
| L28 | 43,871,712 | 3.19 | 44.83 | 36 |
| L25 | 39,696,102 | 2.89 | 44.78 | 36 |
| L34 | 46,575,890 | 3.29 | 44.51 | 36 |
| L35 | 40,123,006 | 2.82 | 44.66 | 36 |
| LL29 | 43,990,552 | 2.79 | 44.64 | 36 |
| L30 | 55,210,666 | 4.06 | 46.07 | 36 |
| L32 | 61,325,454 | 4.32 | 45.55 | 36 |
| L31 | 51,103,864 | 3.70 | 44.8 | 36 |
| L33 | 77,958,916 | 5.53 | 45.05 | 36 |
| L36 | 57,340,714 | 4.09 | 44.62 | 36 |

3.1.3 Análisis de la calidad de secuencias en muestras de estructuras florales de *Cordia sebestena*.

Se realizó un análisis de calidad de las secuencias de muestras de estructuras florales de *Cordia sebestena* con la herramienta FastQC (versión 0.11.9; Andrews, 2010). De acuerdo con el reporte generado por el programa, la distribución de la calidad de las secuencias tuvo un valor Phred promedio de 36 (tabla 3.5) coincidiendo con los valores de calidad reportados por la compañía Novogene en su reporte de calidad (tabla 3.4). Adicionalmente, no se encontró contenido de adaptadores. Se detectaron dos secuencias sobrerrepresentadas, una correspondía a la cadena de poly T y la otra a copolímeros con una longitud de 50 pb, en las muestras BB1, BB10, BB16 y LL29 el porcentaje fue menor al 0.1% de las secuencias totales. Se ejecutó un BLASTp y se empleó la herramienta UNIVeC (base de datos de adaptadores más utilizados) para encontrar algún hit posible relacionado con contaminación sin obtener coincidencias.

Tabla 3.5. Resultados del análisis de calidad de lecturas con la herramienta FastQC en muestras de estructuras florales de *Cordia sebestena*.

| Muestra Código Forward_1/ Reverse_2 | Número de secuencias por muestra. | GC (%) | Puntajes de calidad Phred |
|--|--|---------------|--------------------------------------|
| BB1_1/ BB1_2 | 21,530,539 | 44 | 36/36 |
| B2_1/ B2_2 | 22,878,741 | 45 | 36/36 |
| B5_1/ B5_2 | 21,044,200 | 45 | 36/36 |
| B3_1/ B3_2 | 22,681,405 | 44 | 36/36 |
| B4_1/ B4_2 | 22,906,724 | 44 | 36/36 |
| B6_1/ B6_2 | 23,681,533 | 44 | 36/36 |
| B7_1/ B7_2 | 32,418,074 | 45 | 36/36 |
| B8_1/ B8_2 | 25,136,335 | 45 | 36/36 |
| BB10_1/ BB10_2 | 22,458,140 | 43 | 36/36 |
| B9_1/ B9_2 | 20,551,465 | 44 | 36/36 |
| B11_1/ B11_2 | 24,001,634 | 44 | 36/36 |
| B12_1/ B12_2 | 23,297,340 | 44 | 36/36 |
| B14_1/ B14_2 | 22,389,824 | 45 | 36/36 |
| BB16_1/ BB16_2 | 22,504,702 | 43 | 36/36 |
| BB13_1/ BB13_2 | 21,933,212 | 44 | 36/36 |
| B17_1/ B17_2 | 23,690,927 | 44 | 36/36 |
| LFA_1/ LFA_2 | 22,658,527 | 44 | 36/36 |
| L20_1/ L20_2 | 20,094,702 | 45 | 36/36 |
| LFG_1/ LFG_2 | 21,333,436 | 45 | 36/36 |
| L26_1/ L26_2 | 23,339,392 | 45 | 36/36 |
| L28_1/ L28_2 | 21,935,856 | 44 | 36/36 |
| L25_1/ L25_2 | 19,848,051 | 44 | 36/36 |
| L34_1/ L34_2 | 23,287,945 | 44 | 36/36 |
| L35_1/ L35_2 | 20,061,503 | 44 | 36/36 |
| LL29_1/ LL29_2 | 21,995,276 | 44 | 36/36 |
| L30_1/ L30_2 | 27,605,333 | 46 | 36/36 |

| | | | |
|---------------------|------------|----|-------|
| L32_1/ L32_2 | 30,662,727 | 45 | 36/36 |
| L31_1/ L31_2 | 25,551,932 | 44 | 36/36 |
| L33_1/ L33_2 | 38,979,458 | 45 | 36/36 |
| L36_1/ L36_2 | 28,670,357 | 44 | 36/36 |

3.1.4 Eliminación de adaptadores y secuencias de baja calidad.

Por la baja cantidad de secuencias sobrerrepresentadas (menor al 0.1%) encontradas en solo cuatro de las 30 muestras totales, se decidió no realizar el paso de filtrado con la herramienta Trimmomatic debido a que se tenían lecturas de calidad ≥ 20 según el parámetro Phred.

3.1.5 Ensamblado *de novo* de los transcriptomas de muestras de estructuras florales de *Cordia sebestena*

Se realizó un ensamblado del transcriptoma para cada morfo floral de *Cordia sebestena*, se eligió una muestra por cada tejido (masculino y femenino) y para cada estadio (botón cerrado, botón pre-antesis y flor en antesis) con un total de 6 muestras por cada ensamblado. El criterio de selección fue las muestras con mayor cantidad de reads (cuadro 9 y 10). Para el ensamblado se utilizó un tamaño mínimo de contig de 500 pb en el programa Trinity en su versión 2.11.0 (Haas et al., 2011). Los datos estadísticos de los ensamblados *de novo* fueron obtenidos con el script “trinitystats”. Los valores obtenidos de N50 fueron de 2,678 pb para el morfo brevistilo y 2,560 pb para el morfo longistilo, con un tamaño promedio de contig de 2,045.28 y 1,988.34 pb respectivamente (tabla 3.6).

Tabla 3.6. Muestras elegidas para el ensamblado *de novo* del transcriptoma floral de *Cordia sebestena* y sus estadísticos.

| Morfo brevistilo | | | | |
|-------------------------|--|--------------------------------|---------------|---------------------------------|
| Muestra | # de Lecturas Crudas (Millones) | Tamaño del archivo (Gb) | GC (%) | Puntaje de calidad Phred |
| B2 | 45,757,482 | 3.34 | 44.97 | 36 |
| B6 | 47,363,066 | 3.46 | 44.66 | 36 |
| B7 | 64,836,148 | 4.71 | 45.25 | 36 |
| B11 | 48,003,268 | 3.49 | 43.98 | 36 |
| BB16 | 45,009,404 | 2.92 | 43.64 | 36 |
| B17 | 47,381,854 | 3.47 | 44.53 | 36 |

| Morfo longistilo | | | | |
|-------------------------|--|--------------------------------|---------------|---------------------------------|
| Muestra | # de Lecturas Crudas (Millones) | Tamaño del archivo (Gb) | GC (%) | Puntaje de calidad Phred |
| LFA | 45,317,054 | 3.31 | 44.44 | 36 |
| LFG | 42,666,872 | 3.11 | 45.1 | 36 |
| L26 | 46,678,784 | 3.39 | 45.06 | 36 |
| L34 | 46,575,890 | 3.29 | 44.51 | 36 |
| L32 | 61,325,454 | 4.32 | 45.55 | 36 |
| L33 | 77,958,916 | 5.53 | 45.05 | 36 |

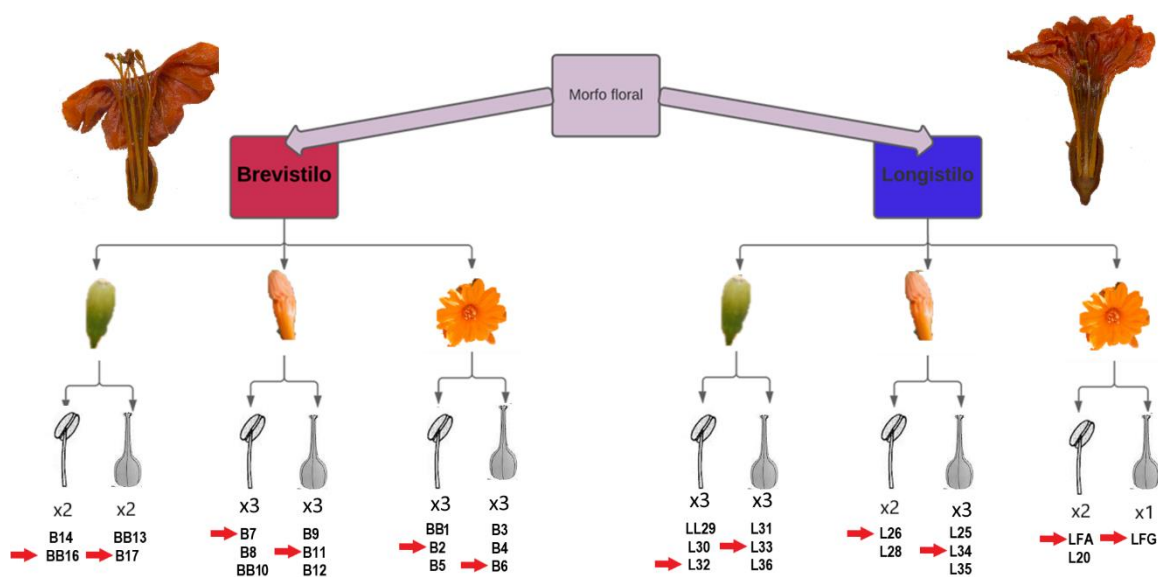


Figura 3.3. Diseño experimental. En rojo se señala las muestras elegidas para el ensamblado *de novo* de los transcriptomas de *Cordia sebestena*.

Tabla 3.7. Datos estadísticos de los transcriptomas ensamblados *de novo* de los morfos florales longistilo y brevistilo de *Cordia sebestena*.

| Resultados | Transcriptomas ensamblados | |
|-----------------------------------|----------------------------|------------------|
| | Morfo brevistilo | Morfo longistilo |
| Total Trinity 'genes': | 44, 869 | 41, 709 |
| Total trinity transcripts: | 126, 260 | 121, 871 |
| Max len | 19,193 | 18,077 |
| Min len | 499 | 499 |
| Percent GC | 39.96 | 40.13 |
| N50 | 2,678 | 2, 560 |
| Average contig | 2, 045.28 | 1, 988.34 |

Por último, se realizó un análisis de agrupamiento de transcritos a través de CD-Hit (Li y Godzik, 2006) para evaluar la redundancia de transcritos en ambos transcriptomas. Como se puede observar en la figura 3.7, posterior a la agrupación de los transcritos, en ambos transcriptomas se agrupó el 30% de los transcritos.

