



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO EN CIENCIAS FORESTALES

ANÁLISIS TRANSCRIPTÓMICO DE LA GERMINACIÓN EN SEMILLAS DE *Cedrela odorata* L. PARA SU CONSERVACIÓN

LORENA GÁLVEZ LÓPEZ

T E S I S
PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:

DOCTORA EN CIENCIAS

MONTECILLO, TEXCOCO, ESTADO DE MÉXICO, MÉXICO

2022



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

La presente tesis titulada: **Análisis transcriptómico de la germinación en semillas de *Cedrela odorata* L. para su conservación**, realizada por el (la) estudiante: **Lorena Gálvez López**, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTORA EN CIENCIAS
CIENCIAS FORESTALES

CONSEJO PARTICULAR

CONSEJERO


DR. JAVIER LÓPEZ UPTON

CO-DIRECTOR


DR. MIGUEL ÁNGEL VALLEJO REYNA

ASESORA


DRA. FLORENCIA TIBERIA AUCÁN GARCÍA CAMPUSANO

ASESOR


DR. MARCOS JIMÉNEZ CASAS

ASESOR


DR. ARNULFO ALDRETE

Montecillo, Texcoco, Estado de México, México, diciembre de 2022

ANÁLISIS TRANSCRIPTÓMICO DE LA GERMINACIÓN EN SEMILLAS DE *Cedrela odorata* L. PARA SU CONSERVACIÓN

Lorena Gálvez López, D.C.
Colegio de Postgraduados, 2022

RESUMEN

Cedrela odorata L. es una especie forestal tropical importante en el mercado de madera para la fabricación de muebles finos. Sus poblaciones se han sobreexplotado lo que la ha llevado a ser protegida. Una manera de usar y conservar la especie es guardar semillas en bancos de germoplasma; sin embargo, su viabilidad decrece en el corto plazo. Los mecanismos fisiológicos y moleculares implicados en el proceso germinativo son poco conocidos; por lo que su estudio es necesario para el desarrollo de nuevas estrategias de conservación y manejo de estas semillas. El objetivo fue identificar genes implicados en el proceso germinativo de semillas de *C. odorata* a través de un estudio basado en los cambios de expresión génica. Se generó un transcriptoma de *novo* considerando tres etapas del proceso germinativo: semilla madura seca (S), hidratada al 95% (E) y germinada (G). Después se realizó el estudio de expresión diferencial utilizando las comparaciones: SvsE, EvsG y SvsG. Para el análisis se usó programación en R y el paquete *DESeq2*. El umbral para considerar un gen diferencialmente expresado (DEG) fue obtener un valor de significancia (p_{adj}) ≤ 0.01 y un valor absoluto de cambio de expresión ($\text{Log}_2 \text{FoldChange}$) ≥ 2 . Las secuencias generadas fueron anotadas de acuerdo con bases de datos públicas preexistentes, lo que permitió identificar putativamente a los genes. Los DEG fueron analizados por medio de los términos GO y su enriquecimiento. El transcriptoma tuvo un total de 52,490 genes, un %GC de 42.62 y un valor N50 de 1,668. Se encontraron 22 grupos funcionales destacando las modificaciones postraduccionales y transcripción. En las categorías procesos biológicos (PB), componentes celulares (CC) y función molecular (FM) de la clasificación GO, predominaron los términos: “regulación de la transcripción-plantilla de ADN”, “núcleo” y “unión de ATP”, respectivamente. El análisis de expresión diferencial evidenció a 8,863 DEG. Durante el proceso germinativo se exhibieron dos etapas: la primera, relacionada con la reactivación metabólica posterior a la imbibición en la cual los genes con mayor cambio son aquellos relacionados con el metabolismo energético, la señalización y el metabolismo de la pared celular. En la segunda se presentaron genes relacionados con el alargamiento del embrión. También se evidenciaron cambios en la expresión de genes relacionados con tolerancia a la desecación como HSP, GGT1 y GOLS2. Los resultados contribuyen al conocimiento de los mecanismos fisiológicos y moleculares presentes en el proceso germinativo de *C. odorata* y genera un punto de partida a investigaciones que incluyan la asociación del genotipo con el manejo de semillas en el almacenamiento.

Palabras clave: Anotación funcional, gen diferencialmente expresado, proceso germinativo, transcriptoma.

TRANSCRIPTOMIC ANALYSIS OF GERMINATION IN SEEDS OF *Cedrela odorata* L. FOR THEIR CONSERVATION

Lorena Gálvez López, D.C.
Colegio de Postgraduados, 2022

ABSTRACT

Cedrela odorata L. is an important tropical forest species in the international wood market, used for manufacturing of fine furniture. Its populations are overexploited, which has led it to be protected. One way to use and conserve the species is by storing seed in germplasm banks, however, its viability decreases in a short time. The physiological and molecular mechanisms involved in the germination process are poorly understood; Therefore, its study is necessary for the development of new strategies for the conservation and management of these seeds. The objective was to identify genes involved in the germination process of *C. odorata* seeds through a study based on changes in gene expression. A de novo transcriptome was generated considering three stages within the germination process: dry mature seed (S), 95% hydrated (E) and germinated (G). Afterwards, the differential expression study was carried out using the comparisons: SvsE, EvsG and SvsG. For the analysis, R programming and the DESeq2 package were used. The threshold to consider a differentially expressed gene (DEG) was to obtain a significance value (p_{adj}) ≤ 0.01 and an absolute value of expression change ($\text{Log}_2 \text{FoldChange}$) ≥ 2 . The generated sequences were annotated according to pre-existing public databases which allowed putative identification of the genes. The DEGs were analyzed through the GO terms and their enrichment. The transcriptome had a total of 52,490 genes, a %GC of 42.62, and an N50 value of 1,668. 22 functional groups were found, highlighting the post-translational modifications and transcription. In the categories, biological processes (PB), cellular components (CC) and molecular function (FM) of the GO classification, the terms: "regulation of transcription-DNA template", "nucleus" and "ATP binding" predominated, respectively. Differential expression analysis evidenced at 8,863 DEG. During the germinative process two stages were exhibited: the first, related to metabolic reactivation after imbibition in which the genes with the greatest change are those related to energy metabolism, signaling and cell wall metabolism are expressed. In the second, genes related to the elongation of the embryo showed up. Changes in the expression of genes related to tolerance to desiccation such as HSP, GGT1 and GOLS2 were also evidenced. The results contribute to the knowledge of the physiological and molecular mechanisms present in the germination process of *C. odorata* and generate a starting point for further research that includes the association of the genotype with the handling of seeds in storage.

Key words: Functional annotation, differentially expressed gene, germinative process, transcriptome.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el financiamiento otorgado para para realizar mis estudios con el objetivo de obtener el grado de Doctora en Ciencias.

Al Colegio de Postgraduados (COLPOS), Campus Montecillo, especialmente a los profesores investigadores que me compartieron gran parte de su conocimiento. También agradezco especialmente **al Postgrado en Ciencias Forestales**, lugar donde desarrolle este grado académico.

Al Centro Nacional de Investigación Disciplinaria en Conservación y Mejoramiento de Ecosistemas Forestales (Cenid Comef) del Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) por el financiamiento otorgado al proyecto número 941634436 bajo el nombre: “Análisis transcriptómico de la germinación en *Cedrela odorata* L. para la identificación de marcadores moleculares para la conservación de semillas”, el cual fue financiado con recursos fiscales del mismo Instituto y permitió el desarrollo de esta investigación.

Al Dr. Miguel Ángel Vallejo Reyna por haberme dado la oportunidad de trabajar con él en este proyecto, fungiendo como director de tesis. Por su tiempo, guía, apoyo, enseñanzas y dedicación.

Al Dr. Javier López Upton, por ser mi consejero durante este periodo de formación. Gracias por su apoyo, consejos, aportes y revisiones.

A cada uno de los integrantes del comité asesor en esta investigación: **Dra. Florencia García Campusano, Dr. Marcos Jiménez Casas y Dr. Arnulfo Aldrete**. Gracias por el apoyo, sugerencias y aportes para mejorar esta investigación.

A mis amigas: Viridiana Aguilera e Hilda Méndez. Muchas gracias por todo su apoyo en el laboratorio. Gracias también por su valiosa amistad y por todos esos momentos muy agradables juntas. **A mis amigos y compañeros:** Gina Isabel, Gardenia, Mary Chuy, Elizabeth, Frida, Marce, Rigo con quienes compartimos momentos muy agradables, gracias por su amistad y apoyo.

A mi familia de Texcoco:

Sra. Carmen Zaragoza y Don Félix Velázquez, quienes siempre me han apoyado y han estado pendiente de mi en todo momento.

A Edith, Patricia, Norma, Xóchilt y Don Salome Rangel (†). Por todo su apoyo brindado y esos momentos de convivencia. Lo extrañamos mucho don Salo, ¡besos hasta el cielo!

Al señor Carlos Zaragoza y su esposa Teresa Venegas, quienes también nos han apoyado durante estos años.

A mi amiga Marilú Gómez, gracias por todos estos años de amistad, apoyo y convivencia.

DEDICATORIA

A Dios, te doy gracias por la vida y la oportunidad de permitirme llegar hasta este momento. Porque sé que sin ti no soy nada. Porque todo lo puedo en Cristo que me fortalece (Filipenses 4:3).

A mis padres: Didier Gálvez y Juanita López. Muchas gracias por darme la vida, por su amor, apoyo, comprensión, cariño, ejemplo y todo lo que hacen por mí. Aunque físicamente estén lejos, mi corazón siempre está con ustedes. ¡Los amo tanto!

A mi esposo Jose Luis Rangel y nuestro pequeño José Said. Gracias a Dios por cruzarnos en el camino y permitirnos escribir esta historia juntos. Gracias por tu amor y todo tu apoyo siempre. Y a nuestro bebé que ha venido a darnos tanto amor y alegría.

¡Te amamos Said!

A mis hermanas: Didi, Chío, Lau, Kary y Caro. Gracias por estar ahí siempre. Por todos sus consejos y apoyo. Gracias a Dios que nos permitió crecer juntas y compartir tantos momentos de alegría.

CONTENIDO

RESUMEN	iii
ABSTRACT	iv
AGRADECIMIENTOS.....	v
DEDICATORIA	vii
LISTA DE CUADROS.....	xi
LISTA DE FIGURAS.....	xiii
SIGLAS Y ABREVIATURAS	xviii
INTRODUCCIÓN GENERAL.....	1
REVISIÓN DE LITERATURA	8
Resumen.....	8
Abstract	9
Introducción	1
El genoma de <i>Cedrela odorata</i> L.	2
Aportaciones de la biología molecular al estudio filogenético de cedro rojo	4
Diversidad genética	5
Mejoramiento genético convencional y asistido por biología molecular	5
Conclusiones	9
CAPÍTULO I. GENERACIÓN Y ANOTACIÓN DEL TRANSCRIPTOMA DE SEMILLAS DE <i>Cedrela odorata</i> L. DURANTE LA GERMINACIÓN.....	10
1.1 RESUMEN	10
1.2 ABSTRACT	11
1.3 INTRODUCCIÓN	12
1.4 MATERIALES Y MÉTODOS	15
1.4.1 Material vegetal y tratamientos	15
1.4.2 Extracción del ARN Total.....	16
1.4.3 Preparación de bibliotecas de ADNc y secuenciación	17
1.4.4 Análisis de calidad de la secuencia	17
1.4.5 Ensamble de <i>novo</i> del transcriptoma.....	17
1.4.6 Anotación del transcriptoma	18

1.5 RESULTADOS	19
1.5.1 Extracción del ARN.....	19
1.5.2 Análisis de calidad de la secuencia	20
1.5.3 Ensamble de <i>novovo</i> del transcriptoma.....	29
1.5.4 Anotación funcional.....	30
1.5.5 Clasificación de las funciones eggNOG.....	31
1.5.6 Clasificación Ontológica.....	32
1.5.7 Base de datos Nr	34
1.6 DISCUSIÓN	35
1.7 CONCLUSIONES	42
CAPÍTULO II. ANÁLISIS DE EXPRESIÓN DIFERENCIAL DE GENES DURANTE LA GERMINACIÓN DE SEMILLAS DE <i>Cedrela odorata</i> L.	43
2.1 RESUMEN	43
2.2 ABSTRACT	44
2.3 INTRODUCCIÓN	45
2.4 MATERIALES Y MÉTODOS	48
2.4.1 Cuantificación de transcritos.....	48
2.4.2 Análisis de expresión diferencial.....	48
2.4.3 Clasificación de la ontología génica de los DEG	49
2.4.4 Mapa de calor y anotación individual de los DEG con mayor y menor expresión	49
2.4.5 Enriquecimiento GO de los DEG con mayor y menor expresión	50
2.5 RESULTADOS	51
2.5.1 Cuantificación de transcritos.....	51
2.5.2. Análisis de la expresión diferencial	51
2.5.3 Clasificación de la Ontología Génica	52
2.5.4 Mapa de calor y anotación individual de los DEG con mayor y menor expresión	55
2.5.5 Enriquecimiento de los términos GO de los DEG con mayor y menor expresión	58
2.6 DISCUSIÓN	60

2.6.1 Reactivación metabólica durante la hidratación de la semilla.....	60
2.6.2 Expansión celular y germinación <i>sensu stricto</i>	66
2.6.3 Expresión diferencial de genes en la comparación de semillas secas vs. germinadas	70
2.7 CONCLUSIONES	75
CONCLUSIONES GENERALES	76
LITERATURA CITADA	78
ANEXOS	94

LISTA DE CUADROS

Cuadro 1. Heredabilidad estimada para las principales variables dasométricas en <i>Cedrela odorata</i> L. en ensayos de progenie establecidos en el campo experimental “El Palmar”, Tezonapa, Veracruz, México.....	8
Cuadro 1.1. Lecturas promedio tomadas con un Nanodrop para evaluar la cantidad y calidad del ARN total en semillas de <i>Cedrela odorata</i> L.	19
Cuadro 1.2. Resumen de la estadística del ensamble de transcriptoma de semillas de <i>Cedrela odorata</i> L.	30
Cuadro 1.3. Análisis estadístico de la anotación con <i>Trinotate</i> del transcriptoma de <i>novus</i> de semillas de <i>Cedrela odorata</i> L. Se muestra el número de secuencias únicas y totales en las distintas bases de datos.	31
Cuadro A1. Resumen de los análisis de calidad de los 18 archivos <i>fastq</i> correspondientes a las tres condiciones evaluadas: seca, embebida al 95% y germinada, cada una con sus tres repeticiones y cada repetición con su respectivo <i>Read 1</i> y <i>Read 2</i> en semillas de <i>Cedrela odorata</i> L.	94
Cuadro A2. Resumen de la estadística básica de los 18 archivos <i>fastq</i> correspondientes a las tres condiciones evaluadas: seca, embebida al 95% y germinada, cada una con sus tres repeticiones y cada repetición con su respectivo <i>Read 1</i> y <i>Read 2</i> en semillas de <i>Cedrela odorata</i> L.	95
Cuadro A3. Anotación funcional con <i>Trinotate</i> del transcriptoma completo de semillas de <i>Cedrela odorata</i> L. en las condiciones seca, embebida al 95% y germinada.....	96
Cuadro A4. Análisis de genes diferencialmente expresados en semillas de <i>Cedrela odorata</i> L. en la condición SvsG.	97
Cuadro A5. Análisis de genes diferencialmente expresados en semillas de <i>Cedrela odorata</i> L. en la condición SvsE.....	99
Cuadro A6. Análisis de genes diferencialmente expresados en semillas de <i>Cedrela odorata</i> L. en la condición EvsG.	101
Cuadro A7. Anotación funcional manual apoyados de las bases de datos de NCBI y TAIR en semillas de <i>Cedrela odorata</i> L.	103

Cuadro A8. Resultados de enriquecimiento GO de los PB de los 100 genes (50 aumentados y 50 disminuidos) con mayor nivel de expresión (<i>log2FoldChange</i>) en la comparación SvsE en semillas de <i>Cedrela odorata</i> L.	104
Cuadro A9. Resultados de enriquecimiento GO de los PB de los 100 genes (50 aumentados y 50 disminuidos) con mayor nivel de expresión (<i>log2FoldChange</i>) en la comparación EvsG en semillas de <i>Cedrela odorata</i> L.	106
Cuadro A10. Resultados de enriquecimiento GO de los PB de los 100 genes (50 aumentados y 50 disminuidos) con mayor nivel de expresión (<i>log2FoldChange</i>) en la comparación SvsG en semillas de <i>Cedrela odorata</i> L.	107

LISTA DE FIGURAS

- Figura 1.** Fases de la germinación de la semilla. Se muestran los distintos eventos que ocurren durante el proceso germinativo y la postgerminación. La línea de en medio muestra la dinámica de absorción de agua en las tres fases (Tomado de Zaynab *et al.*, 2018). 5
- Figura 2.** Representación esquemática de las etapas de un programa de mejoramiento genético de *Cedreia odorata* L..... 6
- Figura 1.1.** Electroforesis en gel de agarosa al 1.5% del ARN total extraído de las semillas S=Seca, E=embebida al 95% y G= germinada. Las dos distintas bandas de ARNr muestran la integridad del ARN. 19
- Figura 1.2.** Estadística básica del análisis de calidad con FastQC. Se muestra el *read* 1 y 2 de los archivos *fastq* que corresponde a la muestra 1 de las semillas secas de *Cedreia odorata* L..... 20
- Figura 1.3.** Calidad de secuencia por base para el *read* 1 y 2 que corresponde a la muestra 1 de las semillas secas de *Cedreia odorata* L. Se presenta un gráfico de cajas para cada base. La línea roja representa la mediana y la azul el valor medio de las bases. El eje X muestra la posición de cada base en una secuencia y el eje Y representa el valor de calidad de cada base (valor Q30), en donde a mayor valor mejor es la calidad. 21
- Figura 1.4.** Mapa de calor de la calidad de la secuencia que muestra las celdas de flujo para el *read* 1 y 2 que corresponde a la muestra 1 de las semillas secas de *Cedreia odorata* L. En el eje de las X se muestran la posición de los *read* y en el eje de las Y las secuencias asociadas a los *reads*. 22
- Figura 1.5.** Estimador de calidad de la secuencia para el *read* 1 y 2 que corresponde a la muestra 1 de las semillas secas de *Cedreia odorata* L. En el eje de las X se muestra la calidad media de las secuencias y en el eje de las Y el número de secuencias que corresponden a esa media..... 23