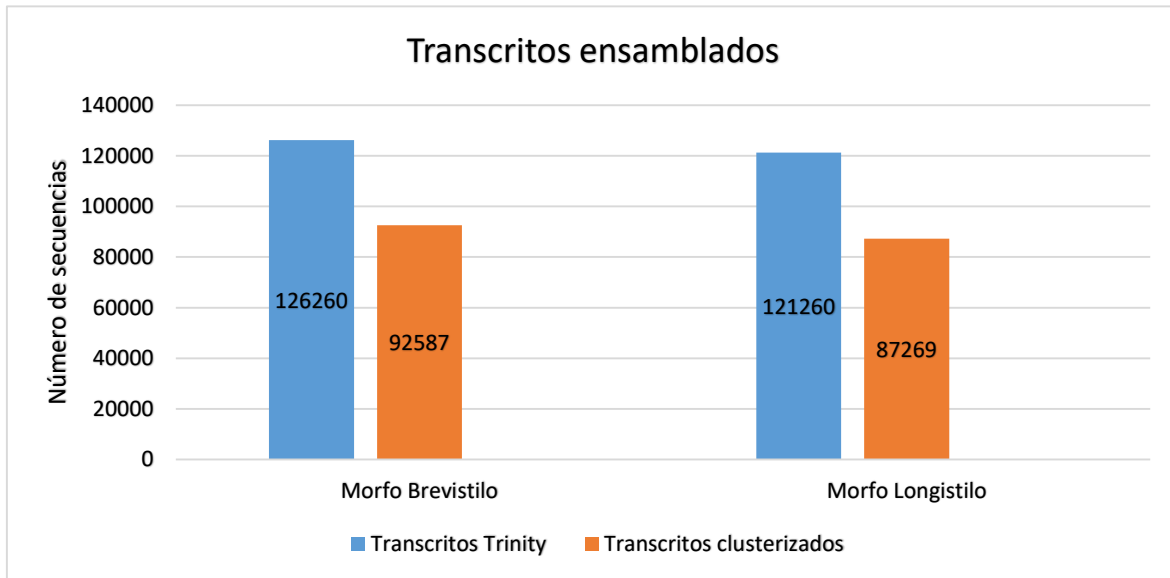


Figura 3.4. Transcritos conservados posterior al agrupamiento a través de CD-Hit v.4.8.1 a través del equipo de súper cómputo Hobón en los transcriptomas de los morfos brevistilo y longistilo de *Cordia sebestena*.

3.1.6 Anotación funcional de los transcriptomas florales de *Cordia sebestena*

Se empleó la herramienta Transdecoder (Haas *et al.*, 2016) para traducir las secuencias ensambladas *de novo* de cada transcriptoma a secuencias de aminoácidos para su posterior alineación contra diferentes bases de datos de proteínas de plantas. Se obtuvo un total de 115,786 y 107, 833 secuencias de aminoácidos traducidas en los transcriptomas del morfo brevistilo y longistilo, respectivamente (tabla 3.8).

Tabla 3.8. Porcentajes de anotación en los transcriptomas del morfo brevistilo y longistilo de *Cordia sebestena* contra bases de datos de proteínas de plantas a través de BLAST versión 2.11.0.

| Resultados | Transcriptomas ensamblados | |
|--------------------------|----------------------------|------------------|
| | Morfo brevistilo | Morfo longistilo |
| Secuencias de a.a. total | 115,786 | 107,833 |
| Araport11 | 59.48% | 59.73% |
| NR Viridiplantae | 43.77% | 36.53% |

Los transcriptomas para los morfos brevistilo y longistilo de *Cordia sebestena* fueron anotados empleando las bases de datos de proteínas de plantas el proteoma de *Arabidopsis thaliana* (TAIR11) (Cheng *et al.*, 2017) (**Figura 3.7** y **Tabla 3.8**). La proporción de secuencias de los transcriptomas de los morfos brevistilo y longistilo de *Cordia sebestena* anotadas contra el proteoma de *Arabidopsis thaliana* fueron muy similares del 59.48% (68,875) y 59.73% (64,415), respectivamente. La base de datos de proteínas de plantas del NCBI (Viridiplantae) mostró proporciones de 43.77% (50,682) y 36.53% (39,398), respectivamente.

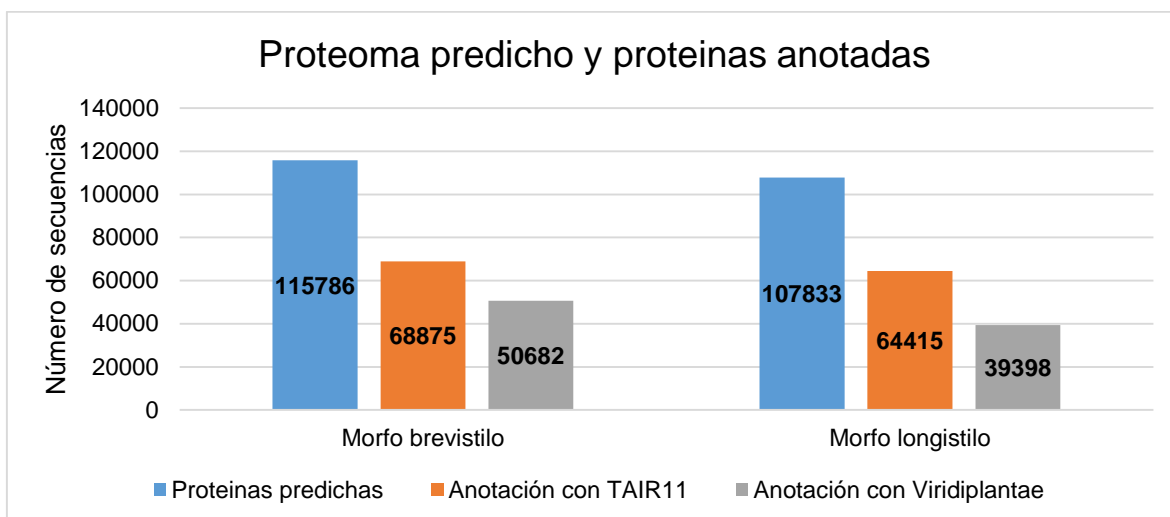


Figura 3.5. Porcentaje de secuencias anotadas contra diferentes bases de datos de proteínas de plantas de los proteomas predichos de los morfos brevistilo y longistilo de *Cordia sebestena* a partir de sus transcriptomas ensamblados.

3.1.7 Análisis de enriquecimiento funcional

Se realizó un análisis de enriquecimiento de Gene Ontology (GO) utilizando la herramienta “gene list analysis” del sitio pantherdb.org. El análisis de enriquecimiento de genes permite realizar una validación, con base en conocimiento biológico previo a través de resultados obtenidos por técnicas de agrupación de genes (Clustering y Biclustering) sobre bases de datos de expresión genética, siendo las características de interés las categorías funcionales de los genes (Mi *et al.*, 2019). La anotación de los términos GO para ambos transcriptomas (brevistilo y longistilo) de *Cordia sebestena*, fue realizada empleando la anotación previamente obtenida tras ejecutar BLASTp contra el proteoma completo de *Arabidopsis Thaliana* (TAIR11). De las 68,875 secuencias previamente anotadas en el transcriptoma del morfo brevistilo contra el proteoma de *Arabidopsis thaliana*, se asignaron términos a 11,818 secuencias (17%). En el caso del transcriptoma del morfo longistilo de las 64,415 secuencias previamente anotadas, se asignaron términos para 11,689 secuencias (18%). Dentro de la categoría de Función Molecular se obtuvieron 5,495 y 5,393 hits en los transcriptomas de los morfos brevistilo y longistilo, respectivamente, distribuidas en 11 grupos (figura 3.9). Actividad catalítica y actividad de unión fueron los grupos con la más alta representación en el transcriptoma del morfo brevistilo con 44.6% y 31.6 (2,449 hits y 1,739 hits, respectivamente) y en el transcriptoma del morfo longistilo con 44.8% y 31.5% (2,417 hits y 1,699 hits, respectivamente).

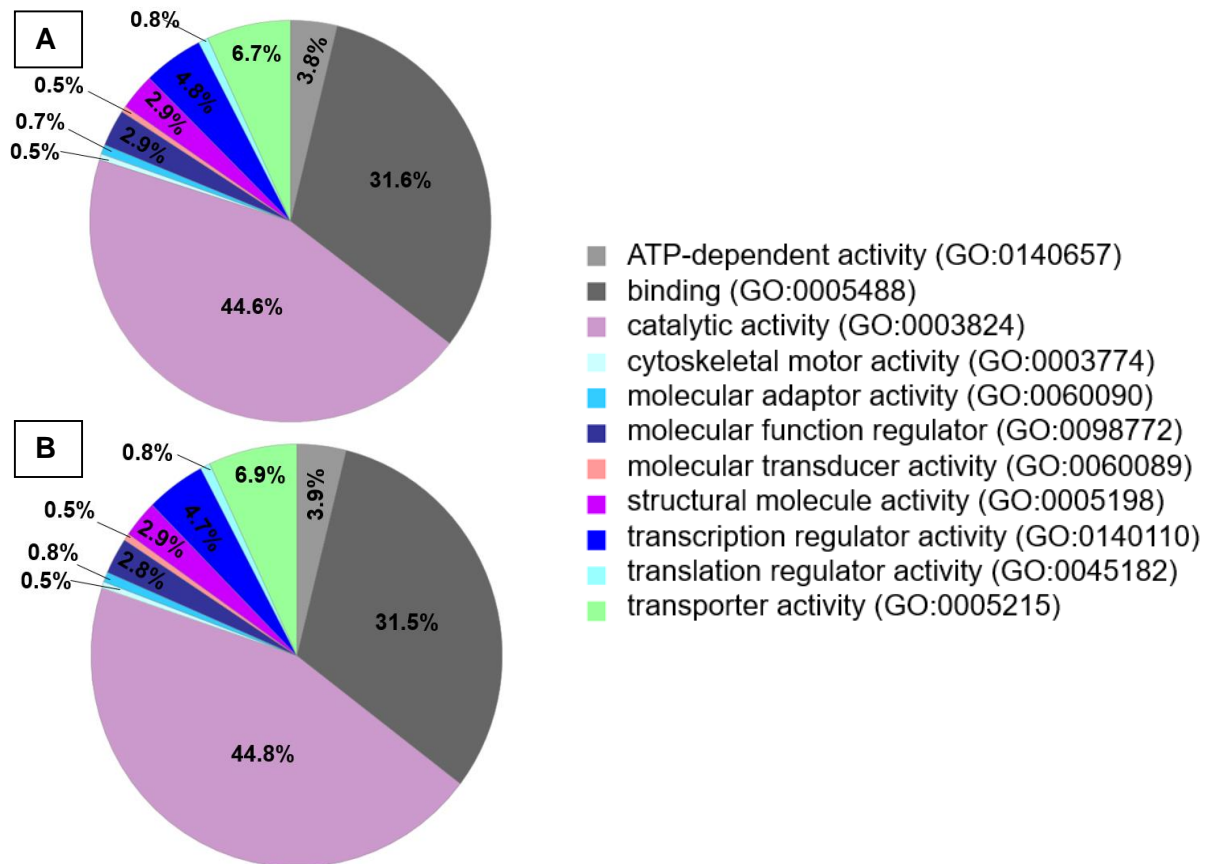


Figura 3.6 Términos de GO asignados en la categoría de Función Molecular en los transcriptomas de los A) morfos brevistilo y B) longistilo *Cordia sebestena* a partir de la anotación funcional realizada con el proteoma predicho de *Arabidopsis thaliana* (TAIR11).

La categoría de Procesos Biológicos arrojó 11,818 términos para el transcriptoma del morfo brevistilo y 11,689 términos para el transcriptoma del morfo longistilo, distribuidos en 17 grupos (figura 3.10). Se observó una proporción similar de los grupos más representativos siendo Procesos Celulares, Procesos metabólicos y Regulación Biológica en ambos transcriptomas. Para el transcriptoma del morfo brevistilo se obtuvieron 3,433 (38.4%), 2,612 (29.2%) y 1,030 hits (11.5%), respectivamente. En el transcriptoma del morfo longistilo fue de 3,380 (38.3%), 2,569 (29.1%) y 1,010 hits (11.4%), respectivamente. Es interesante observar la presencia del grupo Respuesta a Estímulos, siendo este el quinto más representativo con el 6.3% en el morfo brevistilo y 6.5% para el morfo longistilo con 563 hits y 570 hits respectivamente.

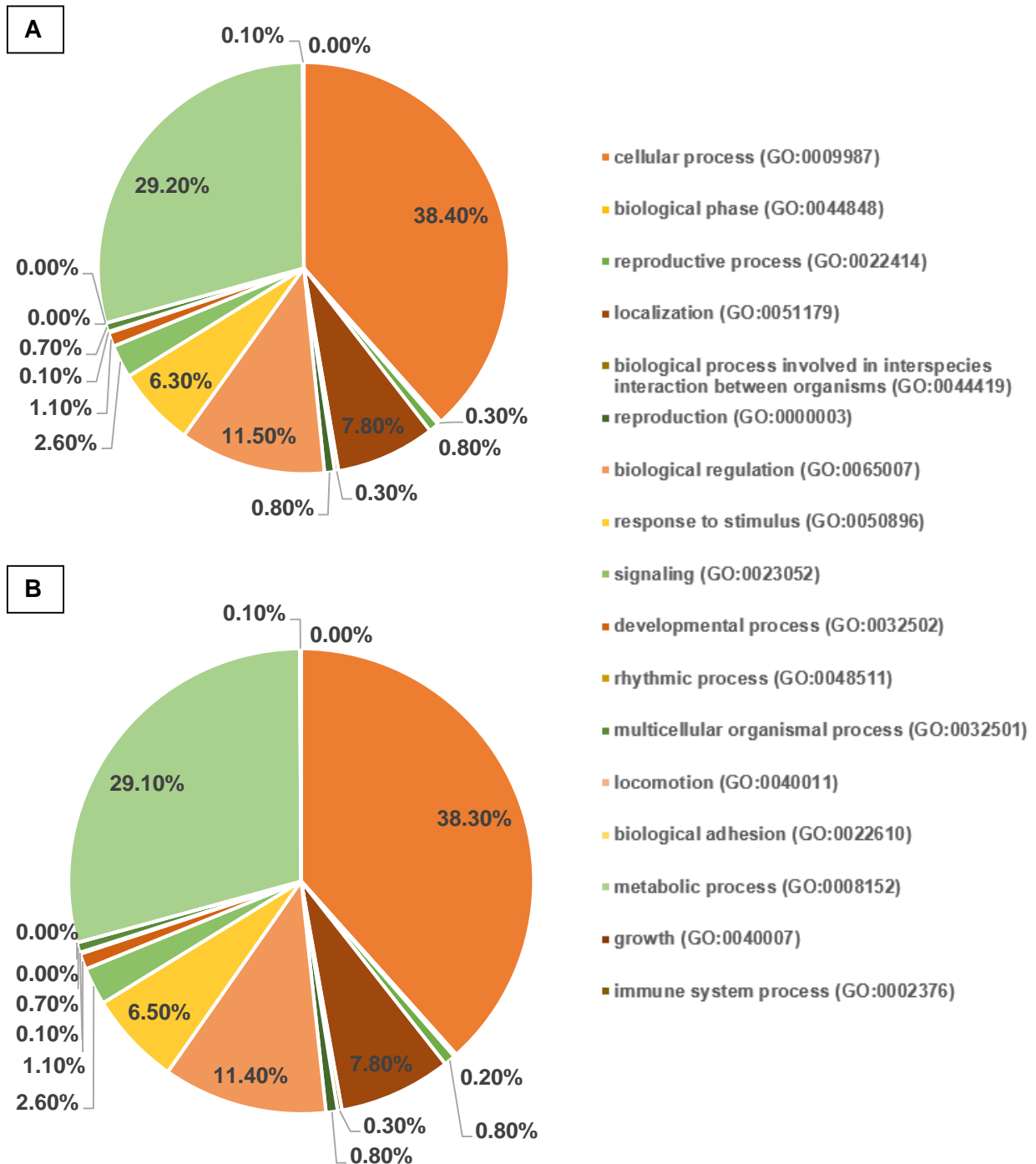


Figura 3.7. Términos GO asignados en la categoría de Procesos Biológicos para el transcriptoma de los morfos A) brevistilo y B) longistilo de *Cordia sebestena* a partir de la anotación funcional realizada con el proteoma de *Arabidopsis thaliana* (TAIR11).

Como se puede observar en ambas figuras (figura 3.9 y figura 3.10), la representación de los términos de GO para ambas categorías (Función Molecular y Proceso Biológico) son muy similares entre ambos morfos.

Se realizó una búsqueda comparativa en los resultados del análisis de enriquecimiento para los transcriptomas florales de los morfos brevistilo y longistilo de *Cordia sebestena*, posteriormente, realizando una revisión bibliográfica se identificaron aquellos relacionados con la distilia (tabla 3.9). Estos transcritos están relacionados a proteínas con funciones en la síntesis de fitohormonas, respuesta a estrés y desarrollo de la pared celular. Los términos de GO más representados dentro de la categoría de función molecular fueron: unión de lavín adenín dinuclótidos y actividad de la monooxygenasa. Para la categoría de procesos biológicos las categorías de GO más representadas fueron: respuesta a estímulo externo, ritmo circadiano, regulación del ritmo circadiano, proceso biosintético de compuestos orgánicos hidroxilados, proceso biosintético hormonal, proceso biosintético de esteroides, proceso metabólico de los fitoesteroides, homeostasis lipídica, desarrollo de organismos multicelulares, proceso metabólico de los esteroides. Finalmente, los términos de GO de componente celular más representados fueron: membrana celular.

En total tres de los ocho genes listados fueron encontrados únicamente en el morfo longistilo y se encuentran relacionados con respuesta a estrés y desarrollo de la pared celular (TRINITY_DN31200_c0_g1_i1.p1, TRINITY_DN1879_c0_g1_i1.p1 y TRINITY_DN9314_c0_g1_i1.p1). En el caso del morfo brevistilo se encontraron cinco genes que no se encontraban en el morfo longistilo, los cuales estaban relacionados principalmente con la regulación de fitohormonas y desarrollo de estructuras florales (TRINITY_DN14736_c0_g1_i1.p1, TRINITY_DN7489_c0_g1_i11.p1, TRINITY_DN2959_c0_g1_i10.p1, TRINITY_DN7040_c1_g1_i2.p1 y TRINITY_DN9281_c0_g1_i1.p1) (tabla 3.9).

Tabla 3.9. Genes únicos identificados en transcriptomas de los morfos brevistilo o longistilo de *Cordia sebestena*. Los genes únicos encontrados en el morfo brevistilo se encuentran en color naranja y en color azul para el morfo Longistilo.

| Identificador | Araport11 | Viridiplantae | GO Función Molecular | GO Proceso Biológico | GO Proceso Celular |
|-----------------------------|---|--|--|--|--------------------|
| TRINITY_DN14736_c0_g1_i1.p1 | NA | BAHD acyltransferase | NA | NA | NA |
| TRINITY_DN31200_c0_g1_i1.p1 | YUCCA10/9/6 | probable indole-3-pyruvate monooxygenase YUCCA10 | Unión de lavín adenín dinucleótido Actividad de monooxygenasa | NA | NA |
| TRINITY_DN7489_c0_g1_i11.p1 | glycolate oxidase 1 | (S)-2-hydroxy-acid oxidase GLO1-like | NA | NA | NA |
| TRINITY_DN2959_c0_g1_i10.p1 | EFL3, ELF4-L3 ELF4-like 3 | protein ELF4-LIKE 4 | NA | Respuesta a estímulo externo, ritmo circadiano, regulación del ritmo circadiano | NA |
| TRINITY_DN7040_c1_g1_i2.p1 | CYP90A1, CPD, DWF3, CBB3, CYP90A, CYP90 | cytochrome P450 90A5, CYP90A1-like protein | Actividad de la monooxygenasa | proceso biosintético de compuestos orgánicos hidroxilados proceso biosintético hormonal | NA |

| | | | | | |
|----------------------------|-------------------------------------|--|----|--|------------------|
| | | | | <p>proceso biosintético de esteroides</p> <p>proceso metabólico de los fitoesteroides</p> <p>homeostasis lipídica</p> <p>desarrollo de organismos multicelulares</p> <p>proceso metabólico de los esteroides</p> | |
| TRINITY_DN9281_c0_g1_i1.p1 | indoleacetic acid-induced protein 8 | auxin-responsive protein IAA9-like | NA | NA | NA |
| TRINITY_DN1879_c0_g1_i1.p1 | homeodomain GLABROUS 2 | homeobox-leucine zipper protein HDG2 isoform X3 | NA | NA | NA |
| TRINITY_DN9314_c0_g1_i1.p1 | Microrchidia 5 | histidine kinase-, DNA gyrase B-, and HSP90-like | NA | NA | Membrana celular |

NA: No Anotado

CAPITULO IV

DISCUSIÓN

4.1 Caracterización del transcriptoma floral de *Cordia sebestena*.

Cordia sebestena es una planta distíllica que pertenece a la familia Cordiaceae. Esta familia es una de las cuatro que presentan hercogamia de aproximación, en comparación con once que presentan hercogamia reversa, en el estudio de la reconstrucción de caracteres ancestrales para el tipo de hercogamia, con métodos de máxima parsimonia (MP) y máxima verosimilitud (MV) que realizó Cohen (2019). La especie es originaria de América (Bahamas, Belice, Cuba, Haití, Honduras, Jamaica, México, Panamá, República Dominicana y Florida (E.U)). Tiene como uso principal el de ornato, la especie se encuentra presente en las guías de arbolado de los estados de Yucatán y Quintana Roo (matriz de arbolado y paleta vegetal, respectivamente). *Cordia sebestena* se propaga de manera sexual vía la polinización de las flores por su morfo floral complementario, esto quiere decir que los únicos cruces legítimos (productores de progenie fértil) son aquellos que ocurren entre los morfos longistilo y brevistilo distintos (Kohn y Barrett, 1992). *Cordia sebestena* posee importancia médica, esto tras aislar el flavonoide hesperetina presente en sus flores. La hesperetina es un compuesto antioxidante que ha demostrado efectividad contra el cáncer cervical (Prakash *et al.* 2020). La hesperetina se une fuertemente al sitio activo de la proteína E6, la proteína E6 es un factor crítico en la carcinogénesis cervical del VPH16 y es una proteína transformadora crucial para la proliferación celular (Prakash *et al.* 2020; Sohel *et al.* 2022).