Figura 1.6. Proporción del contenido de nucleótidos por base que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. En el eje de las X se muestra la posición del *read* y en el eje de las Y el porcentaje de los nucleótidos en cada base de los *reads*. 24

Figura 1.7. Proporción de GC en los *reads* (línea roja) comparándose con una distribución normal teórica (línea azul) que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. El eje X muestra el porcentaje medio del contenido de GC y el eje de las Y el número de *reads*. 25

Figura 1.8. Presencia de N (Nucleótido desconocido) en la secuencia que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. En el eje de las X se muestra la posición del *read* y en el eje Y el porcentaje de N en los *reads*. 25

Figura 1.9. Distribución del tamaño de *reads* que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. En el eje de las X se muestra la longitud de la secuencia y en el eje de las Y el número de secuencias que corresponden a esa longitud. 26

Figura 1.10. Niveles de duplicación de secuencia que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. En el eje de las X se muestran el nivel de secuencias duplicadas y en el eje de las Y el porcentaje asociado a cada uno de los niveles de duplicación. La línea azul indica el total de secuencias y la roja muestra las secuencias duplicadas. 27

Figura 1.11. Secuencias sobrerrepresentadas que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. Solo se muestran las secuencias que constituyen más de un 0.1 % del total. 28

Figura 1.12. Contenido de adaptadores en la secuencia que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. En el eje de las X se muestran la posición en el *read* y en el eje de las Y el porcentaje del contenido de adaptadores asociado a los *reads*. 29

Figura 1.13. Clasificación funcional de proteínas ortólogas obtenidas del transcriptoma de <i>novo</i> del proceso germinativo de <i>Cedrela odorata</i> L.....	32
Figura 1.14. Clasificación de los términos GO obtenidos del transcriptoma de <i>novo</i> del proceso germinativo de <i>Cedrela odorata</i> L.	33
Figura 1.15. Distribución de aciertos en la base de datos Nr del transcriptoma de <i>novo</i> del proceso germinativo de <i>Cedrela odorata</i> L.	34
Figura 2.1. Análisis del número de genes expresados diferencialmente (DEG), identificados en tres etapas de germinación de semillas en <i>C. odorata</i> . A) Diagrama de volcán que muestra el patrón de expresión entre semillas secas contra embebidas (SvsE), B) embebidas c contra germinadas (EvsG) y C) secas contra germinadas (SvsG). El eje X representa el valor $Log_2FolChange$ y el eje Y el nivel de significación según $-\log_{10}(padj)$. Los puntos azules y rojos representan los DEG regulados en forma positiva y negativa, respectivamente, los puntos negros representan genes que no se expresaron de manera diferencial. D) Diagrama de Venn que muestra superposiciones de DEG totales; E) superposición de los genes regulados positivamente y F) genes regulados negativamente. Los números entre paréntesis indican el porcentaje de DEG en cada situación de los diferentes grupos.	53
Figura 2.2. Ontología genética (GO) de los genes expresados diferencialmente (DEG) en las comparaciones: seca vs embebida (SvsE), embebida vs germinada (EvsG) y seca vs germinada (SvsG). A) Clasificación funcional de los términos GO. El eje X representa los términos organizados en tres categorías: componente celular, función molecular y procesos biológicos. El eje Y, en el lado izquierdo, muestra el porcentaje de genes anotados en cada subgrupo de cada categoría y en el lado derecho, el número de genes anotados en cada comparación. B) Diferencias significativas en función del $\log_{10} p\text{-value}$ ($<0,05$) de los procesos biológicos analizados, organizados en orden	

descendente.2.5.4 Mapa de calor y anotación individual de los DEG con mayor y menor expresión 54

Figura 2.3. A) Mapa de calor que muestra el agrupamiento y patrón de expresión de 247 genes expresados diferencialmente (DEG), según el valor *log2FoldChange*. Cada fila representa un gen y cada columna la comparación en la que se expresa el gen: seca vs embebida (SvsE), embebida vs germinada (EvsG) y seca vs germinada (SvsG). Los colores rojo y azul indican DEG regulados al alza y a la baja, respectivamente, mientras que el blanco indica que no hubo cambios. Cuanto más fuerte es el color, mayor es la expresión. **B)** Gráficos hipotéticos que representan la tendencia en el comportamiento de los doce grupos representados en el mapa de calor. Las letras S, E y G, indican las condiciones de tratamiento: seca, embebida y germinada respectivamente, n, indica el número de genes en ese grupo. 56

Figura 2.4. Análisis REVIGO de los procesos biológicos en 100 genes diferencialmente expresados (DEG), regulados al alza y a la baja en las comparaciones: **A)** Seca vs embebida (SvsE), y **B)** embebida vs germinada (EvsG). Cada rectángulo representa un grupo. Los representantes se clasifican en "supergrupos" de términos vagamente relacionados, reflejados por diferentes colores. La frecuencia del término GO se refleja en el tamaño de los rectángulos..... 59

Figura 2.5. Genes diferencialmente expresados durante la germinación de semillas de *C. odorata*. La flecha azul con sus respectivos enunciados indica los procesos y genes que se expresaron durante el paso de la semilla seca a embebida (SvsE), la verde durante el paso de la semilla embebida a germinada (EvsG) y la negra aquellos que se expresaron durante todo el proceso (SvsG). Las flechas verdes y rojas indican que el gen aumento o disminuyo su expresión respectivamente. 74

Figura A1. Análisis REVIGO de los procesos biológicos en 100 genes expresados diferencialmente (DEG), regulados al alza y a la baja en la comparación Seca vs Germinada (SvsG). Cada rectángulo representa

un grupo. Los representantes se clasifican en "supergrupos" de términos vagamente relacionados, reflejados por diferentes colores. La frecuencia del término GO se refleja en el tamaño de los rectángulos..... 109

SIGLAS Y ABREVIATURAS

- ABA:** Ácido abscísico.
- ABI3:** Insensible al ácido abscísico 3.
- ABI5:** Insensible al ácido abscísico 5 .
- ACO1:** Acc oxidasa1.
- ACT11:** Actina-11.
- ADNc:** ADN complementario.
- AG:** Ácido giberélico.
- AGP:** Arabinogalctanos.
- AGP14:** Proteín arabinogalactan 14.
- AGP41:** Proteín arabinogalactano 41.
- AMY3:** Alfa amilasa 3.
- APM1:** Aminopeptidasa M1
- ARNm:** ARN mensajero.
- ARNr:** ARN ribosomal.
- ARNt:** ARN de transferencia.
- HIGD3:** Dominio de gen 3 inducido por hipoxia.
- ATLP-1:**Proteína similar a la taumatina 1.
- ATP:** Adenosín trifosfato.
- BBX18:** Proteína 18 de dedos de Zinc de caja BB-box dominio proteína 18.
- BXL4:** Beta-d-xilosidasa 4
- BZIP61:** Codifica un miembro de la familia BZIP de factores de transcripción.
- CAD7:** Cinamil-alcohol deshidrogenasa 7.
- CC:** Componentes Celulares.
- CITES:** Convención sobre el Comercio Internacional de Especies Amenazadas de Fauna y Flora Silvestres.
- CLE4:** Clavata3/esr-relacionado 4.
- COG:** Grupos de grupos ortólogos.
- COX1:**Citocromo oxidasa 1.
- CSD2:** Superóxido dismutasa cloroplástica de cobre/zinc.
- CYP705A22:** Citocromo P450 familia 70.

DEG: Gen diferencialmente expresado.

E: Semilla hidratada al 95 %.

EggNOG: Database of biological information hosted by the EMBL (Base de datos de información biológica alojada por el EMBL).

ERF: Factor de respuesta al etileno.

ERFVII: Factores de respuesta a etileno del grupo VII.

ESE3: Inducible de etileno y sal 3.

EvsG: Embebida vs germinada.

FDR: Tasa de descubrimiento falso.

FGGY: Familia de carbohidrato quinasa.

FLA: Fasciclina.

FLA6: Arabinogalactano tipo fasciclina 6.

FLA7: Arabinogalactano tipo fasciclina 7.

FLS6: Flavonol sintasa 6.

FM: Función molecular.

FoldChange: Cambio relativo en la expresión.

FUS3: FUSCA3.

G: Semilla germinada.

GC: Guanina-Citosina

GGT: Gamma-glutamyl transferasa.

GGT1: Gamma-glutamyl transpeptidasa 1.

GMC: Proteína de la familia de las oxidoreductasas de glucosa-metanol-colina.

GO: Ontología Génica.

GOLS2: Galactinol sintasa 2.

GSA6: Familia de proteínas regulada por la giberelina.

GSH :Glutación.

HEC2:HECATE 2.

HO1: Heme Oxigenasa 1.

HSP: Proteína de choque térmico.

HSP20: Proteína de choque térmico 20.

HSP70: Proteína de choque térmico 70.

HSP70B: Proteína de choque térmico 70 B.

H2O2: Peróxido de hidrogeno.

IAA: Ácido indol-3-acético.

IAA4: Ácido indol-3-acético inducible 4.

IAA9: Ácido indol-3-acético inducible 9.

IAA14: Ácido indol-3-acético inducible 14.

ITS: Regiones intergénicas.

IUCN: Unión Internacional para la Conservación de la Naturaleza.

KEEG: Kyoto Encyclopedia of Genes and Genomes. (Enciclopedia de genes y genomas de Kyoto).

LACCASE 7: Miembro de la familia de genes lacasa.7.

LEA: Proteínas abundantes en embriogénesis tardía.

LEC1: LEAFY COTYLEDON1.

LEC2: LEAFY COTYLEDON 2.

LTP1: Proteína de transferencia de lípidos 1.

MAN7: Endo-beta-mananasa 7.

MAS: Marker Assisted Selection. (Selección asistida por marcadores).

NCBI: National Center for Biotechnology Information. (Centro Nacional para la Información Biotecnológica).

NGS: Next Generation Sequencing .(Secuenciación de nueva generación).

NO: Óxido nítrico.

Nr: No redundantes.

O2 -°: Anión superóxido.

°OH: Radical hidroxilo.

PAP3: Fosfatasa acida 3.

PB: Procesos biológicos.

PCO: Oxidasa de cisteína vegetal.

PCO2: Oxidasa de cisteína vegetal 2.

PER3: Peroxidasa 3.

PER4: Peroxidasas 4.

Pfam: Protein family. (Familia de proteínas).

PGK2: Fosfoglicerato quinasa 2.

PMEI13: Pectina metilesterasa inhibidor 13.

PME41: Pectina metilesterasa 41

PRX52: Peroxidasa 52.

PSAO: Subunidad O del fotosistema I.

Read 1: Sentido.

Read 2: Antisentido.

RLP51: Proteína similar a receptor 51

RNA-seq: Secuenciación de ARN.

ROS: Especies reactivas de oxígeno.

RPL30B: Proteína ribosómica L7Ae/L30e/S12e/ proteína de la familia Gadd 45.

S: Semilla madura seca.

SAUR: ARN pequeño de auxina.

SAUR63: ARN regulado al alza de auxina pequeña 63.

SAUR46: ARN regulado al alza de auxina pequeña 46.

SEMARNAT: Secretaria de Medio Ambiente y Recursos Naturales.

SOD2: Superóxido dismutasa 2.

STP7: Proteína transportadora de azúcar 7.

ST2A: Sulfotransferasa 2A.

SvsE: Seca vs embebida.

SvsG: Seca vs germinada.

TAIR: *Arabidopsis* Information Resource. (Recurso de información de *Arabidopsis*).

TCP15: Teosinte ramificado1/cycloidea/PCF 15.

TPM: Transcritos por Millón.

TPS02: Terpeno sintasa 02.

TUA2: Alfa-tubulina 2,4.

TUB3: Cadena de tubulina beta-2/beta-3.

USPL1: Proteína del dominio BURP dirigida a las vacuolas.

WC: Contenido de agua.

XTH: Xiloglucano endotransglucosilasa/hidrolasa.

XTH8: Xiloglucano endotransglucosilasa/hidrolasa 8.

XTH9: Xiloglucano endotransglucosilasa/hidrolasa 9

XTH16: Xiloglucano endotransglucosilasa/hidrolasa16.

INTRODUCCIÓN GENERAL

Cedro rojo (*Cedrela odorata* L.) es una especie forestal maderable de la familia meliácea. Se distribuye en México por la vertiente del golfo, desde el Sur de Tamaulipas hasta la península de Yucatán y en la vertiente del Océano Pacífico desde Sinaloa hasta Colima y en la costa de Chiapas (Pennington y Sarukhán, 2005). Tiene alta importancia en la industria forestal debido a distintos usos de su madera, especialmente la fabricación de muebles finos (Cintrón, 1990). Esto la hace vulnerable por la explotación ilegal de los mejores fenotipos en su zona de distribución natural (de la Torre, 2013); además, se suma a ello la reducción de la cubierta vegetal generada por el cambio de uso de suelo para cultivos agrícolas, potreros y asentamientos humanos (Orantes-García *et al.*, 2013). Debido a ello, es catalogada como especie protegida por algunas instituciones como la Convención sobre el Comercio Internacional de Especies Amenazadas de Fauna y Flora Silvestres (CITES) (CITES, 2017), la Unión Internacional para la Conservación de la Naturaleza (IUCN) (IUCN, 2017) y en México por la Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT) a través de la NOM-059 (SEMARNAT, 2010). Por lo tanto, es considerada como una especie prioritaria dentro de los programas de conservación a nivel nacional.

El almacenamiento de semillas en bancos de germoplasma representa una estrategia importante de conservación *ex situ* a largo plazo, ya que favorece resguardar genotipos sobresalientes y preservar la diversidad genética de la especie. Sin embargo, este tipo de almacenamiento no es factible para todas las especies, debido a la variabilidad con la cual envejecen naturalmente las semillas. La pérdida de viabilidad y vigor puede ser gradual o rápida según el origen y genotipo de la semilla a almacenar, lo cual es un reflejo de características intrínsecas como su tolerancia a la desecación, es decir su capacidad de sobrevivir al secado y después rehidratarse completamente para originar una nueva plántula, además de otras particularidades como su tolerancia a las bajas temperaturas (Verdier *et al.*, 2013).

La longevidad de la semilla se adquiere durante la maduración en la planta madre (Verdier *et al.*, 2013; Sano *et al.*, 2016), aumentando progresivamente durante esta

etapa, después de adquirir tolerancia a la desecación (Sano *et al.*, 2016). De manera que se define como longevidad de la semilla al periodo en el cual la semilla sigue siendo viable. Dependiendo del tipo semillas, estas pueden tener una longevidad desde meses, años, décadas hasta milenios (Verdier *et al.*, 2013; Sano *et al.*, 2016). No obstante a lo anterior, la longevidad también dependerá de las condiciones de manejo de las semillas previo a y durante el almacenamiento como la humedad relativa, el contenido de humedad de la semilla, la presión de oxígeno y la temperatura a la cual es expuesta (Sano *et al.*, 2016).

A nivel celular, los factores que influyen en la reducción de la longevidad de la semilla están asociados con la oxidación de ácidos nucleicos, proteínas y lípidos, mediadas por las especies reactivas de oxígeno (ROS). Dos de los métodos para evitar la oxidación (estrés) son: en primer lugar, el mecanismo de protección en el que el citoplasma pasa de un estado líquido a un estado vítreo, minimizando las actividades metabólicas y generando sistemas antioxidantes, lo que evita la acumulación de macromoléculas oxidadas (Sano *et al.*, 2016). Dicho cambio de estado esta mediado por proteínas de choque térmico (*Heat Shock Proteins*, HSP, por sus siglas en inglés), proteínas abundantes en embriogénesis tardía (*late embryogenesis abundant*, LEA, por sus siglas en inglés) y azúcares. Los genes responsables de estas proteínas presentan mayor grado de expresión a medida que aumenta la maduración y secado de las semillas, por lo que se les relaciona con la tolerancia a la desecación (Bewley *et al.*, 2013). El segundo es el sistema de reparación, que elimina el daño acumulado en el ADN, ARN y proteínas después de la imbibición, mediante enzimas como la ADN glicosilasa y la metionina sulfóxido reductasa (Sano *et al.*, 2016) o bien las ADN ligasas y la proteína L-isoaspartil metiltransferasa (Verdier *et al.*, 2013).

Es así como las semillas se clasifican según su tolerancia a la desecación en: ortodoxas, intermedias y recalcitrantes. Las ortodoxas soportan una deshidratación hasta del 5 %, las intermedias entre un 10 y 12.5 % y las recalcitrantes entre un 15 % y 50 % (Magnitskiy y Plaza, 2007). Es decir, las semillas ortodoxas tienen alta tolerancia a la desecación, las intermedias no toleran un secado completo, mientras que las recalcitrantes son sumamente sensibles a la desecación (Dussert *et al.*, 2018) y son

susceptibles a la fuga de electrolitos y daño a la membrana (Obroucheva *et al.*, 2016). Además, durante el almacenamiento, las semillas ortodoxas toleran las bajas temperaturas; mientras que las intermedias y recalcitrantes exhiben mayor sensibilidad (Hay y Probert, 2013; Stavrinides *et al.*, 2020). Dada estas características, las semillas ortodoxas, bajo condiciones adecuadas como la temperatura y contenido de humedad, se pueden almacenar en bancos de semillas por periodos de tiempo prolongados manteniendo su viabilidad; mientras que las semillas no ortodoxas bajo dichas condiciones de almacenamiento pueden sobrevivir uno o hasta dos años (Hay y Probert, 2013). A menudo muchas de las especies tropicales se caracterizan por tener un comportamiento recalcitrante, tal es el caso del coco y el cacao, o bien son semillas intermedias como el café y los críticos (Stavrinides *et al.*, 2020).

Las semillas de *C. odorata* se clasifican como ortodoxas, pero pierden la viabilidad rápidamente. Durante el almacenamiento, el mejor registro para mantener la viabilidad es de aproximadamente 10 meses, almacenadas a 2 °C con un 4 % en contenido de humedad (CONABIO,1983). Dadas las características de las semillas de *C. odorata* como la sensibilidad a las bajas temperaturas, la dificultad de reducir el contenido de humedad y la reducida longevidad, deben ser consideradas como intermedias (García *et al.*, 2015).