A medida que aumenta el número de especies en las que se analiza la distilia, diversos autores coinciden en que se trata de un fenómeno complejo, por ello, no se debe generalizar sobre las características típicas del síndrome y su significado adaptativo (Faivre, 2000; Ferrero 2014; Kappel *et al.* 2017). Además, los recientes avances tecnológicos han permitido profundizar en la comprensión de las bases genético-moleculares de la distilia desde las primeras aproximaciones de Bateson y Gregory (1905) donde establecieron que la distilia en *Primula* estaba controlada por un único locus, denominado S, del que existe un alelo dominante S y un alelo recesivo s. Estos autores propusieron que los individuos de morfo corto son heterocigotos S/s, mientras que los individuos de morfo largo son homocigotas para el alelo recesivo s/s. Con estudios moleculares, se sabe ahora que el

control genético está relacionado con la presencia en hemicigosis del locus S en el morfo brevistilo y su ausencia en el morfo longistilo.

Para identificar las bases moleculares de la distilia se han empleado tres enfoques principales. La expresión diferencial o la abundancia de proteínas o ARNm en distintos morfos se ha utilizado para identificar factores candidatos; la mutagénesis por delección y la búsqueda de homostilos han ayudado a delimitar la localización cromosómica del locus S; la cartografía genética y la clonación posicional se han utilizado para definir la secuencia genómica del locus S (Kappel *et al.*, 2017).

En este estudio se presentan los primeros resultados de secuenciación de los transcriptomas florales de plantas de *Cordia sebestena* obtenidos en el parque Gran Santa Fe del fraccionamiento Ciudad Caucel en Mérida, Yucatán. Se logró obtener un ARN de buena calidad (RIN >4.5) el cual fue secuenciado obteniendo más de 100 Gb de secuencias de alta calidad (Phred >36) (tablas 3.2 y 3.3). Los transcriptomas florales para los morfos brevistilo y longistilo fueron ensamblados *de novo*, donde se logró ensamblar alrededor del 86% de secuencias para el morfo brevistilo y el 75% para el morfo longistilo, de acuerdo con el análisis de calidad. Así mismo el valor del contig N50 y tamaño de contig promedio del morfo brevistilo (2,678 pb y 2045 pb, respectivamente) y del morfo longistilo (2,560 pb y 1988 pb, respectivamente). No existe un tamaño definido de N50 para considerar a un ensamblado como óptimo, tomando como referencia a la planta modelo *Arabidopsis thaliana* la cual tiene transcritos con una longitud promedio de 1,557 bases, se considera que al tener un tamaño N50 $\geq 1,500$ pb es probable que se obtenga una representación razonable de las transcripciones completas en el transcriptoma ensamblado (Góngora-Castillo y Buell, 2013).

Se logró anotar más del 36% de las secuencias de aminoácidos en los transcriptomas de los morfos brevistilo y longistilo de *Cordia sebestena* y los porcentajes de genes ortólogos anotados funcionalmente fueron similares con la base de datos no redundante (NR) de NCBI (Viridiplantae) (44% y 37%), así como en el proteoma predicho de *Arabidopsis thaliana* (TAIR11) (59% y 60%, respectivamente) (Cheng *et al.*, 2017). Estos resultados confirman la buena calidad de nuestros datos, además, hubo un 60% de secuencias en ambos transcriptomas que no pudieron ser anotadas funcionalmente con ninguna base de datos, esto sugiere que podrían ser proteínas novedosas o tal vez ARN largos no codificantes (lncRNA, por sus siglas en inglés), los cuales se ha observado que pueden

tener funciones regulatorias importantes, así como jugar roles clave en procesos biológicos en diversos organismos tales como la floración, splicing alternativo, también actuar como andamios para los complejos modificadores de la cromatina, en la regulación de procesos relacionados a la fertilidad y en respuesta a estímulos externos, además de aquel material exógeno, que, por el origen de las muestras, tomadas de poblaciones silvestres, se encontraban expuestas a visitas de insectos ladrones de polen o polinizadores (Jingjing *et al.*, 2017; Bardou *et al.*, 2014; Di *et al.*, 2014; Ariel *et al.*, 2014; Kim *et al.*, 2012; Dinger *et al.*, 2009).

El análisis de enriquecimiento funcional nos permitió anotar y asignar términos al 45% del total de los transcritos en los transcriptomas de los morfos brevistilo y longistilo, así mismo más del 50% de los transcritos no pudieron ser anotados funcionalmente ni ser asignados en alguna clasificación de GO en ambos transcriptomas. A simple vista no se observaron diferencias en los términos de GO en las categorías de Función Molecular, Proceso Biológico entre los transcriptomas florales de los morfos brevistilos de *Cordia sebestena* y el proteoma predicho de *Arabidopsis thaliana* (TAIR11). Tras una revisión bibliográfica sobre los genes relacionados con distilia se lograron identificar ocho genes que potencialmente pueden ser responsables de diferencias en el desarrollo de los morfos florales.

En total, tres de los ocho genes listados fueron encontrados únicamente en el morfo longistilo y se encuentran relacionados con respuesta a estrés y desarrollo de la pared celular. En el caso del morfo brevistilo se encontraron cinco genes ausentes el morfo longistilo, los cuales estaban relacionados principalmente con la regulación de fitohormonas y desarrollo de estructuras florales. Trabajos como el de Cohen (2016) tuvieron resultados similares con un alto número de genes relacionados con respuesta a estrés, genes que codifican para proteínas reguladoras de brasinoesteroides y auxinas.

En el estudio de Henning y colaboradores (2020) realizado en la especie distílica *Turnera subulata* encontraron que los genes TsBAHD y TsYUC6 se expresan solo en el morfo brevistilo. Por el contrario de lo reportado en *Turnera*, el gen YUCCA10 homólogo a TsYUC6 fue encontrado en el morfo longistilo de *Cordia sebestena*. El gen BAHD (TRINITY_DN14736_c0_g1_i1.p1) es homólogo al gen TsBAHD que en *Turnera* se expresa en los pistilos y determina su longitud y fisiología. BAHD tiene la mayor homología con los miembros de las familias de las aciltransferasas, genes que inactivan los brasinoesteroides

por acilación (Lil et al., 2016; Wang et al., 2012). Por ello, se tiene la hipótesis de que la TsBAHD posee actividad inactivadora de los brasinoesteroides (Huu et al., 2016). El gen YUCCA10 (TRINITY_DN31200_c0_g1_i1.p1) tiene homología con TsYUC6 (Flavin-containing monooxygenase), este gen se encontró únicamente en el morfo longistilo, de acuerdo con la bibliografía este gen codifica un pequeño ligando polipeptídico secretado que interactúa con un receptor para activar una cascada de transducción de señales, lo que conduce a la muerte celular programada en el polen incompatible (Foote et al., 1994; Kakeda et al., 1998 y Ride et al., 1999).

El gen GLO1 (GLOBOSA-like protein) (TRINITY_DN7489_c0_g1_i11.p1), fue identificado en el morfo brevistilo de *Cordia sebestena*, este gen representa un fuerte candidato para el gen que controla la posición de las anteras en *Primula veris* de acuerdo con Keppel y colaboradores (2017) quienes realizaron un trabajo recopilatorio de genes relacionados con distilia para las especies *Turnera*, *Fagopyrum*, *Linum*, y *Primula*. De igual modo, en el mismo morfo se encontró el gen ELF3 (EARLY FLOWERING 3) (TRINITY_DN2959_c0_g1_i10.p1), este gen codifica una proteína relacionada con la longitud del estilo y posiblemente también el comportamiento de incompatibilidad femenina y es homólogo del gen SSG3 presente en *Fagopyrum*.

En el análisis del transcriptoma global de la especie distílica *Prímula oredoxa* realizado por Zhao y colaboradores (2019) se encontraron otros genes asociados únicamente al morfo brevistilo: CYP90A1 y IAA9 (INDOLE-3-ACETIC ACID INDUCIBLE 9). Estos mismos genes fueron identificados en *Cordia sebestena*. El gen CYP90A1 (TRINITY_DN7040_c1_g1_i2.p1) cataliza la biosíntesis del brasinoesteroides (BR). La reducción de castasterona (uno de los principales brasinoesteroides) puede causar la regulación negativa por retroalimentación de los genes implicados en la síntesis de BR (Mathur et al., 1998; Salchert et al., 1998). En el caso del gen IAA9 (TRINITY_DN9281_c0_g1_i1.p1) funciona como represor de los genes de respuesta temprana a la auxina a bajas concentraciones de auxina (Liscum y Reed, 2002).

Además, fue posible identificar los genes HDG2 (TRINITY_DN1879_c0_g1_i1.p1) y HSP90 (TRINITY_DN9314_c0_g1_i1.p1) asociados únicamente al morfo longistilo. En el caso del primero corresponde homeobox-leucine zipper protein, estas proteínas son factores de transcripción exclusivos de las plantas y participan en las respuestas al estrés abiótico y en *A. thaliana* tiene relación con el desarrollo de la pared celular (Elhiti y Stasolla, 2009;

Nakamura et al. 2006). El gen HSP90 codifica para una proteína chaperona que ayuda al correcto plegamiento de otras proteínas, además de estabilizar las proteínas frente al estrés térmico (Li et al., 2012).

Por último, podemos observar un mayor número de genes únicos en el morfo brevistilo, con la presencia de cinco de los ocho genes únicos identificados. La literatura sugiere que los genes encargados de la diferenciación de los morfos se encuentran asociados a un locus S que se encuentra únicamente en el morfo brevistilo y ausente en el longistilo (Keppel et al., 2017; Henning *et al.*, 2020; Zhao *et al.*, 2019; Cocker *et al.*, 2018). Los genes identificados potencialmente podrían ser parte del locus S o tratarse de modificadores de estos. Dado que en el morfo longistilo el número de genes únicos identificados fue menor (tres de los ocho genes) es probable, entonces, que también el control genético sea hemicigota en *Cordia sebestena*.

CAPITULO V

CONCLUSIONES GENERALES Y PERSPECTIVAS

5.1 Conclusiones

El presente estudio es el primero a nivel molecular de *Cordia sebestena*. Como se ha visto a lo largo de este trabajo, gracias a las nuevas tecnologías el número de especies distílicas que han sido estudiadas va en aumento (Cohen *et al.*, 2012; Cohen, 2016; Keppel *et al.*, 2017; Henning *et al.*, 2020; Zhao *et al.*, 2019; Cocker *et al.*, 2018). *Cordia sebestena* es la primera especie miembro de la familia Cordiaceae en el que se analiza la distilia, por lo que este trabajo de investigación representa un recurso valioso e importante en la generación de información genética de plantas con polimorfismos florales tales como la distilia, fenómeno complejo puesto que la base molecular y fisiológica sigue siendo poco conocida en muchas especies.

En este trabajo se utilizó la estrategia de RNA-Seq, recurso importante para el estudio de la información genética en especies no modelo. Las herramientas que facilitan este tipo de análisis son prácticamente inexistentes para las especies no modelo. Una forma de resolver este problema es convertir los conjuntos de datos de especies no modelo en genes de *Arabidopsis*. Al no contar con un genoma secuenciado y anotado, la secuenciación del transcriptoma nos permite obtener gran cantidad de información, de manera que es posible identificar transcritos que codifican a proteínas, isoformas, transcritos desconocidos por su baja expresión, y ayuda identificar una gran variedad de ARNs reguladores (mARN, lncRNA, sRNA, snRNA, siRNA) (Filichkin *et al.*, 2010; Morozova *et al.*, 2009;).

Se obtuvo un ensamblado de buena calidad de los transcriptomas florales de los morfos brevistilo y longistilo de plantas de *Cordia sebestena* donde se identificaron más del 44% de los transcritos ensamblados, así mismo se logró la identificación de genes que se encuentran involucrados en la diferenciación de los morfos florales de especies distílicas modelo presentes en *Cordia sebestena*, en general dichos genes se encontraban relacionados con el desarrollo de estructuras florales y la regulación de fitohormonas. Con respecto a aquellos transcritos que no pudieron ser anotados en ninguna base de datos, presentan una gran oportunidad para el desarrollo de otros estudios a través de estrategias como los genes diferencialmente expresados (DEGs), propuesta y prueba de genes a través de primers para PCR en tiempo real, la minería de datos y homología de secuencias

para poder identificar genes involucrados en la diferenciación de los morfos florales propios de *Cordia sebestena*.

Cordia sebestena es una especie con uso generalizado de ornato, podría considerarse que esta especie es víctima de una selección indirecta debido a que no se prefiere individuos de un morfo sobre otro. Esto sin contar que la falta de estudios que evalúen el reservorio genético de los parientes silvestres y al no ser una especie con importancia económica, tiene como resultado una baja generación de información molecular y una nula representación en las bases de datos. Tomando esto en cuenta y considerando la importancia médica que la especie posee, es necesario el desarrollo de estrategias que permitan caracterizar de una manera más amplia el transcriptoma de la planta y desarrollar estrategias de cultivo que puedan aprovechar mejor aquellos subproductos que la planta ofrece como frutos y flores. Esto sin contar la información que genera el estudio de los genes responsables de la floración en la comprensión del fenómeno distílico, fenómeno que sigue considerándose complejo y único en diversas especies.

Este trabajo permitió generar información genética y sentar las bases moleculares en *Cordia sebestena* a través de la caracterización del transcriptoma floral para sus morfos brevistilo y longistilo.

5.2 Perspectivas

Los resultados obtenidos en este trabajo nos permiten sentar las bases moleculares para comprender de mejor manera el fenómeno distílico en plantas silvestres de *Cordia sebestena* y el papel de los genes en la diferenciación del desarrollo floral. El presente estudio plantea las bases para comprender las bases genéticas de la distilia *Cordia sebestena*. El uso de estrategias como la expresión diferencial de genes, desarrollo de primers y PCR en tiempo real permitirán identificar genes propios de la especie responsables de las diferencias en los morfos florales de *Cordia sebestena*.

Los datos obtenidos con la estrategia RNA-Seq en este trabajo, representan información útil y valiosa para otros estudios como la minería de datos y la homología de secuencias para la identificación de genes novedosos, moléculas regulatorias y proteínas no caracterizadas.

REFERENCIAS

- Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science (New York, N.Y.)*, 309(5737), 1052–1056. <https://doi.org/10.1126/science.1115983>
- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltmann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*, 46(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>
- Aguilar-Bultet, L., & Falquet, L. (2015). Secuenciación y ensamblaje *de novo* de genomas bacterianos: una alternativa para el estudio de nuevos patógenos. *Revista de Salud Animal*, 37, 125–132.
- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology*, 37(4), 420–423. <https://doi.org/10.1038/s41587-019-0036-z>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Alvarez-Buylla, E. R., Benítez, M., Corvera-Poiré, A., Chaos Cador, A., de Folter, S., Gamboa de Buen, A., Garay-Arroyo, A., García-Ponce, B., Jaimes-Miranda, F., Pérez-Ruiz, R. V., Piñeyro-Nelson, A., & Sánchez-Corrales, Y. E. (2010). Flower development. *The arabidopsis book*, 8, e0127. <https://doi.org/10.1199/tab.0127>
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, 31(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

-
- Arellano-Rodríguez, J. A., Flores, J. S., Tun, J., & Cruz-Bojórquez, M. M. (2003). Nomenclatura, forma de vida, uso, manejo y distribución de las especies vegetales de la Península de Yucatán. *Etnoflora Yucatanense*, 20, 1–815.
- Armbruster, W., Perez-Barrales, R., Arroyo, J., Edwards, M., & Vargas, P. (2006). Three-dimensional reciprocity of floral morphs in wild flax (*Linum suffruticosum*): A new twist on heterostyly. *The New Phytologist*, 171, 581–590. <https://doi.org/10.1111/j.1469-8137.2006.01749.x>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, 25(1), 25–29. <https://doi.org/10.1038/75556>
- Athnasiou, A., Khosravi, D., Tamari, F., & Shore, J. S. (2003). Characterization and localization of short-specific polygalacturonase in distylous *Turnera subulata* (turneraceae). *American Journal of Botany*, 90(5), 675–682. <https://doi.org/10.3732/ajb.90.5.675>
- Barrett, S. C. H. (1990). The evolution and adaptive significance of heterostyly. *Trends in Ecology & Evolution*, 5(5), 144–148. [https://doi.org/https://doi.org/10.1016/0169-5347\(90\)90220-8](https://doi.org/https://doi.org/10.1016/0169-5347(90)90220-8)
- Barrett, S. C. H., Shore, J. S. (2008). New Insights on Heterostyly: Comparative Biology, Ecology and Genetics. In: Self-Incompatibility in Flowering Plants. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-68486-2_1
- Barrett, S.C.H. (2019), ‘A most complex marriage arrangement’: recent advances on heterostyly and unresolved questions. *New Phytologist*, 224: 1051-1067. <https://doi.org/10.1111/nph.16026>
- Bateson, W., y Gregory, R. P. (1905). On the inheritance of heterostyly in *Primula*. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 76(513), 581–586. <https://doi.org/10.1098/rspb.1905.0049>
- Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 57 (1). 289-300.

- Bhatia, N., & Heisler, M. G. (2018). Self-organizing periodicity in development: organ positioning in plants. *Development (Cambridge, England)*, 145(3), dev149336. <https://doi.org/10.1242/dev.149336>.
- Bhatia, N., Bozorg, B., Larsson, A., Ohno, C., Jönsson, H., & Heisler, M. G. (2016). Auxin Acts through MONOPTEROS to Regulate Plant Cell Polarity and Pattern Phyllotaxis. *Current biology: CB*, 26(23), 3202–3208. <https://doi.org/10.1016/j.cub.2016.09.044>
- Blake, J. A., & Harris, M. A. (2008). The Gene Ontology (GO) project: structured vocabularies for molecular biology and their application to genome and expression analysis. *Current protocols in bioinformatics, Chapter 7*, 7.2.1–7.2.9. <https://doi.org/10.1002/0471250953.bi0702s23>
- Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., & Galaxy Team (2010). Manipulation of FASTQ data with Galaxy. *Bioinformatics (Oxford, England)*, 26(14), 1783–1785. <https://doi.org/10.1093/bioinformatics/btq281>
- Borhidi, A., Gondár, E., & Orosz-Kovács, Z. (1988). The reconsideration of the genus *Cordia*. *Acta Botanica Hungarica*, 34(3–4), 375–423.
- Bowman, J. L., Smyth, D. R., & Meyerowitz, E. M. (2012). The ABC model of flower development: then and now. *Development (Cambridge, England)*, 139(22), 4095–4098. <https://doi.org/10.1242/dev.083972>
- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Erratum: Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology*, 34(8), 888. <https://doi.org/10.1038/nbt0816-888d>
- Bull-Hereñu, K., Ronse De Craene, L., & Pérez, F. (2016). Flower meristematic size correlates with heterostylous morphs in two Chilean *Oxalis* (Oxalidaceae) species. *Flora*, 221, 14–21. <https://doi.org/https://doi.org/10.1016/j.flora.2016.02.009>
- Bushmanova, E., Antipov, D., Lapidus, A., Suvorov, V., & Prjibelski, A. D. (2016). maQUAST: a quality assessment tool for de novo transcriptome assemblies. *Bioinformatics (Oxford, England)*, 32(14), 2210–2212. <https://doi.org/10.1093/bioinformatics/btw218>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics*, 10, 421. <https://doi.org/10.1186/1471-2105-10-421>

- Canché-Collí, C., & Canto, A. (2014). Distylous traits in *Cordia dodecandra* and *Cordia sebestena* (boraginaceae) from the Yucatan peninsula. *Boletín De La Sociedad Botánica De Mexico*, 92, 289-297.
- Carpenter, R., Copsey, L., Vincent, C., Doyle, S., Magrath, R., & Coen, E. (1995). Control of flower development and phyllotaxy by meristem identity genes in *Antirrhinum*. *The Plant cell*, 7(12), 2001–2011. <https://doi.org/10.1105/tpc.7.12.2001>
- Chanderbali, A. S., Berger, B. A., Howarth, D. G., Soltis, P. S., & Soltis, D. E. (2016). Evolving Ideas on the Origin and Evolution of Flowers: New Perspectives in the Genomic Era. *Genetics*, 202(4), 1255–1265. <https://doi.org/10.1534/genetics.115.182964>
- Charlesworth D. (2015). The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evolutionary applications*, 9(1), 74–90. <https://doi.org/10.1111/eva.12291>
- Chiner-Oms, Á., Berney, M., Boinett, C., Gonzáles-Candela, F., Young, D., Gagneaux, S., Jacobs, W., Parkhill, J., Cortes, T., & Comas, I. (2019). Genome-wide mutational biases fuel transcriptional diversity in the *Mycobacterium tuberculosis* complex. *Nat Commun* 10, 3994 (2019). <https://doi.org/10.1038/s41467-019-11948-6>
- Chu, Y., & Corey, D. R. (2012). RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic acid therapeutics*, 22(4), 271–274. <https://doi.org/10.1089/nat.2012.0367>
- Church S. A. (2003). Molecular phylogenetics of *Houstonia* (Rubiaceae): descending aneuploidy and breeding system evolution in the radiation of the lineage across North America. *Molecular phylogenetics and evolution*, 27(2), 223–238. [https://doi.org/10.1016/s1055-7903\(02\)00446-3](https://doi.org/10.1016/s1055-7903(02)00446-3)
- Coen, E., Meyerowitz, E. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353:31–37. <https://doi.org/10.1038/353031a0>
- Cohen J. I. (2010). "A case to which no parallel exists": The influence of Darwin's Different Forms of Flowers. *American journal of botany*, 97(5), 701–716. <https://doi.org/10.3732/ajb.0900395>