Dentro de la caracterización de estas clasificaciones de semillas se ha observado que las recalcitrantes se desprenden de la planta madre durante la estación húmeda con un contenido de humedad alto respecto a las ortodoxas, lo que las hace tener un metabolismo más activo y mayor movilización de reservas, provocando su rápida germinación (Obroucheva *et al.*, 2016). Así mismo, se caracterizan por tener un mayor tamaño y tejidos externos relativamente delgados (Hay y Probert, 2013; Obroucheva *et al.*, 2016). De igual manera Magnitskiy y Plaza (2007) consideran que las semillas recalcitrantes poseen un alto porcentaje de saturación de ácidos grasos en las membranas (32 y 36 %) lo cual se considera alto respecto a las ortodoxas que tienen un porcentaje entre 16 y 27 %. Sin embargo, Obroucheva *et al.* (2016) consideran que el almacenamiento de lípidos no es un distintivo de las semillas recalcitrantes debido a que el castaño de indias (*Aesculus hippocastanum* L.), siendo una semilla recalcitrante,

tiene un contenido de aceite bajo en cotiledones y ejes embrionarios, 6.2 y 7.9 %, respectivamente. Además, en el *Seed Information Database* (Royal Botanic Gardens Kew, 2022), las semillas recalcitrantes tienen un contenido de aceite que varía de 0.5 % en *Durio zibethinus* Murray y 66 % en *Cocos nucifera* L. En semillas intermedias varía de 7 % para *Dipterocarpus alatus* Roxb. ex G. Don y 64 % para *Corylus avellana* L. y en ortodoxas 2.65 % para semillas de *Abrus precatorius* L. y 49.7 % en *Osteospermum hyoseroides* (DC.) Norl. Otras de las características distintivas en semillas recalcitrantes son los carbohidratos como reservas de almacenamiento predominante, las bajas cantidades de oligosacáridos y alta cantidad en sacarosa (Obroucheva *et al.*, 2016).

De manera general, existe una regulación hormonal y distintas vías de señalización involucradas en la longevidad de las semillas. Por ejemplo, los reguladores transcripcionales *LAFL* que incluye *ABI3* (Abscisic Acid Insensitive3), *FUS3* (*FUSCA3*), *LEC2* (*LEAFY COTYLEDON2*) y *LEC1* (*LEAFY COTYLEDON1*). En semillas de *Arabidopsis* se encontró que la falta de expresión de *ABI3* y *LEC1* hizo que las semillas perdieran su viabilidad pocas semanas después de ser recolectadas, lo que hace augurar que ambos factores influyen en la longevidad de las semillas (Zinsmeister *et al.*, 2020). Así mismo, se han encontrado también otros mecanismos que intervienen en la supervivencia de las semillas cuando están secas, entre ellas se encuentran los azúcares no reductores, proteínas LEA, HSP, defensas antioxidantes como el glutatión, tocoferoles y flavonoides (Verdier *et al.*, 2013). La información en semillas intermedias aún es escasa, por ello, es indispensable la búsqueda del conocimiento que permita comprender a fondo los mecanismos involucrados en la pérdida de viabilidad en semillas de *C. odorata*.

Durante el desarrollo de la semilla se sintetizan grandes cantidades de proteínas (proteosíntesis) de las cuales algunas se degradan y otras permanecen hasta el momento de la rehidratación de la semilla (Matilla, 2008), muchas de las cuales son importantes para activar durante la germinación (Zaynab *et al.*, 2018). Así el ARNt, factores de iniciación y elongación, ribosomas, aminoácidos o aminoacil-ARNt sintetizadas, entre otros, se encuentran almacenados y protegidos en la semilla seca

para realizar una parte de la síntesis proteica necesaria para el inicio de la germinación (Matilla, 2008). Dicho proceso requiere pasar por una serie de fases fisiológicas que exigen de actividad metabólica acelerada (Doria, 2010; Bewley *et al.*, 2013; Zaynab *et al.*, 2018): Fase I: la hidratación y rápida absorción de agua e inicio de actividades metabólicas. Fase II: aumento de las actividades metabólicas, se da lugar a la síntesis de mitocondrias, los ARNm almacenados son degradados o bien traducidos a proteínas e inicia la transcripción de nuevos genes. El rompimiento de la testa por la manifestación de la radícula marca la finalización de la germinación. Fase III: crecimiento de la radícula, que se caracteriza por una mayor absorción de agua como parte del crecimiento de la planta (**Figura 1**).

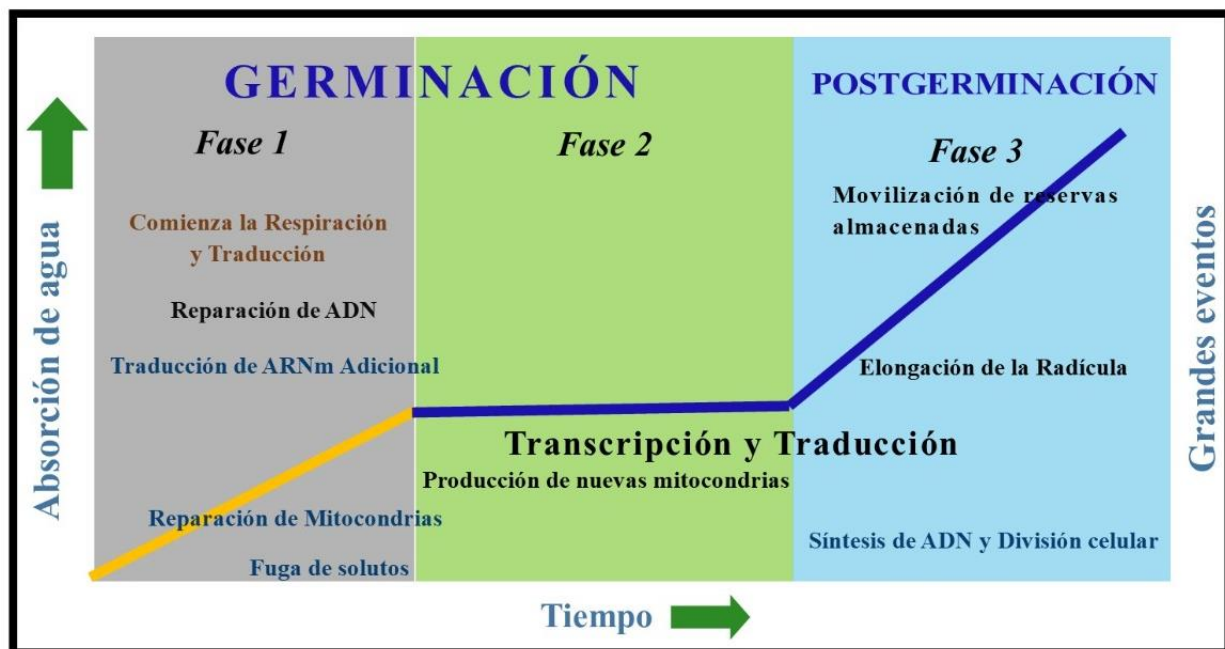


Figura 1. Fases de la germinación de la semilla. Se muestran los distintos eventos que ocurren durante el proceso germinativo y la postgerminación. La línea de en medio muestra la dinámica de absorción de agua en las tres fases (Tomado de Zaynab *et al.*, 2018).

Actualmente existen nuevas tecnologías que permiten realizar estudios de expresión génica, por lo que el análisis del proceso germinativo en semillas puede ser estudiado por medio de ellas. A través de los avances tecnológicos con la secuenciación de nueva generación (*Next Generation Sequencing*, NGS, por sus siglas en inglés) se ha innovado en el campo de la transcriptómica y la genómica, (Soto y López, 2012). La transcriptómica, mediante la técnica de *RNA-seq* (secuenciación de

ARN), se enfoca en analizar la expresión génica a nivel del ARN y brinda información acerca de la estructura y función de los genes dando información de los mecanismos moleculares involucrados en ciertos procesos biológicos (Alonso-Sánchez *et al.*, 2017). Permite además analizar la expresión de genes desconocidos e identificar eventos de *splicing* alternativo, además de que ofrece un rango amplio de cuantificación (Alonso-Sánchez *et al.*, 2017). El transcriptoma se define como el conjunto completo de las transcripciones en una célula o tejido y su cantidad en una etapa específica de desarrollo o condición fisiológica, que incluye el ARN mensajero (ARNm), ARN ribosomal (ARNr) y otros ARN no codificantes (Wang *et al.*, 2009; Alonso-Sánchez *et al.*, 2017). La transcripción es un proceso llevado a cabo en el núcleo de la célula y su activación depende de estímulos intra y extracelulares encendiendo cascadas de señalización para determinar los genes que se expresan o reprimen (Soto y López, 2012).

Los estudios de secuenciación usando tecnología NGS son relativamente recientes, por lo tanto, los trabajos enfocados en realizar estudios transcriptómicos en semillas son poco frecuentes, sobre todo en especies forestales. En *C. odorata* se tiene información del transcriptoma de hoja (Finch *et al.*, 2019), pero no hay información del transcriptoma en semillas. En este estudio el transcriptoma en semillas que parte desde una semilla seca y viable hasta la germinación *sensu stricto* permitirá conocer los mecanismos moleculares predominantes relacionados con los cambios bioquímicos y fisiológicos involucrados durante las fases de la germinación, así como los genes presentes en la regulación de la expresión génica.

Por lo tanto, el objetivo general del presente estudio fue identificar genes implicados en el proceso germinativo de semillas de *C. odorata* L. a través de estudios basados en el análisis de los cambios de expresión génica. Para ello se abordan los objetivos específicos que se desglosan en dos capítulos. En el capítulo I se describe la generación de un transcriptoma de referencia anotado para semillas de *C. odorata* L. En el capítulo II se presenta el análisis de expresión diferencial de genes (DEG) en tres etapas del proceso germinativo: en semillas quiescentes secas, hidratadas a 95 % y germinadas *sensu stricto*. Así mismo, se llevó a cabo una anotación individual de los

100 DEG con mayor y menor expresión en las comparaciones seca vs. embebida (SvsE), embebida vs. germinada (EvsG) y seca vs. germinada (SvsG) así como un análisis de enriquecimiento GO con REVIGO. Este análisis permitió distinguir dos fases distintivas en el proceso germinativo, con cambios en la expresión de genes asociados, primeramente, a la reactivación metabólica, reparación y respuesta a estrés; seguido de aquellos que favorecen los procesos de emergencia de la radícula, como genes implicados en el metabolismo de la pared celular y el alargamiento del embrión.

El conocimiento de los procesos fisiológicos y moleculares que se asocian a una germinación vigorosa en las semillas de *C. odorata* es fundamental para posteriores estudios enfocados en la búsqueda de estrategias para el manejo y conservación de las semillas a corto, mediano y largo plazo.

REVISIÓN DE LITERATURA

Cedrela odorata L. OPORTUNIDADES PARA SU CONSERVACIÓN Y MEJORAMIENTO GENÉTICO¹

Resumen

Cedrela odorata, conocida comúnmente como cedro rojo, es una especie forestal tropical de alto valor comercial en el mercado internacional de la madera. La fragmentación del paisaje en su área de distribución y la tala clandestina con selección disgénica causan erosión genética en sus poblaciones naturales, lo que disminuye su potencial adaptativo y productivo. Debido a esto, es más susceptible a daños por factores abióticos y bióticos como *Hypsipyla grandella* Zeller, su principal plaga y causa de bajo rendimiento en plantaciones comerciales. Es así, que se incluye en las listas de organizaciones internacionales como IUCN y CITES, y en México es protegida por la NOM-059 SEMARNAT-2010; sin embargo, la falta de suficientes evidencias sobre su riesgo real pone en duda su permanencia en esta última. Por ello, aunque se han realizado algunas investigaciones para determinar su diversidad genética en poblaciones de América Central, cariotipo, caracterización de su plastoma y transcriptoma foliar, aún falta información esencial para su manejo eficiente y sostenible. El objetivo de la presente revisión es proporcionar un panorama del conocimiento existente sobre los recursos genéticos de cedro rojo, con el fin de identificar alternativas basadas en herramientas biotecnológicas que permitan la comprensión de la especie y fundamenten el diseño e implementación de programas de mejoramiento genético y conservación a largo plazo.

Palabras clave: *Cedrela odorata* L., conservación, diversidad genética, genoma, marcadores moleculares, mejoramiento genético.

¹ Gálvez López, L., Vallejo Reyna, M. A., Méndez Espinoza, C., & López Upton, J. (2020). *Cedrela odorata* L.: oportunidades para su conservación y mejoramiento genético. *Revista Mexicana de Ciencias Forestales* 11(58). <https://doi.org/10.29298/rmcf.v11i58.622>

Cedrela odorata L.: OPPORTUNITIES FOR ITS CONSERVATION AND GENETIC IMPROVEMENT

Abstract

Cedrela odorata, commonly known as Spanish cedar, is a tropical forest species of high commercial value on the international timber market. Habitat fragmentation and illegal logging with dysgenic selection cause the genetic erosion of its natural populations, reducing its adaptive and productive potential. Therefore, it is increasingly susceptible to damage by abiotic and biotic factors such as *Hypsipyla grandella* Zeller, its main insect pest and the leading cause of low yield in commercial plantations. Spanish cedar is included in the lists of international organizations such as IUCN and CITES, and protected by NOM-059 in Mexico. However, because of the lack of convincing supporting evidence, its permanence in the latter is not guaranteed. Although some research has been done in the past evaluation of its genetic diversity in Central American populations, karyotyping, characterization of its plastome and leaf transcriptome essential information is still lacking to establish efficient and sustainable management strategies. The aim of this review is to present an overview of the current knowledge of genetic resources of *C. odorata* in order to identify biotechnological alternatives that would allow a better understanding of this species and substantiate the planning and implementation of tree breeding and long-term conservation programs.

Key words: *Cedrela odorata* L., conservation, genetic diversity, genome, molecular markers, tree breeding.

Introducción

Cedrela odorata L. (Meliácea), especie conocida como cedro rojo, es una de las principales riquezas de los bosques mesoamericanos (Cortés, 2001). En la república mexicana se distribuye de forma natural en las vertientes del Golfo de México y del Pacífico (Pennington y Sarukhán, 2005); así como en Centro, Sudamérica y las islas del Caribe (Patiño, 1997). El interés en este taxón se debe a las cualidades de su madera: como la alta durabilidad, color, jaspeado y aroma (Mader *et al.*, 2018); por lo que, se utiliza para la fabricación de muebles finos, instrumentos musicales, artesanías (Cintrón, 1990; Romo-Lozano *et al.*, 2017), elaboración de cajas de tabaco (Rodríguez, 2011), además de tener usos medicinales y aportar servicios ambientales (Romo-Lozano *et al.*, 2017). Por su relevancia económica se ha introducido en África, el sudeste asiático y al noreste de Australia (GBIF, 2018) ya que es la segunda especie arbórea tropical más comercializada del mundo, después de la caoba (*Swietenia macrophylla* King. [Meliaceae]) (Patiño, 1997), con un precio de exportación en Norteamérica de \$958 a \$977 USD por metro cúbico de madera aserrada (ITTO, 2019).

En consecuencia, el cedro rojo se ha sobreexplotado sin planes adecuados de manejo, lo que ha ocasionado el decremento y fragmentación de sus poblaciones naturales con la consiguiente restricción de flujo de genes entre las poblaciones (de la Torre, 2013), la pérdida de diversidad genética (Leshner-Gordillo *et al.*, 2018) y de genotipos de alta productividad que son potencialmente resistentes a estrés biótico y abiótico (Cavers *et al.*, 2003a). Además del daño antropogénico, presenta problemas de regeneración natural (Rodríguez, 2011); pérdida rápida de viabilidad de las semillas (García y Abdelnour, 2013); lento crecimiento, ya que tarda alrededor de 40 años en alcanzar la talla para su aprovechamiento maderable (Galván-Hernández *et al.*, 2018); y el ataque de *Hypsipyla grandella* Zeller (barrenador de las meliáceas).

Este último constituye el principal factor que impide el establecimiento y productividad de plantaciones forestales comerciales de la especie (López-Ochoa *et al.*, 2015; Santos *et al.*, 2015). Al respecto, Marquetti (1990) señala la presencia de un híbrido natural, resultado del cruzamiento entre *C. odorata* y *Cedrela cubensis* Bisse, para el cual describe mayor heterosis, crecimiento y resistencia al barrenador respecto

a sus progenitores. Sin embargo, *C. cubensis* se considera una sinonimia de *C. odorata* (The plant list, 2013) y faltan pruebas que corroboren la existencia de esa afirmación.

Es importante resaltar que el comportamiento de las plantaciones está sujeto a la calidad del sitio de su establecimiento, pues los factores bióticos, abióticos y su interacción con el manejo, afectan su capacidad productiva; por lo tanto, a partir del lugar de plantación, los árboles alcanzarán rendimientos en volumen mayores en poco tiempo; mientras que, en otros, será menor (Murillo-Brito *et al.*, 2017).

Debido a lo anterior, *C. odorata* se incluye en la lista roja de la Unión Internacional para la Conservación de la Naturaleza (IUCN) como especie vulnerable (IUCN, 2017), y en la lista de especies protegidas por la Convención sobre el Comercio Internacional de Especies Amenazadas de Fauna y Flora Silvestre (CITES) (CITES, 2017). En México, está catalogada como Sujeta a Protección Especial en la Norma Oficial Mexicana 059-SEMARNAT 2010 (SEMARNAT, 2010); sin embargo, se ha planteado que no cumple con los estándares establecidos que justifiquen su permanencia en dicha disposición (Ruiz-Jiménez *et al.*, 2018). Dada la actual situación problemática que enfrenta *C. odorata*, el objetivo de este trabajo es presentar los principales avances relacionados con la genética del cedro rojo para identificar oportunidades de desarrollo orientadas a su mejoramiento y conservación.

El genoma de *Cedrela odorata* L.

El cedro rojo pertenece a las meliáceas, grupo con un alto y variable número de cromosomas; por ejemplo, el $2n$ de especies de los géneros *Soymida*, *Toona* y *Cipadessa* es igual a 56, lo que sugiere un origen común y paleopoliploide (Styles y Vosa, 1971; Khosla y Styles, 1975). En comparación con otros taxa forestales como los géneros *Picea* y *Pinus* ($2n=24$) (Fuchs *et al.*, 1995; Nkongolo, 1996; Pavy *et al.*, 2008), algunos miembros de la familia Meliaceae poseen un número diploide más de dos veces mayor.

Styles y Vosa (1971) identificaron dos citotipos para *C. odorata*: $2n = 50$ y 56 , y sugieren que las poblaciones con menor número cromosómico se ubican en América Central y las Islas del Caribe, mientras que los de mayor número están en México y

Sudamérica. No obstante, Khosla y Styles (1975) descartaron la relación de los citotipos con su origen geográfico y estimaron una longitud cromosómica de 2.0 a 5.0 μm y 60 μm para el tamaño total de la cromatina, lo que constituyó la primera referencia sobre la dimensión del genoma.