- Cohen J. I. (2014). A phylogenetic analysis of morphological and molecular characters of Boraginaceae: evolutionary relationships, taxonomy, and patterns of character evolution. *Cladistics : the international journal of the Willi Hennig Society*, 30(2), 139–169. <https://doi.org/10.1111/cla.12036>
- Cohen J. I. (2016). *De novo* Sequencing and Comparative Transcriptomics of Floral Development of the Distylous Species *Lithospermum multiflorum*. *Frontiers in plant science*, 7, 1934. <https://doi.org/10.3389/fpls.2016.01934>
- Cohen J. I. (2019). How to build distylous flowers: comparative floral development and evolution of distylous species across the angiosperms. *American journal of botany*, 106(10), 1285–1299. <https://doi.org/10.1002/ajb2.1363>
- Cohen, J. I., Litt, A., & Davis, J. I. (2012). Comparative floral development in *Lithospermum* (Boraginaceae) and implications for the evolution and development of heterostyly. *American journal of botany*, 99(5), 797–805. <https://doi.org/10.3732/ajb.1100329>
- Davidson, J., Durham, S., & Wolf, P. (2014). Breeding system of the threatened endemic *Primula cusickiana* var. *maguirei* (Primulaceae). *Western North American Naturalist*, 71(3), 327–337. <http://www.jstor.org/stable/41718276>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Durán R. (2000). Listado florístico de la península de Yucatán. Centro de Investigación Científica de Yucatán.
- Eddy, S. (2003). The HMMER User's Guide (biological sequence analysis using profile hidden Markov models). Howard Hughes Medical Institute and Dept. of Genetics Washington University School of Medicine, 660 South Euclid Avenue, Box 8232 Saint Louis, Missouri 63110, USA, version 2.3.2 edition. <http://hmmer.wustl.edu/>.
- Elhiti, M., & Stasolla, C. (2009). Structure and function of homeodomain-leucine zipper (HD-Zip) proteins. *Plant Signaling & Behaviour*, 4(2), 86-88. <https://doi.org/10.4161/psb.4.2.7692>

- Estrada, M. E. (2000). Caracterización y cartografía del recurso suelo del Municipio de Hocabá, Yucatán. Tesis de Maestría en Ciencias sobre Manejo y Conservación de Recursos Naturales. Facultad de Medicina Veterinaria y Zootecnia de la Universidad Autónoma de Yucatán. Mérida, Yucatán, México.
- Faegri, Knut. & Pijl, L. Van Der. (1979). The principles of pollination ecology. *Oxford; New York: Pergamon Press*
- Faivre A. E. (2000). Ontogenetic differences in heterostylous plants and implications for development from a herkogamous ancestor. *Evolution; international journal of organic evolution*, 54(3), 847–858. <https://doi.org/10.1111/j.0014-3820.2000.tb00085.x>
- Fang, Z., Martin, J., & Wang, Z. (2012). Statistical methods for identifying differentially expressed genes in RNA-Seq experiments. *Cell & bioscience*, 2(1), 26. <https://doi.org/10.1186/2045-3701-2-26>
- Ferrero, V., Rojas, D., Vale, A., & Navarro, L. (2012). Delving into the loss of heterostyly in Rubiaceae: Is there a similar trend in tropical and non-tropical climate zones? Perspectives in Plant Ecology, *Evolution and Systematics*, 14(3), 161–167. <https://doi.org/https://doi.org/10.1016/j.ppees.2011.11.005>
- Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., Wong, W. K., & Mockler, T. C. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome research*, 20(1), 45–58. <https://doi.org/10.1101/gr.093302.109>
- Foote, H. C., Ride, J. P., Franklin-Tong, V. E., Walker, E. A., Lawrence, M. J., & Franklin, F. C. (1994). Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas* L. *Proceedings of the National Academy of Sciences of the United States of America*, 91(6), 2265–2269. <https://doi.org/10.1073/pnas.91.6.2265>
- Fosket, Donald E. (1994). Plant Growth and Development: A Molecular Approach. San Diego: Academic Press. pp. 498-509. ISBN 0-12-262430-0.
- Franch Sarto, M. (2019). Bioinformática funcional y su aplicación en genómica, proteogenómica y reposicionamiento de fármacos. Tesis de la Universidad Complutense de Madrid, Facultad de Farmacia, Departamento de Microbiología y Parasitología
- Ganders, F. R. (1979). The biology of heterostyly. *New Zealand Journal of Botany*, 17(4), 607–635. <https://doi.org/10.1080/0028825X.1979.10432574>

- Góngora-Castillo, E., & Buell, C. R. (2013). Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Natural product reports*, 30(4), 490–500. <https://doi.org/10.1039/c3np20099j>
- Góngora-Castillo, E., Nic-Can, G. I., Galaz-Ávalos, R. M., & Loyola-Vargas, V. M. (2018). Elaboration of Transcriptome During the Induction of Somatic Embryogenesis. *Methods in molecular biology (Clifton, N.J.)*, 1815, 411–427. https://doi.org/10.1007/978-1-4939-8594-4_29
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351. <https://doi.org/10.1038/nrg.2016.49>
- Gottschling, M., Hilger, H. H., Wolf, M., & Diane, N. (2001). Secondary Structure of the ITS1 Transcript and its Application in a Reconstruction of the Phylogeny of Boraginales. *Plant Biology*, 3(6), 629–636. <https://doi.org/https://doi.org/10.1055/s-2001-19371>
- Gottschling, M., Miller, J. S. (2007). Generic Classification in the Cordiaceae (Boraginales): Resurrection of the Genus *Varronia* P. Br. *International Association for Plant Taxonomy (IAPT)*. Vol. 56, No. 1 (Feb., 2007), pp. 163-169
- Gottschling, M., Miller, J. S., Weigend, M., & Hilger, H. H. (2005). Congruence of a Phylogeny of Cordiaceae (Boraginales) Inferred from ITS1 Sequence Data with Morphology, Ecology, and Biogeography. *Annals of the Missouri Botanical Garden*, 92(3), 425–437. <http://www.jstor.org/stable/40035480>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- Griffith, M., Griffith, O. L., Mwenifumbo, J., Goya, R., Morrissy, A. S., Morin, R. D., Corbett, R., Tang, M. J., Hou, Y. C., Pugh, T. J., Robertson, G., Chittaranjan, S., Ally, A., Asano, J. K., Chan, S. Y., Li, H. I., McDonald, H., Teague, K., Zhao, Y., Zeng, T., ... Marra, M. A. (2010). Alternative expression analysis by RNA sequencing. *Nature methods*, 7(10), 843–847. <https://doi.org/10.1038/nmeth.1503>

- Haas, B. (2019). Trinotate. <https://github.com/Trinotate/Trinotate>. [Acceso marzo 15 2022].
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, 8(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Hamès, C., Ptchelkine, D., Grimm, C., Thevenon, E., Moyroud, E., Gérard, F., Martiel, J. L., Benloch, R., Parcy, F., & Müller, C. W. (2008). Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *The EMBO journal*, 27(19), 2628–2637. <https://doi.org/10.1038/emboj.2008.184>
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R; Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic acids research*, 32(Database issue), D258–D261. <https://doi.org/10.1093/nar/gkh036>
- Henning, P. M., Shore, J. S., & McCubbin, A. G. (2020). Transcriptome and Network Analyses of Heterostyly in *Turnera subulata* Provide Mechanistic Insights: Are S-Loci a Red-Light for Pistil Elongation?. *Plants (Basel, Switzerland)*, 9(6), 713. <https://doi.org/10.3390/plants9060713>
- Herrera, C. M. (1996). Floral Traits and Plant Adaptation to Insect Pollinators: A Devil's Advocate Approach. In D. G. Lloyd y S. C. H. Barrett (Eds.), *Floral Biology: Studies on Floral Evolution in Animal-Pollinated Plants* (pp. 65–87). Springer US. https://doi.org/10.1007/978-1-4613-1165-2_3
- Heuch, I., & Lie, R. T. (1985). Genotype frequencies associated with incompatibility systems in tristylous plants. *Theoretical Population Biology*, 27, 318–336.

- Huu, C. N., Kappel, C., Keller, B., Sicard, A., Takebayashi, Y., Breuninger, H., Nowak, M. D., Baurle, I., Himmelbach, A., Burkart, M., Ebbing-Lohaus, T., Sakakibara, H., Altschmied, L., Conti, E., & Lenhard, M. (2016). Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. *eLife*, 5, e17956. <https://doi.org/10.7554/eLife.17956>
- Johnston, I. M. (1948). Studies in the Boraginaceae, XVII. *Journal of the RNAold Arboretum*, 30(1), 85–110. <https://doi.org/10.5962/bhl.part.18047>
- Kakeda, K., Jordan, N. D., Conner, A., Ride, J. P., Franklin-Tong, V. E., & Franklin, F. C. (1998). Identification of residues in a hydrophilic loop of the *Papaver rhoeas* S protein that play a crucial role in recognition of incompatible pollen. *The Plant cell*, 10(10), 1723–1732. <https://doi.org/10.1105/tpc.10.10.1723>
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., & Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic acids research*, 32(Database issue), D277–D280. <https://doi.org/10.1093/nar/gkh063>
- Kappel, C., Huu, C. N., & Lenhard, M. (2017). A short story gets longer: recent insights into the molecular basis of heterostyly. *Journal of experimental botany*, 68(21-22), 5719–5730. <https://doi.org/10.1093/jxb/erx387>
- Kerner, A., Busk, M., Macdonald, M. F., Mary, F., & Oliver, F. W. (1895). The natural history of plants, their forms, growth, reproduction, and distribution: from the German of Anton Kerner von Marilaun. (Vol. 1). H. Holt. <https://www.biodiversitylibrary.org/item/57617>
- Khosravi, D., Yang, E. C. C., Siu, K. W. M., & Shore, J. S. (2004). High Level of α -Dioxygenase in Short Styles of Distylous *Turnera* Species. *International Journal of Plant Sciences*, 165(6), 995–1006. <https://doi.org/10.1086/423871>
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, 14(4), R36. <https://doi.org/10.1186/gb-2013-14-4-r36>
- Kohn, J. R., & Barrett, S. C. H. (1992). Experimental studies on the functional significance of heterostyly. *Evolution; international journal of organic evolution*, 46(1), 43–55. <https://doi.org/10.1111/j.1558-5646.1992.tb01983.x>