Se desconoce la secuencia del genoma nuclear del cedro rojo. La única meliácea caracterizada es *Azadirachta indica* A. Juss. (nim), cuyo tamaño es de 364 Mpb con poco más de 20 000 genes (Krishnan *et al.*, 2012). Con esta referencia e información de taxones con filogenia cercana como Rutaceae (*Citrus sinensis* (L.) Osbeck, GenBank NC_023046-54), Malvaceae (*Theobroma cacao* L.; GenBank GCF_000208745), Brassicaceae (*Arabidopsis thaliana* (L.) Heynh.; GenBank NC_003070-76) o Vitaceae (*Vitis vinifera* L.; GenBank NC_012007-25) se esperaría que el tamaño del genoma de *C. odorata* sea de alrededor de 500 Mpb (Jaillon *et al.*, 2007; Argout *et al.*, 2011; Xu *et al.*, 2013). Sin embargo, hay que considerar que *A. indica* tiene un número cromosómico $2n = 28$; es decir, la mitad que el de cedro rojo y si se supone que su genoma es producto de una poliploidía (Khosla y Styles, 1975), entonces podría tener una dimensión dos veces mayor que la del nim.

Por otra parte, al describir el plastoma (genoma del cloroplasto) de *C. odorata* por medio de la plataforma de secuenciación *Illumina*, se determinó que su tamaño es de 158 558 pb (GenBank MG724915), en el que las regiones de copia sencilla larga (*Long Single Copy*, LSC) y corta (*Short Single Copy*, SSC) miden 86 390 pb y 18 380 pb, respectivamente; mientras que la región invertida (*Inverted Region*, IR) mide 26 894 pb. La lista de genes anotados, aquellos con funciones putativas identificadas, muestra un total de 112 genes diferentes de los cuales 78 codifican para proteínas, 30 para ARNt (ARN de transferencia) y 4 para ARNr (ARN ribosomal) (Mader *et al.*, 2018). Del cloroplasto también se ha caracterizado la región intergénica *rrn16-rrn23* con la que se construyó el vector *pCBL5* que permite su transformación directa por recombinación homóloga. Este tipo de transformación reduce el riesgo de contaminación por polen de transgenes a individuos silvestres, debido a que los plastidios son heredados por vía materna. Lo anterior tiene el potencial de generar variedades transgénicas resistentes a plagas, como el barrenador de las meliáceas (López-Ochoa *et al.*, 2015).

Con respecto al transcriptoma, mediante secuenciación de ARN de tejido foliar se identificaron 52 181 modelos de genes cuya longitud varía entre 200 pb y 37 635 pb (Finch *et al.*, 2019), lo que aporta otra evidencia relacionada a que cedro rojo tiene más del doble de los 20 000 genes identificados en el genoma de *A. indica* (Krishnan *et al.*, 2012). De esos genes putativos, entre 65 % y 70% presentan homología con aquellos caracterizados en genomas secuenciados de árboles como *Populus trichocarpa* Torr. & A. Gray ex Hook. Y *Quercus lobata* Née. Esto representa el primer estudio de la especie por medio de tecnología de secuenciación de última generación (Finch *et al.*, 2019).

Aportaciones de la biología molecular al estudio filogenético de cedro rojo

Las especies del género *Cedrela* son consideradas crípticas; es decir, la delimitación entre ellas es ambigua con base en rasgos morfológicos, de modo que la divergencia evolutiva del genoma no se manifiesta en el fenotipo (Cavers *et al.*, 2013). Por lo tanto, los marcadores moleculares son una alternativa para resolver ese tipo de problemas, ya que permiten caracterizar la estructura poblacional, la diversidad genética, relaciones filogenéticas, rutas migratorias y los procesos de hibridación, entre otros (Alía *et al.*, 2003).

Algunas de las principales aportaciones y avances obtenidos con la aplicación de herramientas moleculares en estudios evolutivos de *Cedrela* se abordan a continuación. El análisis de variación de las regiones intergénicas (ITS) 18S-26S y *trnS-trnG* y algunos genes del cloroplasto (ejem.: *psbB*, *psbT*, *psbN*) ha concluido que la divergencia de los géneros *Cedrela* y *Toona* ocurrió hace 46.6 a 50.6 millones de años. Asimismo, durante el Mioceno, Centroamérica fue el punto de origen de la diversificación de *Cedrela*, lo que resultó en la aparición de dos linajes que migraron a Norte y Sudamérica (Muellner *et al.*, 2010; Cavers *et al.*, 2013; Koecke *et al.*, 2013). También, se han reclasificado diversas especies del género: de ocho taxones a 17 y otras seis que aún no se aceptan (Cavers *et al.*, 2013; Koecke *et al.*, 2013; The plant list, 2013).

En otro estudio, por medio de marcadores moleculares de cloroplasto se identificaron cinco haplotipos en Centroamérica, la zona de contacto entre los dos grandes linajes del norte y sur de América (Cavers *et al.*, 2003b). Posteriormente, en el análisis se incluyeron poblaciones del Caribe y Sudamérica, en el cual mediante ITS y microsatélites nucleares y de cloroplasto se caracterizaron 22 haplotipos, lo que justificó la propuesta de reclasificación de algunos ecotipos, así como de nuevas especies de *Cedrela* (Cavers *et al.*, 2013). Con dichos antecedentes, se confirma la ocurrencia de al menos tres divergencias evolutivas que concuerdan con la colonización de Mesoamérica por la flora sudamericana. Las primeras dos sucedieron antes y después de la formación del Istmo de Panamá, y la tercera, a finales del Pleistoceno.

Diversidad genética

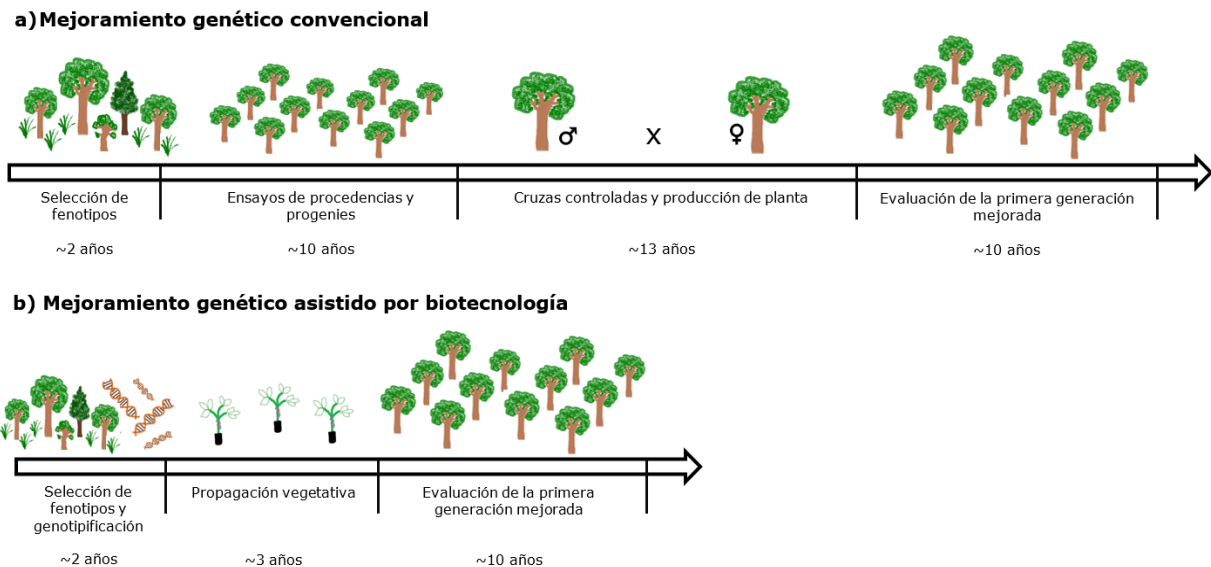
La diversidad genética de *C. odorata* se ha evaluado en poblaciones de Centro y Sudamérica con el uso de marcadores moleculares de tipo RAPD, AFLP, microsatélites y secuencias de cloroplastos, con lo que se evidenció una mayor diversidad interpoblacional que dentro de las poblaciones, la cual se relaciona con la precipitación media anual y diferencia entre poblaciones de regiones secas (norte de la Península de Yucatán) de aquellas de zonas húmedas (Costa Rica) (Gillies *et al.*, 1997; Cavers *et al.*, 2003a, 2003b; Navarro *et al.*, 2005; de la Torre *et al.*, 2008).

La diversidad genética es fundamental para la evolución y conservación a largo plazo, así como para los programas de mejoramiento y manejo de los recursos genéticos (Jump *et al.*, 2008). En particular, con los efectos del cambio climático resulta indispensable preservar el flujo génico y las variaciones en el material hereditario, ya que estas aumentan el potencial de adaptación a condiciones de estrés abiótico y biótico (Jiménez y Collada, 2000; Jump *et al.*, 2008).

Mejoramiento genético convencional y asistido por biología molecular

El mejoramiento genético tiene como objetivo desarrollar variedades de alto valor económico, ecológico o social e implica procesos reiterados de actividades como selección, cruzamientos dirigidos y evaluaciones genéticas (White *et al.*, 2007). Una de

las principales estrategias del mejoramiento genético clásico consiste en la selección fenotípica de los mejores individuos maduros, seguida del establecimiento de ensayos de procedencias y progenies para estimar los componentes genéticos y ambientales (Zobel y Talbert, 1988). Sin embargo, los métodos tradicionales son limitantes por los largos ciclos reproductivos de la mayoría de las especies forestales, lo que representa varios años e incluso décadas para la obtención de resultados, en función de su biología (Figura 2) (Grattapaglia, 2017).



a) Mejoramiento genético convencional. El tiempo estimado para un ciclo completo es de 35 años, debido a que en *C. odorata* la madurez sexual ocurre cerca de los 12-15 años a los que se adicionan de tres a siete más para realizar la evaluación temprana de los rasgos fenotípicos de interés. b) Mejoramiento genético asistido por biotecnología. Con el uso de herramientas como el cultivo de tejidos y marcadores moleculares, se estima una reducción aproximada de 50 % de tiempo con respecto al mejoramiento genético convencional, ya que los rasgos genotípicos asociados al fenotipo de interés pueden evaluarse en plántulas (Elaboración propia).

Figura 2. Representación esquemática de las etapas de un programa de mejoramiento genético de *Cedrela odorata* L.

Además, en especies forestales este tipo de mejoramiento genético se basa principalmente, en características de rasgos económicos para obtener mejor rendimiento en madera. La mayor parte de los estudios y, especialmente en especies tropicales no se basan en ideotipos específicos, que sean como una referencia al mejoramiento genético de estas, las cuales son importantes para la calidad y

heterogeneidad de los materiales establecidos y que, de manera adicional, puedan considerarse como unidades productoras de germoplasma forestal (Dickmann *et al.*, 1994; Solís-Guillén *et al.*, 2017). Esta selección se dificulta aún más, cuando los caracteres de interés son complejos; por ejemplo, las propiedades de la madera, resistencia a plagas y enfermedades, así como la tolerancia al estrés abiótico (Nehra *et al.*, 2005).

La biotecnología, en particular la biología molecular, ofrece oportunidades para genotipificar individuos superiores, realizar la selección asistida por marcadores (*Marker Assisted Selection*, MAS, por sus siglas en inglés), además de estudiar la organización del genoma y el funcionamiento de genes asociados a caracteres de importancia económica, entre otras. Una de las principales ventajas de las nuevas tecnologías consiste en reducir la duración de los ciclos de mejoramiento; por ejemplo, la MAS permitiría la selección temprana de caracteres que se expresan en la madurez (Grattapaglia, 2017); o bien, las predicciones con base en la genómica cuantitativa ahorrarían tiempo durante la etapa de evaluación (Figura 1) (Park *et al.*, 2016), tal como se estima para la conífera *Picea abies* (L.) Karst, cuyo ciclo de mejoramiento convencional requiere más de 28 años y que es factible reducirlo a menos de 10 por medio de estas herramientas (Lenz *et al.*, 2020).

En Latinoamérica, se han realizado diversos estudios orientados al mejoramiento genético tradicional de *C. odorata*. Así, en Costa Rica se evaluó el desempeño temprano y resistencia a *H. grandella* de 115 familias de 21 procedencias de México, Guatemala, Honduras, Costa Rica y Panamá en condiciones de asociación con cultivo de café (Navarro *et al.*, 2004). Las variables de interés fueron altura, diámetro, resistencia y número de rebrotes: todas fueron altamente significativas a nivel de procedencia. Los coeficientes de heredabilidad obtenidos en sentido estricto (h^2) para las familias dentro de las procedencias fueron de 0.12 y 0.20 para diámetro y altura, respectivamente. Además, los árboles de cedro rojo asociados al cultivo de café presentaron mejor crecimiento en diámetro y altura, así como más resistencia al barrenador de las meliáceas y mayor producción de brotes (Navarro *et al.*, 2004).

En México, el Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) es una de las pocas instituciones dedicadas a la domesticación del cedro rojo. Los avances incluyen el establecimiento de dos ensayos de progenies en los que evaluaron un total de 210 familias de 26 procedencias del Golfo de México y sureste del país (Sánchez *et al.*, 2003; Hernández-Máximo *et al.*, 2016). En ambos experimentos, se estimó la heredabilidad individual de las principales variables cuantitativas de importancia económica: altura, diámetro a la altura del pecho (DAP) y volumen (Cuadro 1); asimismo, la correlación genética estimada para los caracteres estudiados fue alta (>0.90). Por lo tanto, se concluyó que las variables de crecimiento tienen un control genético de moderado a alto.

Cuadro 1. Heredabilidad estimada para las principales variables dasométricas en *Cedrela odorata* L. en ensayos de progenie establecidos en el campo experimental “El Palmar”, Tezonapa, Veracruz, México.

Variable evaluada	Heredabilidad (h^2_i) en árboles de diferentes edades (años) al momento de la evaluación				Referencia
	3	5	7	11	
Altura (m)		0.65			
DAP (cm)		0.38			(Sánchez <i>et al.</i> , 2003)
Volumen (dm ³)		0.54			
Índice de conicidad		0.38			
Altura (m)	0.16	0.22	0.25	0.22	(Hernández-Máximo <i>et al.</i> , 2016)
DAP (cm)	0.16	0.21	0.27	0.27	
Volumen (dm ³)	0.12	0.18	0.27	0.29	

h^2_i = Heredabilidad individual; DAP = Diámetro a la altura del pecho.

El proceso desde la selección de fenotipos para el establecimiento de los ensayos de progenie hasta la obtención de dichos parámetros genéticos requirió más de 20 años, a partir de los cuales se seleccionaron los 40 mejores individuos y se establecieron en un ensayo clonal en el que se calculó una ganancia genética de 82 % para el volumen, con base en una intensidad de selección de 10 % (Sampayo-Maldonado *et al.*, 2019).

Conclusiones

A diferencia de otras especies forestales como aquellas de los géneros *Populus*, *Pinus* o *Eucalyptus*, los estudios genéticos sobre cedro rojo son incipientes. En ellos, algunas herramientas biotecnológicas han permitido determinar parte de sus características genéticas. Las investigaciones se han enfocado, principalmente, a poblaciones de la región comprendida entre Veracruz, México y el norte de Sudamérica, mismas que han evidenciado la estructura poblacional y su relación con las condiciones climáticas prevalecientes en su área de distribución, lo que sugiere que su acervo genético aún posee un importante potencial de adaptación.

No obstante, la fragmentación de sus poblaciones impide el flujo génico y la tala clandestina pone en riesgo este recurso, que, a pesar de su valor en la región Neotropical, en México no se ha logrado que coadyuve al desarrollo económico de las comunidades rurales aledañas a su área de aprovechamiento. Por ello, es imprescindible llevar a cabo estudios intensivos y extensivos que conlleven a la caracterización del patrimonio genético de *C. odorata* a lo largo de su área de distribución en el país; así como investigaciones enfocadas a la fisiología de su respuesta al estrés biótico y abiótico y a la generación de tecnologías para su mejoramiento genético y aprovechamiento sostenible.

Con el apogeo de la era posgenómica y su relación con la bioquímica, biotecnología, biología molecular, bioinformática, bioestadística, entre otras, se ha revolucionado la forma de estudiar la interacción genotipo-ambiente y se dispone de nuevas alternativas para el mejoramiento y conservación de las especies forestales, lo que no implica el reemplazo de las técnicas convencionales, sino su complementación para hacer más eficientes dichos procesos y proteger la diversidad genética.

CAPÍTULO I. GENERACIÓN Y ANOTACIÓN DEL TRANSCRIPTOMA DE SEMILLAS DE *Cedrela odorata* L. DURANTE LA GERMINACIÓN

1.1 RESUMEN

Cedrela odorata L. es una especie forestal tropical maderable intensamente explotada. Una alternativa para proteger la especie y su diversidad genética es a través de la conservación *ex-situ* de las semillas en bancos de germoplasma. Sin embargo, las semillas de esta especie se caracterizan por la pérdida rápida de su viabilidad lo que impide su almacenamiento a largo plazo. Por ello, es importante entender los mecanismos fisiológicos y moleculares que permiten mantener la viabilidad durante el almacenamiento y que conducen a una germinación vigorosa. Actualmente estos mecanismos pueden ser estudiados a través de tecnologías “NGS”, que incluye el estudio del transcriptoma y permite detectar patrones de expresión génica presentes en un tejido y condición específica. El objetivo fue generar un transcriptoma de *novo* a través de tres fases importantes del proceso germinativo de semillas de *C. odorata* usando: semilla madura seca, hidratada al 95% y germinada. Se extrajo el ARN en cada condición, se generaron bibliotecas de ADNc y se secuenciaron de acuerdo con el protocolo estándar de Illumina. Se realizó un análisis de calidad de los datos de secuenciación y después se realizó el ensamble del transcriptoma de *novo* y su anotación funcional. El transcriptoma generado tuvo un total de 95,739 transcritos y 52,490 genes únicos. El % de Guanina-Citosina fue de 42.62 y el valor N50 de 1,668 bases. El alineamiento por *Blastx* generó 57,148 secuencias nucleotídicas únicas y el *Blastp*, 38,157. Se localizaron 22 grupos funcionales por eggNOG donde destacan las modificaciones postraduccionales, recambio de proteínas, chaperonas y transcripción. En las categorías, PB, CC y FM de la clasificación GO predominaron los términos: regulación de la transcripción-plantilla de ADN, núcleo y unión de ATP, respectivamente. En este estudio se generó el primer transcriptoma de *novo* de semillas de *C. odorata* analizando el proceso germinativo. Los resultados permiten conocer los mecanismos moleculares presentes en dichas etapas y facilitan posteriores investigaciones enfocadas a conservar las semillas en el almacenamiento.

Palabras clave: Anotación funcional, ensamble de transcriptoma, germinación, secuenciación.

GENERATION AND ANNOTATION OF THE *Cedrela odorata* L SEED TRANSCRIPTOME OF. DURING GERMINATION

1.2 ABSTRACT

Cedrela odorata L. is an intensely exploited tropical forest species. One alternative to protect the species and its genetic diversity is through the *ex-situ* conservation of seeds in germplasm banks. However, the seeds of this species are characterized by the rapid loss of their viability, which prevents their long-term storage. Therefore, it is important to understand the physiological and molecular mechanisms that maintain viability during storage and that lead to vigorous germination. Currently these mechanisms can be studied through “NGS” technologies, which include the study of the transcriptome and allows the detection of gene expression patterns present in a specific tissue and condition. The objective was to generate a *de novo* transcriptome through three important phases of the germination process of *C. odorata* seeds using: dry mature seed, 95% hydrated and germinated. RNA was extracted in each condition, cDNA libraries were generated and sequenced according to standard Illumina protocol. Quality analysis of the sequencing data was performed followed by *de novo* transcriptome assembly and functional annotation. The generated transcriptome had a total of 95,739 transcripts and 52,490 unique genes. The % Guanine-Cytosine was 42.62 and the N50 value was 1,668 bases. *Blastx* alignment generated 57,148 unique nucleotide sequences and *Blastp*, 38,157. 22 functional groups were located by eggNOG, where post-translational modifications, protein turnover, chaperones and transcription stood out the most. In the categories, PB, CC and FM of the GO classification, prevailing terms were: regulation of transcription-DNA template, nucleus and ATP binding, respectively. In this study, the first *de novo* transcriptome of *C. odorata* seeds was generated by analyzing the germination process. The results allow us to know the molecular mechanisms present in these stages and facilitate further research focused on preservation seeds in storage.

Key words: Functional annotation, transcriptome assembly, germination, sequencing.

1.3 INTRODUCCIÓN

La semilla se define como la estructura que contiene al embrión, cuyo material genético garantiza el desarrollo del próximo ciclo de vida de las plantas y sus procesos evolutivos (Sreenivasulu y Wobus, 2013). La alteración de los ecosistemas con la consecuente pérdida de biodiversidad vegetal y posiblemente la diversidad genética, hace necesaria la búsqueda de estrategias de conservación de las especies tanto en forma *in-situ* como *ex-situ* (Werden *et al.*, 2020). *Cedrela odorata* L. es una especie forestal tropical maderable que está siendo explotada en forma continua (SEMARNAT, 2010; CITES, 2017; IUCN, 2017) por lo que el desarrollo de estrategias de conservación *ex-situ* es importante para proteger la especie. Almacenar las semillas en bancos de germoplasma es una forma efectiva de conservar la biodiversidad (Whitehouse *et al.*, 2020); no obstante, las semillas de *C. odorata* se caracterizan por la rápida pérdida de su viabilidad, lo que impide su conservación a largo plazo y uso para producir plantas (Sautu *et al.*, 2006; García y Abdelnour, 2013). Por lo tanto, es importante entender los mecanismos fisiológicos y moleculares que permiten mantener la viabilidad durante el almacenamiento y que conducen a una germinación vigorosa.

Una de las formas más modernas para analizar los mecanismos moleculares en organismos biológicos es a través de herramientas genómicas novedosas como la tecnología de secuenciación de próxima generación (*Next Generation Sequencing*, NGS, por sus siglas en inglés) (Aguilar-Bultet y Falquet, 2015). NGS es un grupo de tecnologías proyectadas para secuenciar cantidades masivas y en paralelo de segmentos de ADN utilizando una menor cantidad de tiempo y costo por base que las técnicas anteriores (Rubio *et al.*, 2020). La secuenciación puede ser de lectura única (*single-end*) o lectura pareada (*paired-end*) y se almacenan en un archivo de texto plano cuyo formato es llamado *fastq* (Cortese *et al.*, 2021). La tecnología NGS tiene la ventaja de generar grandes volúmenes de datos de forma rápida y económica, en el que podría llegarse a millones o billones de lecturas (*reads*). Para ello es importante el empleo de diversas herramientas bioinformáticas que permitan manipular de forma eficaz cantidades enormes de datos para su análisis (Aguilar-Bultet y Falquet, 2015).

Las aplicaciones tecnológicas NGS se extienden a distintos campos como lo es la genómica, transcriptómica, epigenética y metagenómica (Aguilar-Bultet y Falquet, 2015). Entre éstas, la transcriptómica (también llamada *RNA-Seq*) permite la secuenciación masiva y profunda de ARNs que entre otras cosas permite detectar patrones de expresión génica y con ello, estudiar los mecanismos bioquímicos y moleculares de diversos procesos fisiológicos o de desarrollo (Rodríguez *et al.*, 2014). Esta tecnología se genera a través de lecturas de secuencia de ADN, que proceden de la molécula del ARN completa y determina la secuencia primaria y abundancia relativa de cada ARN (Caicedo-Montoya *et al.*, 2019). La técnica *RNA-Seq* puede realizarse a través de distintas plataformas como SOLID, 454, Ion Torrent, Illumina (Rodríguez *et al.*, 2014), además de PacBio, Oxford Nanopore Technologies y GS111 (Foux *et al.*, 2021).

Por lo anterior, la transcriptómica es uno de los métodos más adecuados para conocer la expresión de genes que están involucrados en el proceso de germinación de *C. odorata*. Posterior a la secuenciación es necesario ensamblar el transcriptoma, para lo cual existen dos alternativas computacionales: las lecturas resultantes se pueden alinear con un genoma o transcriptoma de referencia cuando existen datos de secuenciación previos (Caicedo-Montoya *et al.*, 2019); o el el transcriptoma puede ensamblarse de *novo*, cuando se carece de genomas o transcriptomas de referencia, para lo cual se usan las lecturas para ensamblar las transcripciones directamente (Grabherr *et al.*, 2011; Caicedo-Montoya *et al.*, 2019). Para la generación de dicho ensamble existen diversas aplicaciones bioinformáticas, por ejemplo: Trinity, Oases, Trans-ABYSS, SOAPdenovo-Trans, IDBA-Tran, Bridger, BinPacker, Shannon, SPAdes-sc y SPAdes-rna. Entre estos uno de los algoritmos más novedosos ha sido Trinity (Hölzer y Marz, 2019), que resulta un método eficiente, con un paquete de software robusto que consta de tres módulos para el análisis: *Inchworm*, *Chrysalis* y *Butterfly*, los cuales son aplicados de forma sucesiva para procesar volúmenes grandes de lecturas de *RNASeq* (Grabherr *et al.*, 2011).

La secuenciación del genoma nuclear de *C. odorata* aún no se ha generado, por ello el objetivo que se busca en este capítulo es generar un transcriptoma de semilla ensamblado de *novo* utilizando tres fases importantes del proceso germinativo de

semillas de *C. odorata*: semilla seca (quiescente), hidratada al 95% y germinada. Los resultados permitirán conocer una descripción general del transcriptoma de la semilla, incluyendo el número de transcritos y genes involucrados en dicho proceso, además de las principales clasificaciones funcionales. Todo ello con el propósito de mejorar nuestro entendimiento respecto a los mecanismos fisiológicos y moleculares involucrados en el proceso de la germinación de semillas de *C. odorata*.

1.4 MATERIALES Y MÉTODOS

1.4.1 Material vegetal y tratamientos

Se recolectaron frutos maduros pero indehiscentes de *C. odorata* en el año 2017. La recolecta se realizó de un solo árbol tomado al azar procedente de una plantación en el sitio experimental “El Palmar” del Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP), en Veracruz, México con coordenadas 18° 32´L. N., 96° 47´L. Las cápsulas se secaron en media sombra y a temperatura ambiente durante aproximadamente tres días hasta que se liberaron las semillas, después se limpiaron y se transportaron al laboratorio de Biotecnología Forestal del Centro Nacional de Investigación Disciplinaria en Conservación y Mejoramiento de Especies Forestales (CENID-COMEF-INIFAP) donde se almacenaron en recipientes plásticos y se mantuvieron en refrigeración (4-10 °C), hasta su uso. Los ensayos realizados según las reglas internacionales de la Asociación Internacional de Análisis de Semillas (ISTA, 2016) mostraron que las semillas tenían un 95 % de germinación y un contenido de agua (WC) de 6.9 %.

Los tratamientos consistieron en utilizar tres fases en la germinación de semillas: semilla madura seca (S), semilla hidratada al 95 % (E) y semilla germinada (G). Las semillas secas (S) se congelaron en nitrógeno líquido y se almacenaron a -80 °C. La hidratación al 95% (E) consistió en tomar el peso inicial de cada semilla y colocarlas de forma individual en tubos de 1.5 ml que contenían un hisopo de algodón humedecido con agua destilada para hidratar la semilla. La fórmula utilizada para determinar el porcentaje del contenido de agua fue: $WC = ((\text{peso húmedo} - \text{peso inicial}) / \text{peso inicial}) \times 100$ (Sánchez-Coronado *et al.*, 2007). Una vez que las semillas alcanzaron el WC mencionado anteriormente, se congelaron en nitrógeno líquido y se almacenaron a -80 °C.

Para la germinación (G) las semillas se colocaron en cajas germinadoras de plástico que contenían papel filtro humedecido con agua destilada. Las cajas se incubaron a 28 °C en un horno marca Binder (BD15-UL) hasta obtener la germinación *sensu stricto*, es decir, cuando se presentó la protrusión de la radícula y alcanzó

aproximadamente 2 mm de longitud (Barone y Luna, 2016). Luego de la preparación de cada tratamiento, las muestras fueron sumergidas en nitrógeno líquido y posteriormente almacenadas a -80°C, hasta su procesamiento.

1.4.2 Extracción del ARN Total

El ARN total se obtuvo usando el método de Wang *et al.* (2008) con algunas modificaciones. Se utilizaron 150 mg de semilla (aproximadamente 9 semillas) congelada, por triplicado para cada condición. La semilla se maceró con un mortero usando nitrógeno líquido y recuperó en 800µl del buffer de extracción [2M NaCl, 25 mM EDTA (ácido etilendiaminotetraacético), 200 mM Tris (pH 8.0), 20 mM de Borato de Sodio, 2% de PVP40 (w/v) (Polivinilpirrolidona), 2% de CTAB (w/v) (bromuro trimetil amonio de cetilo) y 2% b-mercaptoethanol (v/v)] precalentado a 65°C en un Termomixer compact marca Sigma-Aldrich, durante 10 minutos. Después, se agregaron 600 µl de cloroformo/alcohol isoamílico (24:1) y se mezcló con ayuda de un vórtex, seguido de una centrifugación a 15,000 g durante 15 minutos a 4°C. La fase acuosa se transfirió a otro tubo al cual se le agregó un volumen de fenol/cloroformo/alcohol isoamílico (25:24:1) y se centrifugó a 12,000 g por 10 minutos, repitiendo el paso dos veces más. Posteriormente se recuperó la fase acuosa en un nuevo tubo y se agregó un volumen de cloroformo/ alcohol isoamílico (24:1) centrifugando a 12,000 g por 10 minutos a 4°C. La fase acuosa se recuperó nuevamente en un tubo nuevo de 1.5 ml y se agregó 0.5 del volumen obtenido de LiCl (cloruro de litio) a 8M. La muestra se incubó toda la noche a -20°C. Pasando este tiempo se centrifugó a 15,000 g durante 15 minutos a 4°C. Se decantó el sobrenadante cuidadosamente para no llevarse la pastilla y se enjuagó con etanol al 75%, seguido de centrifugación a 12,000 g por 5 minutos (este paso se repitió tres veces). Finalmente, la pastilla pegada al tubo se secó al aire y se resuspendió en 30 µL de agua DEPC (Dietilpirocarbonato).

Después de obtener el ARN, se realizó un tratamiento de purificación con el objetivo de eliminar restos de ADN de la muestra, para ello se utilizó el kit de Dnasa I y RNaseOUT (Invitrogen ®) siguiendo las indicaciones del fabricante.

La calidad y la cantidad del ARN purificado se determinaron en función de las proporciones de absorbancia a 260/280 nm (A260/A280) y 260/230 nm (A260/A230) utilizando un Nanodrop (Thermo Scientific). La integridad del ARN se evaluó mediante electroforesis en geles de agarosa al 1.5 % con tampón TAE 1X (40 mM Tris, 20 Mm ácido acético, 1mM EDTA), para el cual se utilizó un voltaje de 80v durante 45 minutos.

1.4.3 Preparación de bibliotecas de ADNc y secuenciación

La secuenciación masiva de ARN se llevó a cabo en la Unidad Universitaria de Secuenciación Masiva y Bioinformática (UUSMB) del Instituto de Biotecnología de la UNAM empleando un secuenciador Genome Analyzer Iix System de acuerdo con el protocolo estándar de Illumina (TruSeq RNA Library Prep Kit v2), empleando tres replicas biológicas para cada tratamiento: S, E y G, obteniendo un total de 9 bibliotecas de ADNc (ADN complementario). Se obtuvieron lecturas apareadas (*paired-reads ends*) de 150 pb.

1.4.4 Análisis de calidad de la secuencia

Después de finalizar el proceso, se obtuvieron archivos de texto plano en formato *fastq* (Deorowicz y Grabowski, 2011) correspondiente a las tres condiciones evaluadas (S, E y G) cada una con sus con tres réplicas. En total se tuvieron 18 archivos que conforman las lecturas *Read 1* (sentido) y *Read 2* (anti-sentido) de las 9 muestras. Todos los archivos fueron analizados con el programa bioinformático FastQC v.0.11.8 (Babraham Institute, 2020), el cual entrega diversos parámetros estadísticos que permiten conocer la calidad de las secuencias, entre ellos se encuentra el valor *Phred*, que refleja la probabilidad de que la base sea correcta (Hernández *et al.*, 2020).

1.4.5 Ensamble de *novo* del transcriptoma

Posterior al análisis de calidad se ejecutó el ensamble del transcriptoma de *novo* con el software Trinity-v2.9.1 (Grabherr *et al.*, 2011) en un ordenador con dos procesadores de 20 núcleos cada uno (Intel Xeon E5-2600v4) con 64GB de memoria

RAM. Para ello se usaron los 18 archivos *fastq* en *paired-end* (lecturas pareadas) de las tres muestras: S, E y G. El comando y opciones utilizados fueron los siguientes:

```
$Trinity --seqType fq --max_memory 60G --bflyCalculateCPU --output \  
/direccion/de/los/archivos/de/salida --CPU 8 --full_cleanup --verbose \  
--left \ FC1_R1.fastq,FC2_R1.fastq,FC3_R1.fastq,FC4_R1.fastq,FC5_R1.fastq,FC6_R1.fastq,FC7_R1.fastq,FC8_R1.fastq,FC9_R1.fastq \  
--right \ FC1_R2.fastq,FC2_R2.fastq,FC3_R2.fastq,FC4_R2.fastq,FC5_R2.fastq,FC6_R2.fastq,FC7_R2.fastq,FC8_R2.fastq,FC9_R2.fastq
```

1.4.6 Anotación del transcriptoma

Para la anotación funcional del transcriptoma se utilizó el software trinitate v.3.2.1; utilizando las especies predeterminadas y para la predicción y codificación de las secuencias se utilizó el software transdecoder v.5.5.0. Los transcritos fueron anotados por similitud de secuencia usando *BLASTx* y *BLASTp* v.2.11.0 en bases de datos públicas: *Pfam* (Protein family), *Swiss-Prot*, *GO* (Gene ontology), *KEEG* (Kyoto Encyclopedia of Genes and Genomes) y *eggNOG* (database of biological information hosted by the EMBL) con un valor de corte de 1×10^{-3} . Las anotaciones generadas se cargaron en una hoja de cálculo en formato *.xls el cual se utilizó para la generación del resumen y gráficos de la anotación utilizando un lenguaje de programación R v.4.0.5, a través de la interfaz RStudio v.1.3.1093.

1.5 RESULTADOS

1.5.1 Extracción del ARN

El resultado de la extracción posterior a la limpieza del ARN mostró una concentración promedió de 421.14 ng/ μ l; una pureza cuantificada con la relación de absorbancia 260nm/280nm de 2.07 y la relación 260nm/230nm de 2.06 (**Cuadro 1.1**).

Cuadro 1.1. Lecturas promedio tomadas con un Nanodrop para evaluar la cantidad y calidad del ARN total en semillas de *Cedrela odorata* L.

Muestra	Concentración (ng/ μ l)	A260/280	A260/230
S	370	2.07	2.02
E	485.17	2.08	2.10
G	408.27	2.06	2.06

S=Semilla seca, E= Semilla Embebida al 95%, G= Semilla germinada

El ARN de las muestras se obtuvo de alta calidad. En el gel se puede visualizar la integridad de las bandas obtenidas. Las flechas muestran las bandas de ARN ribosomal (ARNr) 28S y 18S (Figura 1.1). Por lo tanto, la calidad e integridad de las muestras fueron adecuadas para generar las bibliotecas de ADNc y posteriormente la secuenciación.

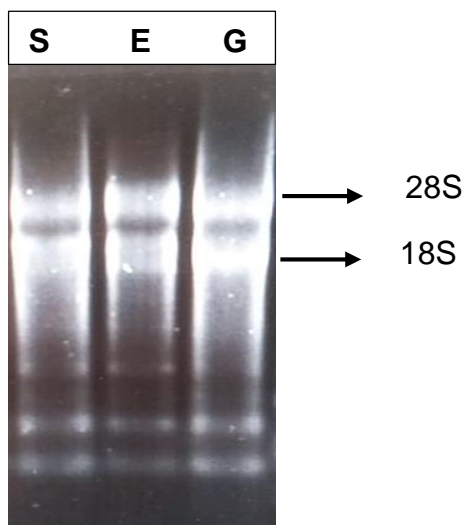


Figura 1.1. Electroforesis en gel de agarosa al 1.5% del ARN total extraído de las semillas S=Seca, E=embebida al 95% y G= germinada. Las dos distintas bandas de ARNr muestran la integridad del ARN.