-
- Kramer, E. M. (2019). Plus ça change, plus c'est la même chose: The developmental evolution of flowers. *Current topics in developmental biology*, 131, 211–238. <https://doi.org/10.1016/bs.ctdb.2018.11.015>
- Kurian, V., & Richards, A. J. (1997). A new recombinant in the heteromorphy 'S' supergene in *Primula*. *Heredity*, 78(4), 383–390. <https://doi.org/10.1038/hdy.1997.61>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Li, B. & Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 12(1), 323. <https://doi.org/10.1186/1471-2105-12-323>
- Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A., & Dewey, C. N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics (Oxford, England)*, 26(4), 493–500. <https://doi.org/10.1093/bioinformatics/btp692>
- Li, J., Cocker, J. M., Wright, J., Webster, M. A., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., Oosterhout, C. V., & Gilmartin, P. M. (2016). Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. *Nature plants*, 2(12), 16188. <https://doi.org/10.1038/nplants.2016.188>
- Li, J., Soroka, J., & Buchner, J. (2012). The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones. *Biochimica et biophysica acta*, 1823(3), 624–635. <https://doi.org/10.1016/j.bbamcr.2011.09.003>
- Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., & Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics (Oxford, England)*, 25(15), 1966–1967. <https://doi.org/10.1093/bioinformatics/btp336>
- Liscum, E., & Reed, J. W. (2002). Genetics of Aux/IAA and ARF action in plant growth and development. *Plant molecular biology*, 49(3-4), 387–400.
- Lister, R., Gregory, B. D., & Ecker, J. R. (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Current opinion in plant biology*, 12(2), 107–118. <https://doi.org/10.1016/j.pbi.2008.11.004>

- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS computational biology*, 13(5), e1005457. <https://doi.org/10.1371/journal.pcbi.1005457>
- Marguerat, S., & Bähler, J. (2010). RNA-seq: from technology to biology. *Cellular and molecular life sciences : CMLS*, 67(4), 569–579. <https://doi.org/10.1007/s00018-009-0180-6>
- Mathur, J., Molnár, G., Fujioka, S., Takatsuto, S., Sakurai, A., Yokota, T., Adam, G., Voigt, B., Nagy, F., Maas, C., Schell, J., Koncz, C., & Szekeres, M. (1998). Transcription of the Arabidopsis CPD gene, encoding a steroidogenic cytochrome P450, is negatively controlled by brassinosteroids. *The Plant journal : for cell and molecular biology*, 14(5), 593–602. <https://doi.org/10.1046/j.1365-313x.1998.00158.x>
- McClure B. (2016). Reproduction: The genetic basis of heterostyly. *Nature plants*, 2(12), 16184. <https://doi.org/10.1038/nplants.2016.184>
- McDill, Joshua & Replinger, Miriam & Simpson, Beryl & Kadereit, Joachim. (2009). The Phylogeny of *Linum* and Linaceae Subfamily Linoideae, with Implications for Their Systematics, Biogeography, and Evolution of Heterostyly. *Systematic Botany*. 34. 386-405. 10.1600/036364409788606244.
- Mendoza-González, P., & Encina-García, A. E. (2018). Bases moleculares de la floración. *Ambiociencias*, 31–42. <https://doi.org/10.18002/ambioc.v0i0.5557>
- Mi, H., & Thomas, P. (2009). PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods in molecular biology (Clifton, N.J.)*, 563, 123–140. https://doi.org/10.1007/978-1-60761-175-2_7
- Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature protocols*, 8(8), 1551–1566. <https://doi.org/10.1038/nprot.2013.092>
- Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic acids research*, 47(D1), D419–D426. <https://doi.org/10.1093/nar/gky1038>

- Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., & Thomas, P. D. (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature protocols*, *14*(3), 703–721. <https://doi.org/10.1038/s41596-019-0128-8>
- Michaels, S. D., & Amasino, R. M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *The Plant cell*, *11*(5), 949–956. <https://doi.org/10.1105/tpc.11.5.949>
- Miller, J. S. (1985). Systematics of the genus *Cordia* (Boraginaceae) in Mexico and Central America. Saint Louis University. 1372 pp.
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic acids research*, *49*(D1), D412–D419. <https://doi.org/10.1093/nar/gkaa913>
- Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S. Y., Bateman, A., Punta, M., Attwood, T. K., Sigrist, C. J., Redaschi, N., Rivoire, C., ... Finn, R. D. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic acids research*, *43*(Database issue), D213–D221. <https://doi.org/10.1093/nar/gku1243>
- Morozova, O., Hirst, M., & Marra, M. A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annual review of genomics and human genetics*, *10*, 135–151. <https://doi.org/10.1146/annurev-genom-082908-145957>
- Müller, H. (1876). Fertilisation of Flowers by Insects. *Nature*, *14*(347), 173–175. <https://doi.org/10.1038/014173b0>
- Nakamura, M., Katsumata, H., Abe, M., Yabe, N., Komeda, Y., Yamamoto, K. T., & Takahashi, T. (2006). Characterization of the class IV homeodomain-Leucine Zipper gene family in *Arabidopsis*. *Plant physiology*, *141*(4), 1363–1375. <https://doi.org/10.1104/pp.106.077388>

- O'Leary NA, Wright MW, Brister JR, Ciufu S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*. 44(D1), D733–D745. <https://doi.org/10.1093/nar/gkv1189>
- Opler, P. A., Baker, H. G., & Frankie, G. W. (1975). Reproductive Biology of Some Costa Rican *Cordia* Species (Boraginaceae). *Biotropica*. 7(4), 234–247. <https://doi.org/10.2307/2989736>
- Orellana, R., Espadas-Manrique, C., & Islebe, G. (2003). Presente, pasado y futuro de los climas de la Península de Yucatán. *Naturaleza y sociedad en el área maya*. pp. 37–52
- Ovalle, R. T. (2018). Análisis transcriptómico asociado a la producción de β -caroteno en Yuca. Tesis para obtención del grado Master en Bioinformática y Bioestadística. Universitat Oberta de Catalunya. Cataluña, España.
- Percival, M. (1974). Floral Ecology of Coastal Scrub in Southeast Jamaica. *Biotropica*, 6(2), 104–129. <https://doi.org/10.2307/2989824>
- Pinar, S. (2009). Introducción al libro Las distintas formas de las flores en plantas de la misma especie de Charles Darwin. Biblioteca Darwiniana. Academia Mexicana de Ciencias. 288 pp.
- Prakash, Shanmugam & Nagaraj, Elavarasan & Subashini, Kasivisvanathan & Kanaga, Selvaraj & Dhandapani, Ramamurthy & Magudeeswaran, Sivanandam & Kumaradhas, Poomani & Thirunavukkarasu, Dr. Chinnasamy & Sujatha, Venugopal. (2020). Isolation of hesperetin - A flavonoid from *Cordia sebestena* flower extract through antioxidant assay guided method and its antibacterial, anticancer effect on cervical cancer via *in vitro* and *in silico* molecular docking studies. *Journal of Molecular Structure*. 1207. 127751. [10.1016/j.molstruc.2020.127751](https://doi.org/10.1016/j.molstruc.2020.127751).

-
- R Core Team. (2016). R: A Language and Environment for Statistical Computing, Journal. R Foundation for Statistical Computing, Vienna, Austria.
- Richards, J. H., & Koptur, S. (1993). Floral Variation and Distyly in *Guettarda scabra* (Rubiaceae). *American Journal of Botany*, 80(1), 31–40. <https://doi.org/10.2307/2445117>
- Richards, J.H., Barrett, S.C.H. (1992). The Development of Heterostyly. In: Barrett, S.C.H. (eds) Evolution and Function of Heterostyly. *Monographs on Theoretical and Applied Genetics*, vol 15. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-86656-2_4
- Rico-Gray, V., Chemás, A., & Mandujano, S. (1991). Uses of tropical deciduous forest species by the Yucatecan Maya. *Agroforestry Systems*, 14(2), 149–161. <https://doi.org/10.1007/BF00045730>
- Ride, J. P., Davies, E. M., Franklin, F. C., & Marshall, D. F. (1999). Analysis of *Arabidopsis* genome sequence reveals a large new gene family in plants. *Plant molecular biology*, 39(5), 927–932. <https://doi.org/10.1023/a:1006178511787>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rodriguez, A., Perlaza-Jiménez, L., & Bernal, A. (2014). RNA-Seq Data Analysis in Prokaryotes: A Review for Non-experts. *Acta Biologica Colombiana*. 19, 131–142. <https://doi.org/10.15446/abc.v19n2.41010>
- Rollano Peñaloza, Oscar M, & Mollinedo Portugal, Patricia. (2017). Análisis bioinformático de Arn-Seq con una perspectiva para Bolivia. *Revista Boliviana de Química*, 34(2), 50-55.
- Romero-Arias, J. R., Barrio, R. Á. Álvarez-Buylla R., Elena, V., C., & Hernández-Machado, A. (2010). Modelo de diferenciación celular en la floración de *Arabidopsis thaliana*. *Tip Revista especializada en ciencias químico-biológicas*. 13(2), 91-102.
- Ruiz-Martín, J. (2011). Algo que fascinó a Darwin: La evolución del polimorfismo floral en el género *Linum* (Linaceae). In *Chronica naturae*. 1, 46-64.

- Salchert, K., Bhalerao, R., Koncz-Kálmán, Z., & Koncz, C. (1998). Control of cell elongation and stress responses by steroid hormones and carbon catabolic repression in plants. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 353(1374), 1517–1520. <https://doi.org/10.1098/rstb.1998.0307>
- Sampson, D. A., & Krebs, R. A. (2013). Developmental variation and the evolution of distyly in *Hedyotis caerulea* (Rubiaceae). *SpringerPlus*, 2, 383. <https://doi.org/10.1186/2193-1801-2-383>
- Sayou, C., Nanao, M. H., Jamin, M., Posé, D., Thévenon, E., Grégoire, L., Tichtinsky, G., Denay, G., Ott, F., Peirats Llobet, M., Schmid, M., Dumas, R., & Parcy, F. (2016). A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. *Nature communications*, 7, 11222. <https://doi.org/10.1038/ncomms11222>
- Schoch, C. L., Ciufo, S., Domrachev, M., Hottton, C. L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., Sharma, S., Soussov, V., Sullivan, J. P., Sun, L., Turner, S., & Karsch-Mizrachi, I. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database : the journal of biological databases and curation*, 2020, baaa062. <https://doi.org/10.1093/database/baaa062>
- Schoen, D. J., Johnston, M. O., L'Heureux, A. M., & Marsolais, J. V. (1997). Evolutionary history of the mating system in *Amsinckia* (Boraginaceae). *Evolution; international journal of organic evolution*, 51(4), 1090–1099. <https://doi.org/10.1111/j.1558-5646.1997.tb03956.x>
- Schopfer, C. R., Nasrallah, M. E., & Nasrallah, J. B. (1999). The male determinant of self-incompatibility in *Brassica*. *Science (New York, N.Y.)*, 286(5445), 1697–1700. <https://doi.org/10.1126/science.286.5445.1697>
- Shore, J. S., & Barrett, S. C. H. (1985). The genetics of distyly and homostyly in *Turnera ulmifolia* L. (Turneraceae). *Heredity*, 55, 167–174. <https://doi.org/10.1038/hdy.1985.88>
- Silberman, A., León, P., & Orellana, R. (2000). El Atlas de procesos territoriales de Yucatán. *Investigaciones Geográficas*, 171–174.
- Simonini, S., & Østergaard, L. (2019). Female reproductive organ formation: A multitasking endeavor. *Current topics in developmental biology*, 131, 337–371. <https://doi.org/10.1016/bs.ctdb.2018.10.004>