1.5.2 Análisis de calidad de la secuencia

El programa FastQC permitió analizar once parámetros de las secuencias que facilitan comprobar la calidad de los datos resultantes de un sistema de secuenciación de Illumina. En los siguientes apartados se muestran los resultados de estos once parámetros para los archivos *fastq* que corresponden a las lecturas, *Read 1* y *Read 2* de la muestra 1 de semilla seca (FC1_R1.fq y FC1_R2.fq) y muestran la interpretación de los resultados. En el Cuadro A1 se muestra el resumen de los resultados de los 18 archivos *fastq* correspondiente a las tres condiciones evaluadas (S, E y G) cada una con sus tres réplicas y cada réplica con sus respectivas lecturas (*Read 1* y *Read 2*), a través de semáforos que indican la calidad. El verde significa que el parámetro evaluado es de buena calidad, el naranja indica una señal de advertencia y el rojo una mala calidad.

1.5.2.1 Estadística básica

En la Figura 1.2 se muestra que el número total lecturas (*reads*) en esta muestra, fue de 12,952,485, con una longitud de 75pb y un porcentaje de GC de 49%. Este porcentaje se considera aceptable cuando supera el 45%. Se aconseja que la longitud de las lecturas sea la misma para evitar complicaciones en el resto de las comprobaciones de calidad (Sánchez, 2015). En el Cuadro A2 se muestra la información de los 18 archivos analizados.



Read 1		Read 2	
 Basic Statistics		 Basic Statistics	
Measure	Value	Measure	Value
Filename	FC1_R1.fastq	Filename	FC1_R2.fastq
File type	Conventional base calls	File type	Conventional base calls
Encoding	Sanger / Illumina 1.9	Encoding	Sanger / Illumina 1.9
Total Sequences	12952485	Total Sequences	12952485
Sequences flagged as poor quality	0	Sequences flagged as poor quality	0
Sequence length	75	Sequence length	75
%GC	49	%GC	50

Figura 1.2. Estadística básica del análisis de calidad con FastQC. Se muestra el *read 1* y *2* de los archivos *fastq* que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L.

1.5.2.2 Calidad de secuencia por base

La Figura 1.3 permite visualizar la calidad de la secuencia por base (nucleótidos) de los *reads*. Generalmente en el análisis de calidad este gráfico es el que recibe mayor importancia (Sánchez, 2015). Se debe tomar en cuenta que la parte roja corresponde a la zona de menor calidad (Q30 menor a 20), la naranja es una calidad regular (Q30 menor a 28 y mayor a 20) y la verde muestra una buena calidad (Q30 mayor a 28) (Cortese *et al.*, 2021). En este caso, la mayoría de las muestras se ubican en la zona verde, por arriba del valor 30, motivo por el que el semáforo se pinta de verde lo que indica que las lecturas tienen una buena calidad.

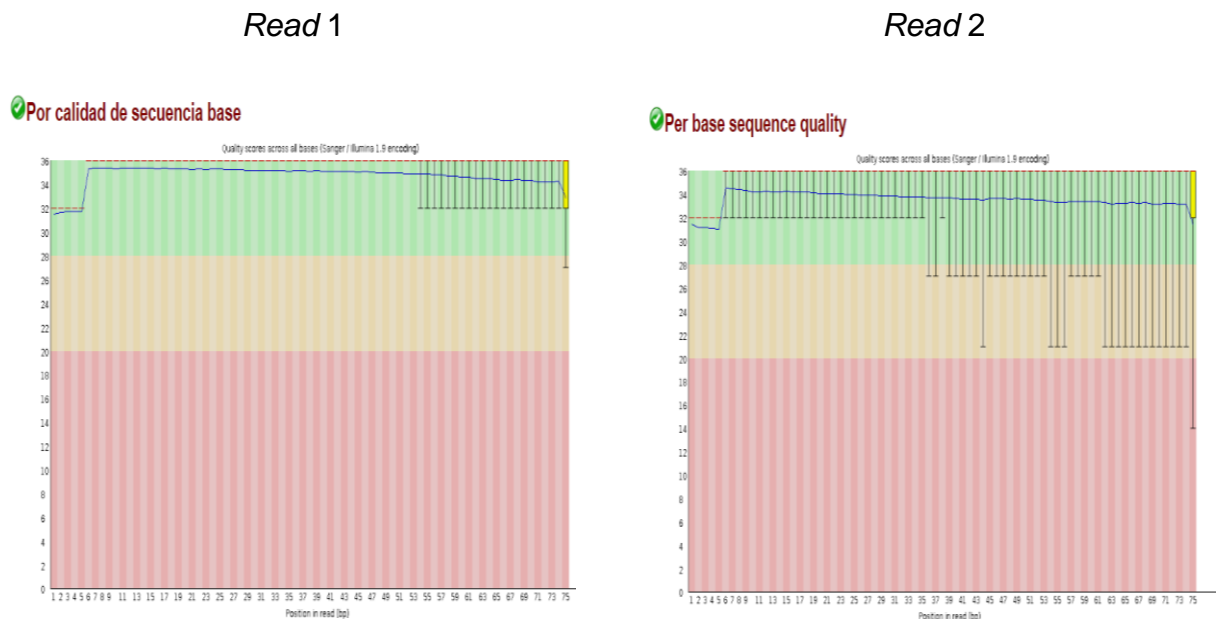


Figura 1.3. Calidad de secuencia por base para el *read 1* y *2* que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. Se presenta un gráfico de cajas para cada base. La línea roja representa la mediana y la azul el valor medio de las bases. El eje X muestra la posición de cada base en una secuencia y el eje Y representa el valor de calidad de cada base (valor Q30), en donde a mayor valor mejor es la calidad.

1.5.2.3 Calidad de secuencia por mosaico

La Figura 1.4 permite visualizar las celdas de flujo, es decir, los canales donde se pone la secuencia. Cuando la calidad es baja, se muestran marcas y cuadros rojos en

el fondo (García, 2018). Por lo tanto, al obtener un azul homogéneo en ambos gráficos indica que no hay pérdida en la calidad.

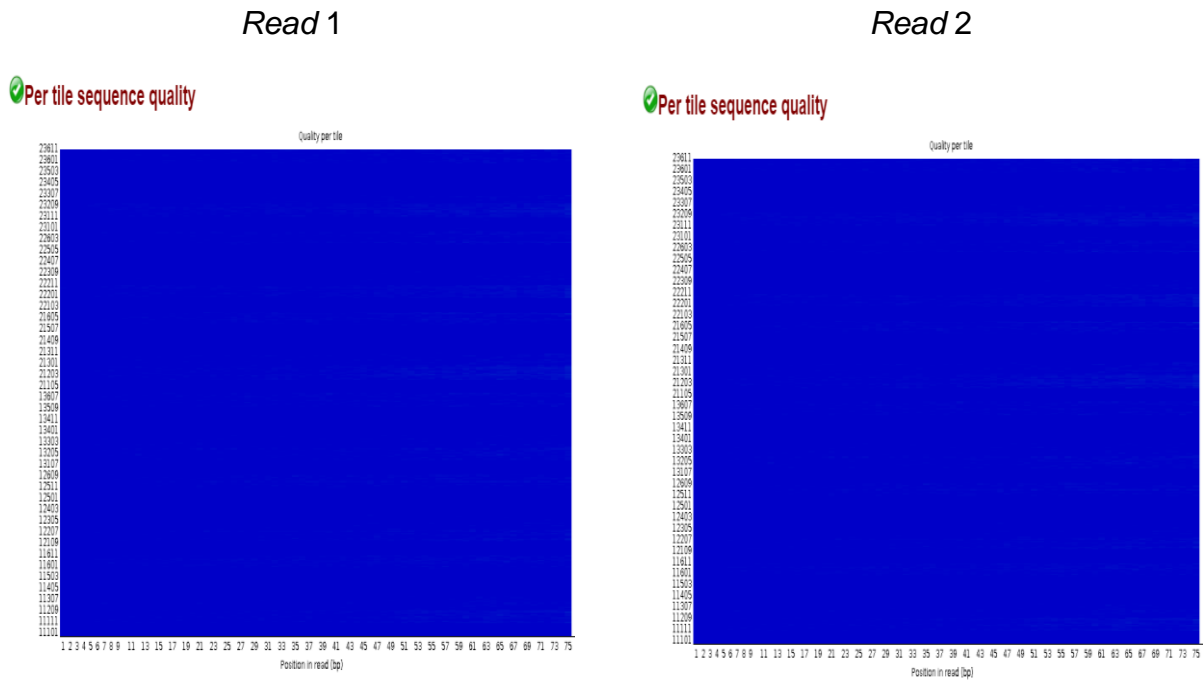


Figura 1.4. Mapa de calor de la calidad de la secuencia que muestra las celdas de flujo para el *read* 1 y 2 que corresponde a la muestra 1 de las semillas secas de *Cedreia odorata* L. En el eje de las X se muestran la posición de los *read* y en el eje de las Y las secuencias asociadas a los *reads*.

1.5.2.4 Niveles de calidad por secuencia

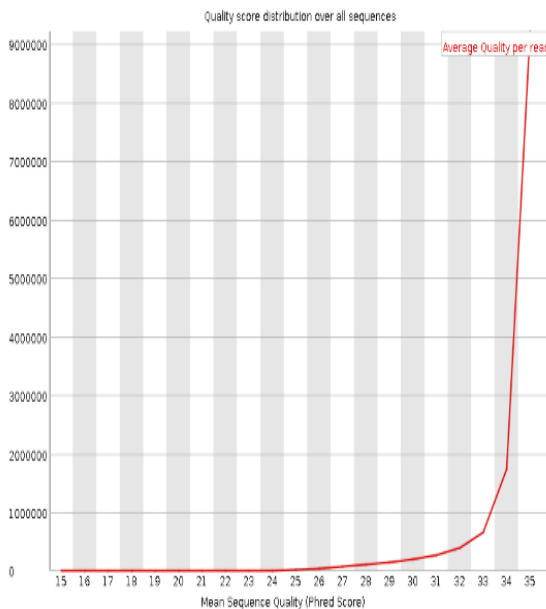
La Figura 1.5 muestra los niveles de calidad por secuencia. En el caso del *read* 1 se muestra que entre 2 y 9 millones de lecturas presentan una calidad media de entre 34 y 35; por tanto, esta calidad se toma como alta considerando que la calidad de secuencia por base mostrada en la Figura 1.3 toma como calidad alta cuando las bases están por arriba de 30.

Se indica una señal de advertencia cuando la calidad media observada con mayor frecuencia es menor de 27, esto equivale a una tasa de error de 0.2 % .Se señala fracaso cuando la calidad media observada con mayor frecuencia se encuentra por debajo de 20 lo que equivaldría a una tasa de error del 1% (Sánchez, 2015).

Read 1

Read 2

✔ Per sequence quality scores



✔ Per sequence quality scores

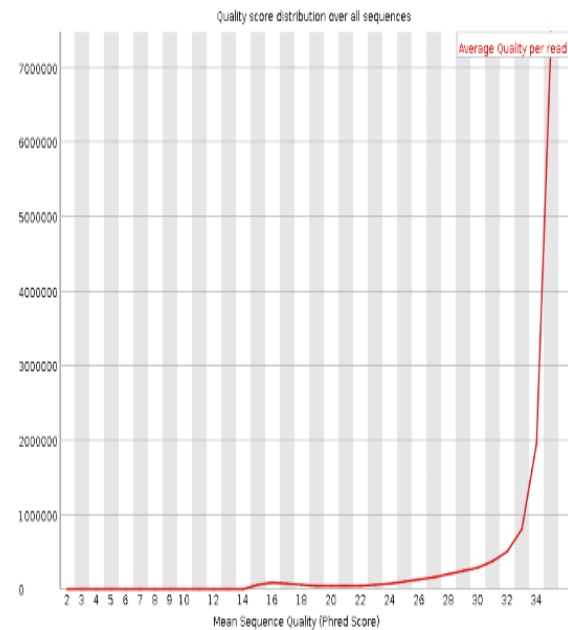


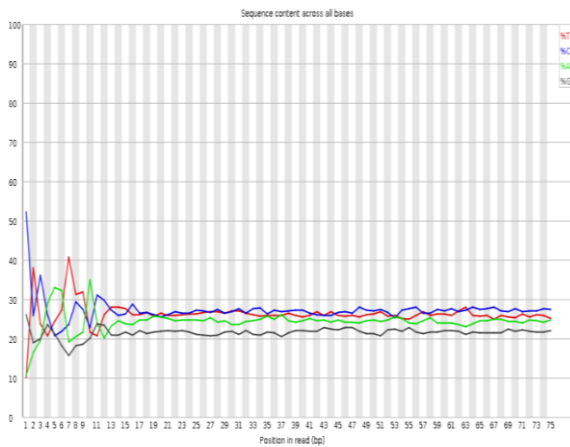
Figura 1.5. Estimador de calidad de la secuencia para el *read* 1 y 2 que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. En el eje de las X se muestra la calidad media de las secuencias y en el eje de las Y el número de secuencias que corresponden a esa media.

1.5.2.5 Contenido de secuencia por base

La Figura 1.6 representa la proporción de nucleótidos (A, T, G, C) en cada base de los *reads*. Lo ideal es que no haya mucha diferencia entre las bases de una secuencia de ejecución, por lo tanto, las líneas deberían ser paralelas (Babraham Institute, 2020). Se indica una señal de advertencia cuando la diferencia entre A y T, o C y G, es mayor al 10% para cualquier posición, y señalará un fracaso cuando dicha diferencia sea mayor que 20% en cualquier posición (Sánchez, 2015). En este caso *en el Read 1* se obtiene mala calidad y el *Read 2* de calidad media en los primeros *reads* (1-11 aproximadamente.) *Read 1* rebasa diferencias del 20% en algunas lecturas y *Read 2* rebasa a más del 10%, motivo por el cual se muestra los semáforos de señalamiento.

Read 1

Per base sequence content



Read 2

Per base sequence content

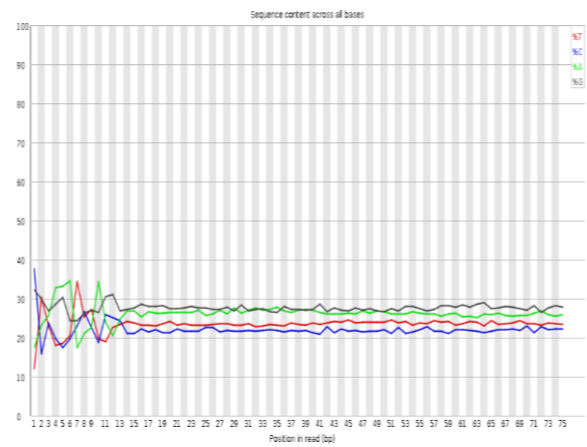


Figura 1.6. Proporción del contenido de nucleótidos por base que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. En el eje de las X se muestra la posición del *read* y en el eje de las Y el porcentaje de los nucleótidos en cada base de los *reads*.

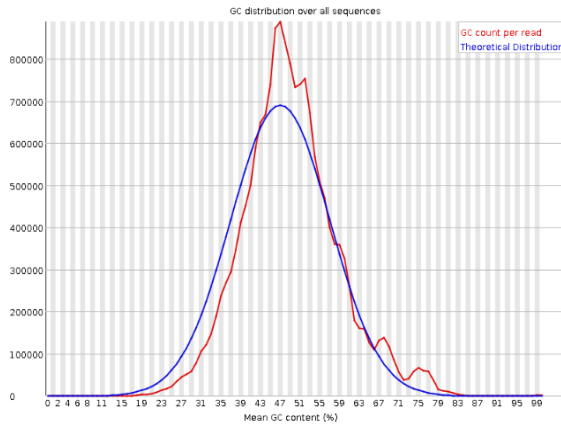
1.5.2.6 Contenido de GC por secuencia

La Figura 1.7 muestra el porcentaje de GC en los *reads*. Lo ideal es observar una distribución normal como se marca en la línea azul. No obstante que en estos resultados marcados con la línea roja se observa una tendencia a la normalidad, sin embargo, se distinguen unos picos que se salen de esta línea de distribución normal, lo que ocasiona la señal de advertencia. Dicha señal se da cuando la suma de las desviaciones de la distribución normal representa más del 15 % de las *reads* y señalará un fracaso si dicha suma representa más del 30 % (Sánchez, 2015).

Read 1

Read 2

👉 Contenido de GC por secuencia



👉 Contenido de GC por secuencia

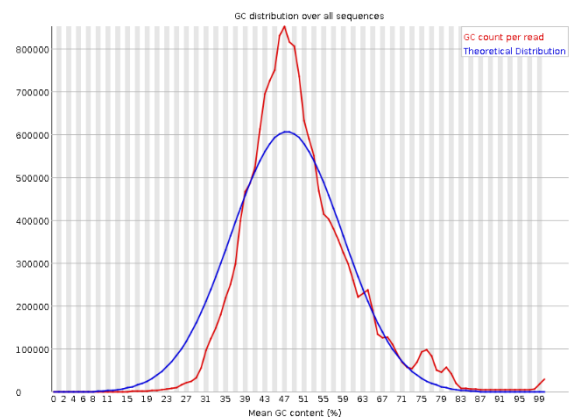


Figura 1.7. Proporción de GC en los *reads* (línea roja) comparándose con una distribución normal teórica (línea azul) que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. El eje X muestra el porcentaje medio del contenido de GC y el eje de las Y el número de *reads*.

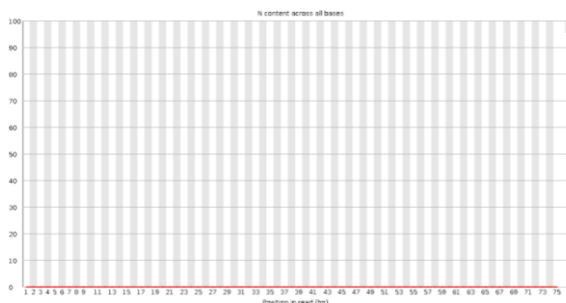
1.5.2.7 Contenido de N por base

La Figura 1.8 representa la proporción de N (Nucleótido desconocido) en la muestra. Cuando el secuenciador no tiene certeza de que una base sea correcta incorpora una N (Sánchez, 2015). En estos resultados no se encontraron nucleótidos desconocidos.

Read 1

Read 2

👉 Per base N content



👉 Per base N content

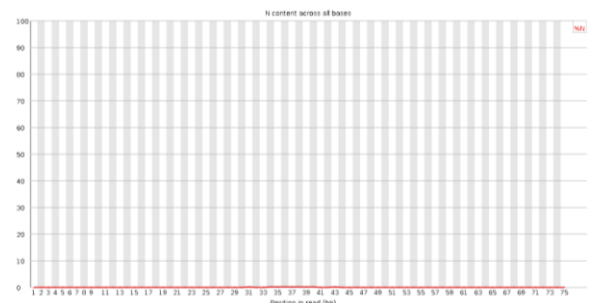


Figura 1.8. Presencia de N (Nucleótido desconocido) en la secuencia que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. En el eje de las X se muestra la posición del *read* y en el eje Y el porcentaje de N en los *reads*.

1.5.2.8 Distribución de longitud de secuencia

La Figura 1.9 muestra la distribución del tamaño de los *reads* del archivo analizado. Lo adecuado es tener *reads* de tamaños iguales, sin embargo, contar con *reads* de distintos tamaños no es un obstáculo, razón por la cual este gráfico no toma demasiada importancia en el análisis (Sánchez, 2015). En este caso los *reads* tienen distribución homogénea y todas presentan longitudes de 75 pb.

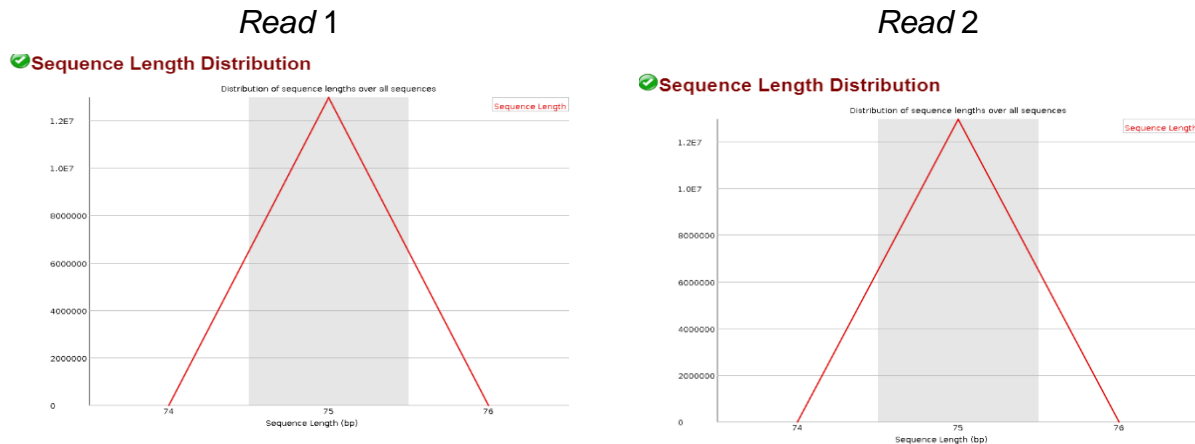


Figura 1.9. Distribución del tamaño de *reads* que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. En el eje de las X se muestra la longitud de la secuencia y en el eje de las Y el número de secuencias que corresponden a esa longitud.

1.5.2.9 Niveles de duplicación de secuencia

La Figura 1.10 muestra el porcentaje de *reads* duplicados. Se da una señal de advertencias cuando el porcentaje de secuencias duplicadas es mayor al 20 % del total de los *reads* y señalara fracaso si dicho porcentaje afecta a más del 50 % (Sánchez, 2015). En este caso se muestra una señal de advertencia debido al nivel de duplicidad superior al 50 %. La explicación de los picos mostrados en el gráfico es el siguiente:

- >10: Corresponde al primer pico, indica que más de 10 % de las secuencias se repiten 10 veces.
- >100: Corresponde al segundo pico indica que más del 10 % de las secuencias se repiten 100 veces.
- 1K: Tercer pico: Indica que el 15 % de las secuencias se repite 1k de veces.

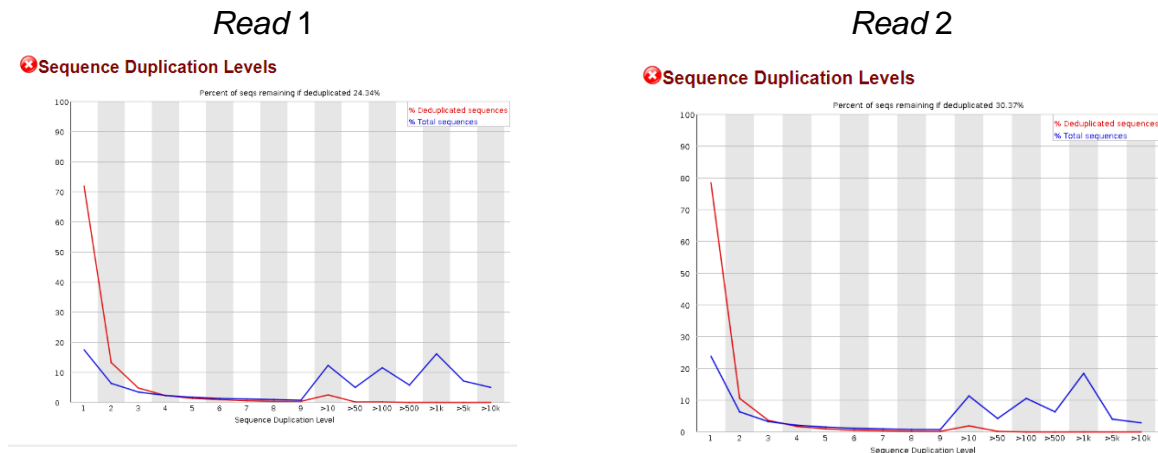


Figura 1.10. Niveles de duplicación de secuencia que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. En el eje de las X se muestran el nivel de secuencias duplicadas y en el eje de las Y el porcentaje asociado a cada uno de los niveles de duplicación. La línea azul indica el total de secuencias y la roja muestra las secuencias duplicadas.

1.5.2.10 Secuencias sobrerrepresentadas

En la Figura 1.11 se muestran las secuencias sobrerrepresentadas. Uno de los motivos por los cuales se presenta esta situación es por la presencia de adaptadores. Entre más grande el *read* los adaptadores son más grandes y por tanto la presencia de estos es mayor (Sánchez, 2015). Sin embargo, en la Figura 1.12 se observa que no hay presencia de dichos adaptadores, por lo que podría estar dándose por otros parámetros. Se da una señal de advertencia cuando el porcentaje que representa una secuencia repetida es mayor al 0.1 % del total de secuencias y señalará fracaso cuando este porcentaje supere el 1 % (Sánchez, 2015). En este caso se muestra una señal de advertencia debido a que hay secuencias que se representan más del 0.1 %.

Read 1

Overrepresented sequences

Sequence	Count	Percentage	Possible Source
CTTGTCCGTACCAAGTCTCTGAGTGCAGCTTTGACGACCCGGGAAGGCCCCGAGGGGGCGTCCCAAGTCCGTCC	34172	0.26382582184036496	No Hit
CCGACATCGAAGGATCAAAAAGCAACGTCGCTATGAACGTTGGCTGCCAAGCCAGTTATCCCTGGGTAAGT	31982	0.24691786942814448	No Hit
CTCTAAGAAGCTGGCCACGGAGGGATACCTCCGCATAGCTAGTTAGCAGGCTGAGGCTCGTTCGTTAACGGAAAT	28226	0.21791957296225395	No Hit
CTGCAAAAGGATTCTACCCGCGCTCGGTGGGAATTTAGCTTCAAGGCGGCCCTCGCGCGCTTCCGCGCGGGGG	26257	0.20271785684368676	No Hit
CTCAAACCTTCTTGGCTAAGCGCCATAGTCCCTCTAAGAAGCTGGCCACGAGGGATACCTCCGCATAGCTAG	25858	0.19957568267392704	No Hit
CACGCTTTCACGGTTCGTATTCTGACTGAAAATCAGAATCAAAACGAGCTTTTACCCTTTTGTCCACACGAGATT	25685	0.19768487375109873	No Hit
CTCCTACTCATCGGGCTCGCGCTTGCCTCGACGCGCGGGTATAGTTCGCGCGCTTCAGCGCCATCCATTTTCG	21526	0.16619204731756107	No Hit
CCGGAACCCAAAGACTTTTGATTTCTCATAAGGTGCCACGGAGTCCATAAAGCAACATCCGCTGATCCCTGGTCG	20961	0.1618299580857325	No Hit
CTAAGAAGCTGGCCACGGAGGGATACCTCCGCATAGCTAGTTAGCAGGCTGAGGCTCGTTCGTTAACGGAAATTA	19418	0.14991717885582186	No Hit
CTTTTGTCCACACGAGATTTCTGTTCTCGTTGAGCTCATCTAGGACACCTGGCTTATCTTTAACAGATGTGC	17549	0.13548751455801725	No Hit
GTCAATATCGCTGCGGGCTCCACACAGATTTCTCTGGCTTCGCCCCCTCAGGCATAGTTACCATCTTTTCGG	17444	0.13467685930537654	No Hit
CCCGACTGCTCCGTTAATCATTACTCCGATCCCGAAGGCCAACAGAAATAGSACCGAAATCCTATGATGTTATCC	16642	0.12848499728044463	No Hit
CTCGTTGAAGACCAACAAATTGCAATGATCTATCCCCATCAGATGAAATTTCAAGATTACCCGGGCTGTCGCG	16225	0.12526553784852867	No Hit
GCCGCAAGCTCACTCTGGTGGTGGCTTCCGTCAATTTCTTAAAGTTTCAAGCTTGCACCATACTCCCCCGG	15921	0.12291849787897843	No Hit
GCAAAAGGATTCTACCCGCGCTCGGTGGGAATTTAGCTTCAAGGCGGCCCTCGCGCGCTTCCGCGCGGGGGCT	15248	0.11772258373586227	No Hit
CGCTATTTGGAGCTGSAATTAACCGCGCTGCTGGCACCAGACTTGGCTTCAATGGATCTCTGTTAAGGGATTTAG	15148	0.11695883111429969	No Hit
CGCGTATTTAAGTGTGCAAAAGGATTCTACCCGCGCTCGGTGGGAATTTAGCTTCAAGGCGGCCCTCGCGCG	14251	0.11082521989888334	No Hit
CTCAAGTCAATTTCAAAAGCTGGACTAGAGTCAAGTCAACAGGGTCTTTTCCCGCGTATCCGCCAAGCCC	14195	0.10959288963888829	No Hit
CGACAATTTCAAGCACTTTTACTCTTTTCAAGTCCCTTTCATCTTCCCTCGCGGACTTGTGTTGCTATC	13956	0.10774766386527372	No Hit
CCTATACCAAGTCAAGCAAGGATTTGACAGTATCGCTGCGGGCTCCACCAAGATTCTTCCCTGCGCTTCG	13589	0.10491423874413985	No Hit
CTTTGACTCTCTTTTCAAAAGTCTTTTTCATCTTTCCCTCGCGGACTTGTGTTGCTATCGGTCTCTCGCCGATT	13348	0.10385358392617324	No Hit
GGGCAATCGGGCGGGCGCACGCGTCTAGCCCGGATTCTGACTAGAGGCGTTCAGTCATAATCCAGCG	13307	0.10273784235133259	No Hit
CTCGTTCCGGGCTCCAGCTATCCTGAGGAAACTCGGAGGAAACCAAGCTACTAGAGGCTGATTAAGTCTTTTC	13182	0.10177197657437935	No Hit

Read 2

Overrepresented sequences

Sequence	Count	Percentage	Possible Source
CTGACATGTGTGCGAGTCAACGGGCCAGTAAACCCGTAAGGCGCAAGGAAGCTGACCGCGGGATCCCTCGCGGG	57036	0.44834793323443333	No Hit
GTTAGTTTTACTACTGATGACAGCGTCGTAATAGTAATCAACCTAGTACGAGAGGAACCGTTGATTCGCACA	32069	0.2475895552889039	No Hit
CTTACCAGGTCAGACATAGTAAGGATTGACAGACTGAGAGCTTTCTTGTATTATGGTGGTGGTGCATGGC	28275	0.21829787874681963	No Hit
GTCTGACATGTGTGCGAGTCAACGGGCCAGTAAACCCGTAAGGCGCAAGGAAGCTGACCGCGGGATCCCTCGCG	26876	0.20749686257115915	No Hit
ACCAGGTCAGACATAGTAAGGATTGACAGACTGAGAGCTTTCTTGTATTATGGTGGTGGTGCATGGCCGT	23817	0.1838797728775598	No Hit
GG	20879	0.1611968685685117	No Hit
CAGGTCCAGACATAGTAAGGATTGACAGACTGAGAGCTTTCTTGTATTATGGTGGTGGTGCATGGCCGTTT	20303	0.15674984375585872	No Hit
CCTGCCAGTAGTCATATGCTGTCTCAAAGATTAAAGCCATGATGTGAAGTATGAACATAATTCAGACTGTGAAA	16778	0.12953498884576975	No Hit
GCCAGTAGTCATATGCTGTCTCAAAGATTAAAGCCATGATGTGAAGTATGAACATAATTCAGACTGTGAAA	13995	0.10884876438768314	No Hit
CTGCCAGTAGTCATATGCTGTCTCAAAGATTAAAGCCATGATGTGAAGTATGAACATAATTCAGACTGTGAAA	13674	0.10557847547246724	No Hit
CTTGGATTTATGAAAGCAAGCAACTGCGAAAGCATTGCAAGGATGTTTTCTTAATCAAGAACGAAAGTTGG	13597	0.10497599495386406	No Hit
GCGGGATTACCCCGTAGTTAAGCATATCAATAAGCGGAGGAAAAGAACTTACCAGGATTTCCCTAGTAACGG	13236	0.10218888499002314	No Hit

Figura 1.11. Secuencias sobrerrepresentadas correspondientes a la muestra 1 de las semillas secas de *Cedreia odorata* L. Solo se muestran las secuencias que constituyen más de un 0.1 % del total.

1.5.2.11 Contenido de adaptadores

En la Figura 1.12 se detecta posibles adaptadores en los *reads* y en cierta forma complementa a la Figura 1.11. Lo normal es que en cada posición de los *reads*

aparezca un nucleótido distinto entre ellos. El porcentaje de cada uno de los cuatro nucleótidos en cada posición de los *reads* debe ser de 25 % aproximadamente, pero en el caso de repetición abundante se obtiene un 50 % en las primeras posiciones de los *reads* (Sánchez, 2015). Se dará una señal de advertencia cuando una secuencia aparezca repetida en más del 5 % del total de *reads* y señalará fracaso si aparece repetida en más del 10 % (Sánchez, 2015). En este caso no se muestran contenido de adaptadores en las secuencias.

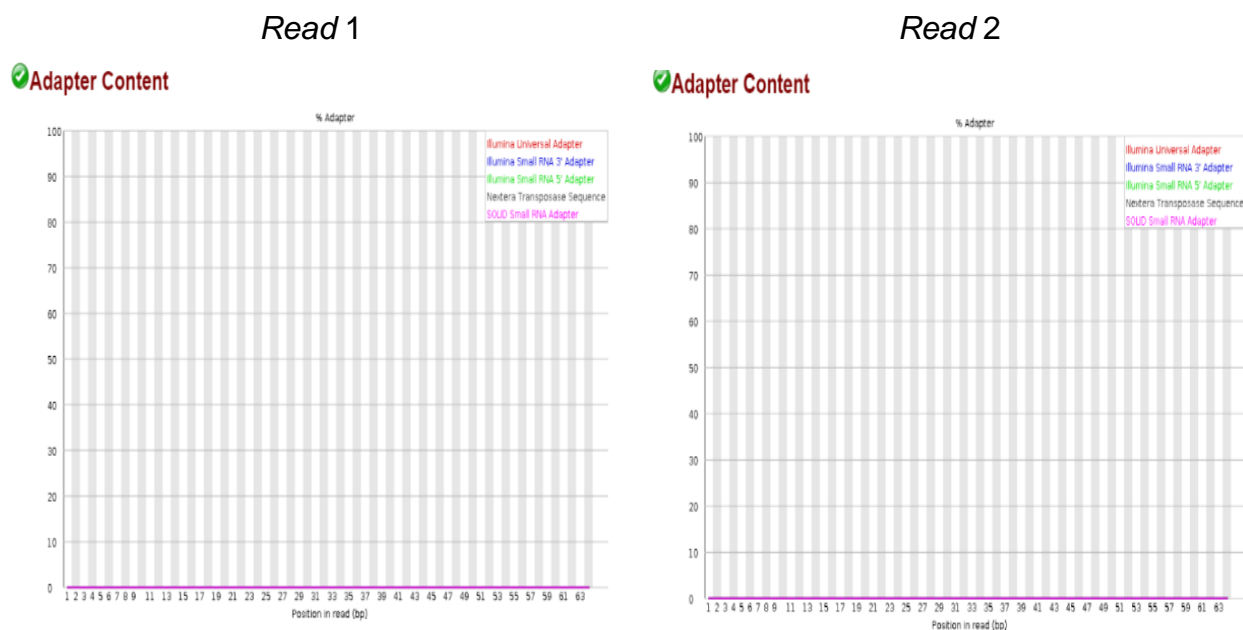


Figura 1.12. Contenido de adaptadores en la secuencia que corresponde a la muestra 1 de las semillas secas de *Cedrela odorata* L. En el eje de las X se muestran la posición en el *read* y en el eje de las Y el porcentaje del contenido de adaptadores asociado a los *reads*.

1.5.3 Ensamble de *novo* del transcriptoma

El ensamble de *novo*, el transcriptoma de semillas de *C. odorata* obtenido con Trinity, tuvo un total de 102,216,934 bases ensambladas y 95,739 transcritos, de los cuales 52,490 eran genes únicos. Así mismo, se obtuvo un porcentaje de GC (Guanina-Citosina) de 42.62. De acuerdo con las longitudes de los cóntigos ensamblados en el transcriptoma, se determinó que el 50 % de los nucleótidos del ensamble (N50) se encuentran con al menos una longitud de 1,668 bases y un promedio de 1,067.66 bases por cóntigo. En las estadísticas basadas en las isoformas

más larga por gen se obtuvo un total de 42,238,489 bases ensambladas, un N50 de 1,457 y un promedio de 804.70 bases por cóntigo (Cuadro 1.2).

Cuadro 1.2. Resumen de la estadística del ensamble de transcriptoma de semillas de *Cedrela odorata* L.