- Simpson, B. B., & Neff, J. L. (1983). Evolution of diversity of floral rewards. In Jones, C. E. & Little, R. J. (Eds.). *Handbook of experimental pollination biology*, pp. 142-159. Van Nostrand Reinhold Co., Inc., New York.
- Sims, D., Sudbery, I., Illott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature reviews. Genetics*, 15(2), 121–132. <https://doi.org/10.1038/nrg3642>
- Sohel, M., Sultana, H., Sultana, T., Al Amin, M., Aktar, S., Ali, M. C., Rahim, Z. B., Hossain, M. A., Al Mamun, A., Amin, M. N., & Dash, R. (2022). Chemotherapeutic potential of hesperetin for cancer treatment, with mechanistic insights: A comprehensive review. *Heliyon*, 8(1), e08815. <https://doi.org/10.1016/j.heliyon.2022.e08815>
- Taroda, N. (1984). Taxonomic studies on brazilian species of *Cordia* L. (Boraginaceae). Tesis para obtener el grado de Doctorado. Universidad de St. Andrews
- Theissen, G., & Saedler, H. (2001). Plant biology. Floral quartets. *Nature*, 409(6819), 469–471. <https://doi.org/10.1038/35054172>
- Thomson, B., & Wellmer, F. (2019). Molecular regulation of flower development. *Current topics in developmental biology*, 131, 185–210. <https://doi.org/10.1016/bs.ctdb.2018.11.007>
- Tippery, N. P., Les, D. H., Padgett, D. J., & Jacobs, S. W. L. (2008). Generic Circumscription in Menyanthaceae: A Phylogenetic Evaluation. *Systematic Botany*, 33(3), 598–612.
- Tomlinson, P. B. (1974). Breeding mechanisms in trees native to tropical florida — A morphological assessment. *Journal of the Arnold Arboretum*, 55(2), 269–290.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, 7(3), 562–578. <https://doi.org/10.1038/nprot.2012.016>
- Truyens, S., Arbo M. M. & Shore, J. S. (2005) Phylogenetic relationships, chromosome and breeding system evolution in *Turnera* (Turneraceae): inferences from ITS sequence data. *American Journal of Botany*. 92,1749–1758.
- Uhlén, M., Hallström, B. M., Lindskog, C., Mardinoglu, A., Pontén, F., & Nielsen, J. (2016). Transcriptomics resources of human tissues and organs. *Molecular systems biology*, 12(4), 862. <https://doi.org/10.15252/msb.20155865>

- Ushijima, K., Nakano, R., Bando, M., Shigezane, Y., Ikeda, K., Namba, Y., Kume, S., Kitabata, T., Mori, H., & Kubo, Y. (2012). Isolation of the floral morph-related genes in heterostylous flax (*Linum grandiflorum*): The genetic polymorphism and the transcriptional and post-transcriptional regulations of the S locus. *Plant Journal*, 69(2), 317–331. <https://doi.org/10.1111/j.1365-313X.2011.04792.x>
- Van Verk, M. C., Hickman, R., Pieterse, C. M., & Van Wees, S. C. (2013). RNA-Seq: revelation of the messengers. *Trends in plant science*, 18(4), 175–179. <https://doi.org/10.1016/j.tplants.2013.02.001>
- Wang, M., Liu, X., Wang, R., Li, W., Rodermeil, S., & Yu, F. (2012). Overexpression of a putative *Arabidopsis* BAHD acyltransferase causes dwarfism that can be rescued by brassinosteroid. *Journal of experimental botany*, 63(16), 5787–5801. <https://doi.org/10.1093/jxb/ers227>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1), 57–63. <https://doi.org/10.1038/nrg2484>
- Ward, J. A., Ponnala, L., & Weber, C. A. (2012). Strategies for transcriptome analysis in nonmodel plants. *American journal of botany*, 99(2), 267–276. <https://doi.org/10.3732/ajb.1100334>
- Webb, C. J., & Lloyd, D. G. (1986). The avoidance of interference between the presentation of pollen and stigmas in angiosperms II. Herkogamy. *New Zealand Journal of Botany*. 24(1), 163–178. <https://doi.org/10.1080/0028825X.1986.10409726>
- Weigel, D., & Meyerowitz, E. M. (1994). The ABCs of floral homeotic genes. *Cell*, 78(2), 203–209. [https://doi.org/10.1016/0092-8674\(94\)90291-7](https://doi.org/10.1016/0092-8674(94)90291-7)
- Wigge, P. A., Kim, M. C., Jaeger, K. E., Busch, W., Schmid, M., Lohmann, J. U., & Weigel, D. (2005). Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science (New York, N.Y.)*, 309(5737), 1056–1059. <https://doi.org/10.1126/science.1114358>
- Wu, M. F., Yamaguchi, N., Xiao, J., Bargmann, B., Estelle, M., Sang, Y., & Wagner, D. (2015). Auxin-regulated chromatin switch directs acquisition of flower primordium founder fate. *eLife*, 4, e09269. <https://doi.org/10.7554/eLife.09269>

-
- Wu, T. D., & Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics (Oxford, England)*, *26*(7), 873–881. <https://doi.org/10.1093/bioinformatics/btq057>
- Yasui, Y., Hirakawa, H., Ueno, M., Matsui, K., Katsube-Tanaka, T., Yang, S. J., Aii, J., Sato, S., & Mori, M. (2016). Assembly of the draft genome of buckwheat and its applications in identifying agronomically useful genes. *DNA research: an international journal for rapid publication of reports on genes and genomes*, *23*(3), 215–224. <https://doi.org/10.1093/dnares/dsw012>
- Yasui, Y., Mori, M., Aii, J., Abe, T., Matsumoto, D., Sato, S., Hayashi, Y., Ohnishi, O., & Ota, T. (2012). S-LOCUS EARLY FLOWERING 3 is exclusively present in the genomes of short-styled buckwheat plants that exhibit heteromorphic self-incompatibility. *PLoS one*, *7*(2), e31264. <https://doi.org/10.1371/journal.pone.0031264>
- Zdobnov, E. M., & Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics (Oxford, England)*, *17*(9), 847–848. <https://doi.org/10.1093/bioinformatics/17.9.847>
- Zhao, Z., Luo, Z., Yuan, S., Mei, L., & Zhang, D. (2019). Global transcriptome and gene co-expression network analyses on the development of distyly in *Primula oreodoxa*. *Heredity*, *123*(6), 784–794. <https://doi.org/10.1038/s41437-019-0250-y>
- Zik, M., & Irish, V. F. (2003). Flower development: initiation, differentiation, and diversification. *Annual review of cell and developmental biology*, *19*, 119–140. <https://doi.org/10.1146/annurev.cellbio.19.111301.134635>

ANEXOS.

ANEXO I. RNeasy® Plant Mini Kit QIAGEN

Guía de inicio rápido

El kit RNeasy Plant Mini (números de catálogo 74903 y 74904) puede ser almacenado a temperatura ambiente (15–25°C) hasta por 9 meses a menos que se diga lo contrario en el etiquetado.

Notas antes de empezar

- El kit RNeasy Plant Mini proporciona opciones de buffers de lisis. El buffer RLT suele ser el buffer de lisis de elección, sin embargo, el buffer RLT puede causar la solidificación de algunas muestras, dependiendo de la cantidad y el tipo de metabolitos secundarios presentes en el tejido. En estos casos, se debería usar el buffer RLC.
- Agregar 10 µl β-mercaptoethanol (β-ME) o 20 µl 2 M dithiothreitol (DTT), a 1 ml de Buffer RLT o Buffer RLC antes de usarse. Buffers con DTT ó β-ME pueden ser almacenadas a temperatura ambiente hasta por 1 mes.
- Agregar 4 volúmenes de etanol (96–100%) al buffer RPE para la solución de trabajo.

1) Moler un máximo de 100 mg de tejido vegetal de acuerdo con el paso 1a o 1b.

1a. Macerado con mortero y pistilo

Cubrir la muestra de tejido de manera inmediata en nitrógeno líquido. Triturar concienzudamente. Decantar el polvo resultante en un tubo de microcentrifuga (no suministrado) libre de RNAsas, enfriado en nitrógeno de 2 ml. Decant tissue powder and liquid nitrogen into RNase-free, liquid-nitrogen-cooled, 2 ml. Permitir la evaporación del nitrógeno líquido, pero que la muestra no se descongele. Proceder inmediatamente al paso 2.

1b. Macerado usando el TissueLyser II, TissueLyser LT, o TissueRuptor®

Para información detallada en el macerado de tejido vegetal para la purificación de RNA, consultar el manual de TissueLyser, TissueLyser LT o TissueRuptor según corresponda.

- 2)** Agregar 450 µl de buffer RLT o Buffer RLC a un máximo de 100 mg de polvo del tejido. Agitar por Vortex vigorosamente.
- 3)** Transferir el lisado a una columna spin (lila) montada en un tubo de recolección de 2 ml. Centrifugar por 2 min a máxima velocidad. Transferir el sobrenadante del flujo a un nuevo tubo de microcentrifuga (no suministrado) sin perturbar el pellet.
- 4)** Agregar 0.5 volúmenes de etanol (96–100%) al lisado clarificado y mezclar inmediatamente por pipeteo. No centrifugar. Proceder inmediatamente al paso 5.

- 5) Transferir la muestra (usualmente 650 μ l), con cualquier precipitado, a una columna spin RNeasy Mini (rosada) montada en un tubo de recolección de 2 ml (suministrado). Cerrar la tapa y centrifugar por 15 s a $\geq 8000 \times g$ ($\geq 10,000$ rpm). Desechar el flujo.
- 6) Agregar 700 μ l de buffer RW1 a la columna spin RNeasy. Cerrar la tapa y centrifugar por 15 s a $\geq 8000 \times g$. Desechar el flujo.
- 7) Agregar 500 μ l de buffer RPE a la columna spin RNeasy. Cerrar la tapa y centrifugar por 15 s a $\geq 8000 \times g$. Desechar el flujo.
- 8) Agregar 500 μ l de buffer RPE a la columna spin RNeasy. Cerrar la tapa y centrifugar por 2 min a $\geq 8000 \times g$.

Opcional: Colocar la columna spin RNeasy en un nuevo tubo de recolección (suministrado). Centrifugar a máxima velocidad por 1 min para secar la membrana.

- 9) Colocar la columna spin RNeasy en un tubo nuevo de 1.5 ml (suministrado). Agregar 30–50 μ l de agua libre de RNAsas directamente a la membrana de la columna spin. Cerrar la tapa y centrifugar por 1 min a $\geq 8000 \times g$ para eluir el RNA.
- 10) Si el rendimiento esperado del RNA es $>30 \mu$ g, repetir el paso 9 usando otra cantidad similar de agua libre de RNAsas (30–50 μ l). Alternativamente, use el eluido del paso 9 (si una alta concentración de RNA es requerida). Rehúse el tubo de colecta del paso 9.