Recuentos de transcripciones, etc.	
Total de genes triniy	52,490
Transcripciones totales de trinity:	95,739
Porcentaje GC:	42.62
Estadísticas basadas en TODOS los cóntigos de transcripción:	
Cóntigo N10:	3,427
Cóntigo N20:	2,676
Cóntigo N30:	2,249
Cóntigo N40:	1,932
Cóntigo N50:	1,668
Longitud mediana de cóntigo:	779
Cóntigo promedio:	1,067.66
Bases totales ensambladas:	102,216,934
estadísticas basadas solo en la isoforma más larga por 'GEN':	
Cóntigo N10:	3,303
Cóntigo N20:	2,521
Cóntigo N30:	2,088
Cóntigo N40:	1,756
Cóntigo N50:	1,457
Longitud mediana de Cóntigo:	420
Cóntigo promedio	804.7
Bases totales ensambladas:	42,238,489

1.5.4 Anotación funcional

El resultado de la anotación funcional del transcriptoma de *novovo* de *C. odorata* obtenido con *Trinotate* se puede visualizar en el archivo *Trinotate.xls* (Cuadro A3). El resumen estadístico de dicha anotación se observa en el Cuadro 1.3. Los resultados obtenidos del *Blastx* muestran un total de 57,905 secuencias nucleotídicas, de los cuales 57,148 fueron únicas y similares a otras ya anotadas. Mientras tanto, en los resultados del *Blastp* se obtuvieron 43,303 secuencias peptídicas, de éstas 38,157 fueron únicas. Así mismo, las secuencias obtenidas también se anotaron en las bases de datos de *kegg*, *eggNOG* y *pfam* y se encontraron una mayor cantidad de aciertos en

la base de datos de *keeg*, donde se muestran 48,989 rutas metabólicas de las cuales 14,514 fueron únicas.

Cuadro 1.3. Análisis estadístico de la anotación con *Trinotate* del transcriptoma de *novo* de semillas de *Cedrela odorata* L. Se muestra el número de secuencias únicas y totales en las distintas bases de datos.

	Únicos	Totales
Id_gen	52,490	95,739
Id_transcrito	95,739	95,739
sprot_Top_BLASTX_hit	57,148	57,905
gene_ontology_BLASTX	12,063	55,654
prot_id	55,080	55,080
prot_coords	38,195	55,080
Kegg	14,514	48,989
eggNOG	5,080	45,225
sprot_Top_BLASTP_hit	38,157	43,303
Pfam	34,384	42,176
gene_ontology_BLASTP	9,277	41,506
gene_ontology_Pfam	1,789	27,952

1.5.5 Clasificación de las funciones eggNOG

Se utilizó la base de datos eggNOG para la descripción y clasificación funcional de las proteínas ortólogas. Los resultados muestran que se asignaron 45,225 proteínas de las cuales 5,080 fueron únicas (Cuadro 1.3). Las secuencias fueron clasificadas en 22 grupos funcionales entre las que destacan: modificaciones postraduccionales, recambio de proteínas, chaperonas (O), transcripción (K), mecanismos de transducción de señales (T), traducción, estructura de ribosomas y biogénesis (J), tráfico intracelular, secreción y transporte vesicular (U) (Figura 1.13).

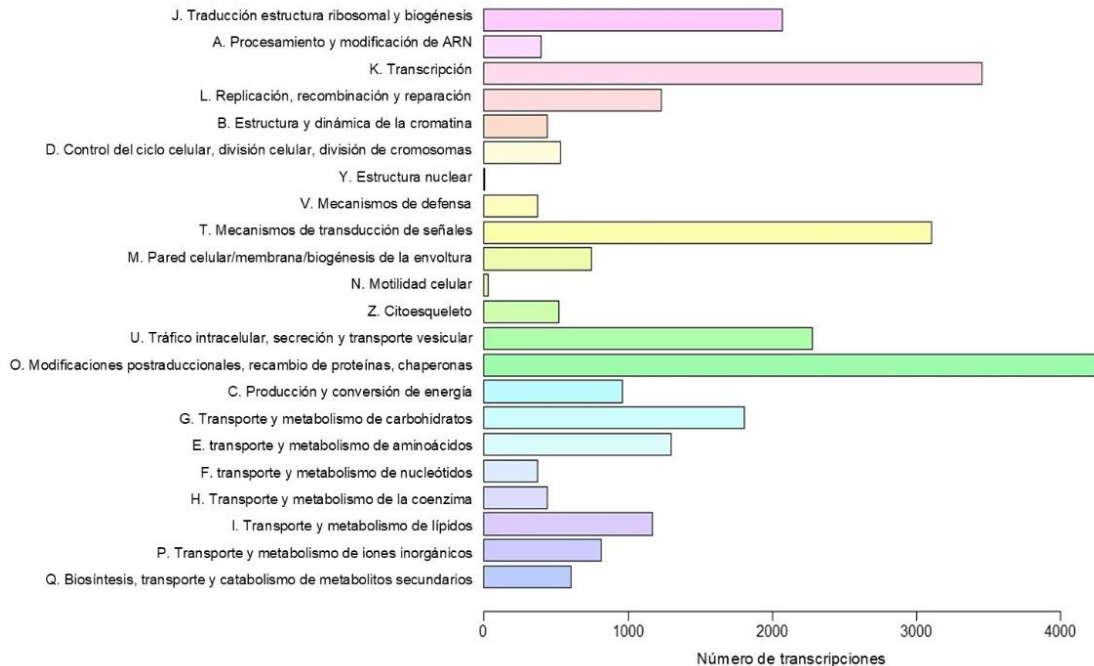


Figura 1.13. Clasificación funcional de proteínas ortólogas obtenidas del transcriptoma de *novu* del proceso germinativo de *Cedrelela odorata* L.

1.5.6 Clasificación Ontológica

Del total de transcritos obtenidos en el transcriptoma, sólo el 58 % (55,654) se localizó en la base de datos de Ontología Génica (GO) y de éstas, el 28 % (12,063) fueron únicas (Cuadro 1.3). Estos transcritos se distribuyeron en 100 grupos funcionales organizados en tres categorías: procesos biológicos (PB), componentes celulares (CC) y función molecular (FM). En la categoría PB predominan los términos regulación de la transcripción-plantilla de ADN, traducción y ubiquitinación de proteínas. En los CC, predominaron los términos núcleo, componente integral de la membrana y citoplasma, mientras tanto en la categoría FM predominaron los términos Unión de ATP, unión de iones metálicos y unión al ADN (Figura 1.14).

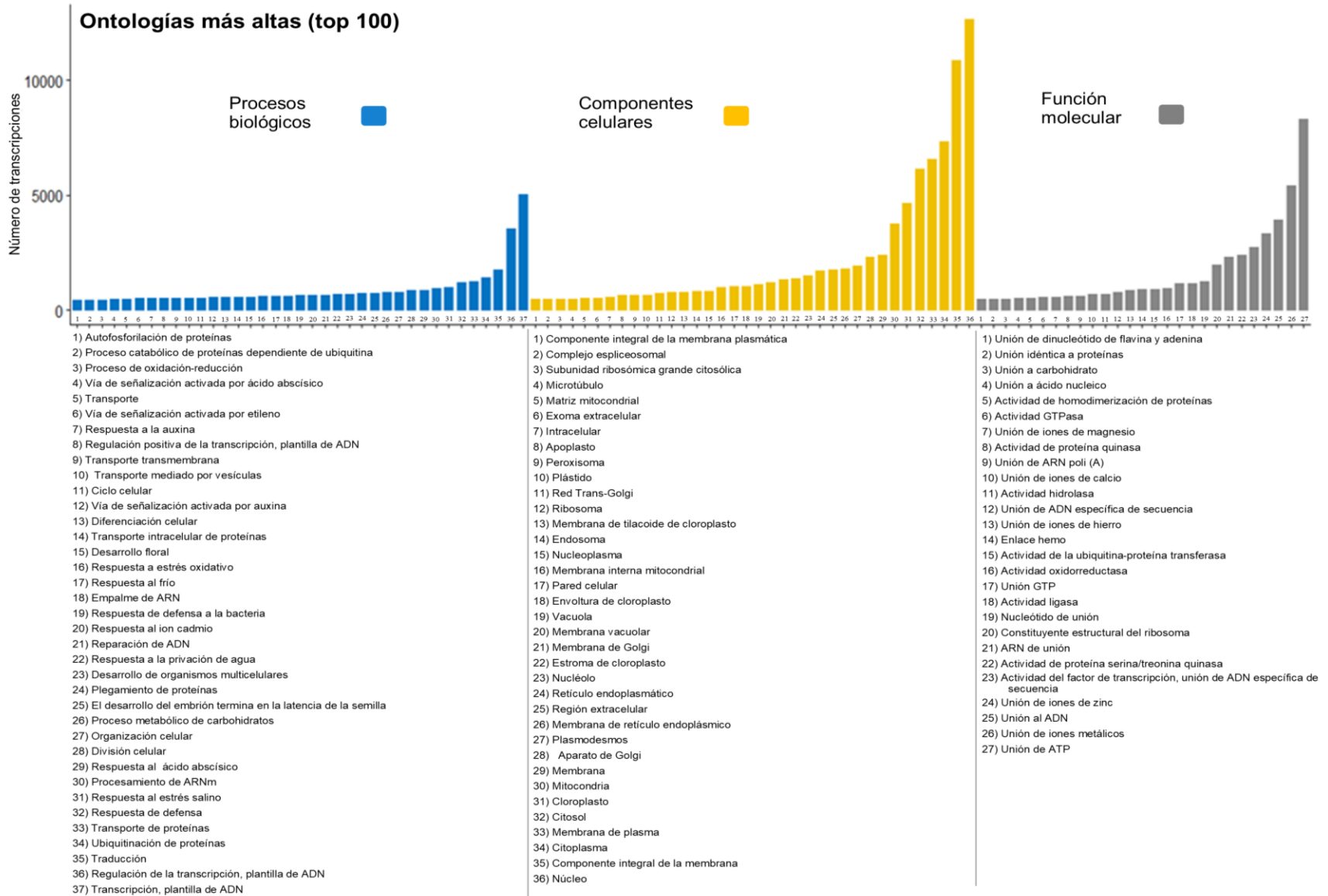


Figura 1.14. Clasificación de los términos GO obtenidos del transcriptoma de *novo* del proceso germinativo de *Cedrela odorata* L.

1.5.7 Base de datos Nr

Los resultados de la base de datos Nr (no redundantes) muestran que la mayor cantidad de transcripciones anotadas en *C. odorata* coinciden con *Arabidopsis*, seguido por el género *oryza* y otros diferentes de plantas como *Homo* y *Mus* (Figura 1.15).

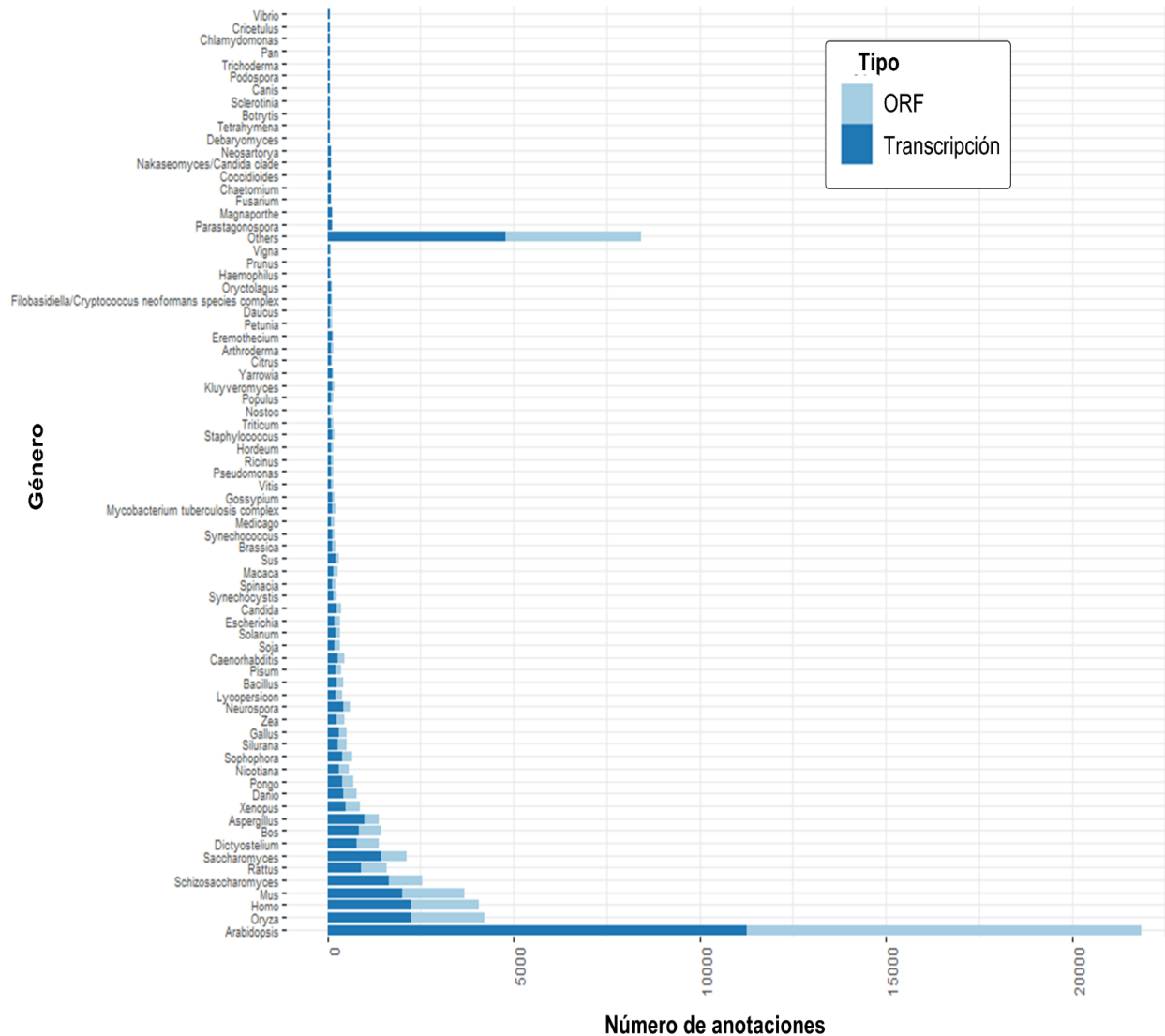


Figura 1.15. Distribución de aciertos en la base de datos Nr del transcriptoma de *novo* del proceso germinativo de *Cedrele odorata* L.

1.6 DISCUSIÓN

El laboratorio de secuenciación entregó lecturas limpias por lo que no fue necesario hacer el recorte de lecturas (*Trimming*), que está indicado para eliminar regiones de baja calidad y así aumentar la calidad de las secuencias (Del Fabbro *et al.*, 2013). Lo anterior se refleja en los distintos parámetros evaluados, entre ellos la calidad de secuencia por base (Figura 1.3) y los niveles de calidad por secuencia (Figura 1.5) que están por arriba del valor Q30. Así mismo, las celdas de flujo muestran que no hay contaminación en la muestra entre otros parámetros.

Los nucleótidos de baja calidad pueden derivarse de distintos errores en la secuenciación, por ejemplo, las burbujas de aire o el mal funcionamiento de laser/lente etc. Ignorar el análisis de calidad y generar el ensamble con dichas secuencias puede dar lugar a interpretaciones falsas de los datos, motivo por el cual el análisis y el procesamiento del recorte de lecturas toma gran importancia (Del Fabbro *et al.*, 2013). En este trabajo las secuencias utilizadas mostraron ser de alta calidad por lo tanto fueron adecuadas para generar el ensamble de *novo* y posteriormente el transcriptoma.

En este estudio se presenta el primer transcriptoma *de novo* de semillas de *C. odorata* en el que se obtuvieron 52, 490 modelos de genes. Este número es muy similar a lo reportado por Finch *et al.* (2019) quienes analizan el transcriptoma de hojas de esta misma especie. Así mismo, al comparar la anotación de genes en ambos estudios se observa que muchos de estos coinciden aun siendo transcriptomas de diferentes tejidos, por lo que puede observarse una dinámica transcripcional de estos genes. De esta forma, ambos estudios toman gran importancia al complementarse y enriquecer la base de datos de secuencias para la especie, así mismo, esta información se vuelve una herramienta importante para futuras investigaciones interesados en estudiar la especie.

Las investigaciones con tecnologías *RNASeq* enfocadas al estudio de la germinación en semillas no ortodoxas de especies forestales son poco frecuentes; entre estos destaca el trabajo publicado por Zhang *et al.* (2019) donde se analiza el proceso germinativo de las semillas de *Populus euphratica* Oliver y *Populus pruinosa*

Schrenk, bajo el contexto de estrés por salinidad; así como el trabajo de Qu *et al.* (2020) donde se estudia a *Populus x xiaohei* T.S. Hwang et Liang, analizando los cambios de producción de energía durante la etapa de la germinación de la semilla; así mismo el trabajo de Gómez-Maqueo *et al.* (2020) donde se examinan transiciones fisiológicas en lotes de semilla de *Ceiba aesculifolia* (Kunth) Britt. & Baker f., recolectadas en distintos años con respuestas fenotípicas contrastantes a un tratamiento de cebado. Del mismo modo otra especie como *Garcinia mangostana* L. que es un árbol frutal y tiene dificultades en el almacenamiento de semillas al ser de tipo recalcitrante, al respecto ellos evalúan distintos puntos de tiempo (basándose en la curva de absorción de agua) durante la germinación de la semilla (Goh *et al.*, 2019).

Por otro lado, para la evaluación de la calidad del transcriptoma de semilla de *C. odorata* se utilizó la métrica N50 que es definida como la longitud del cóntigo más grande del total de cóntigos ordenados de forma ascendente y representa el 50 % de la longitud del conjunto. Por lo tanto, en este trabajo la medida N50 mostro que el 50 % de la longitud total de la secuencia contiene secuencias iguales o superiores a 1,668 pb. Cuando el número de cóntigos es menor y mayor el valor de N50 se dice que el ensamble es mejor (Cortese *et al.*, 2021). Dado los resultados obtenidos con valor medio de cóntigos de 1,067 y un N50 de 1,668, se puede decir que en este trabajo se generó un buen ensamble.

En el análisis de clasificación por eggNOG se encontró que los grupos más representados corresponden a “modificaciones postraduccionales, recambio de proteínas, chaperonas” (O) y “mecanismos de transducción de señales” (T), estos grupos también han sido localizados en la germinación de semillas de *G. mangostana* (Go *et al.*, 2019). Se ha visto que las modificaciones postraduccionales responden a situaciones de estrés cuando su entorno ha cambiado, su función radica en adicionar una molécula pequeña y modificar la ubicación, estabilidad o función de la proteína (Han *et al.*, 2022). Algunas de las modificaciones postraduccionales más sobresalientes en plantas son: la fosforilación, ubiquitinación, acetilación, glicosilación, nitrosilación, metilación y lipidación (Yu *et al.*, 2021).

Otro ejemplo de las funciones de las modificaciones postraduccionales es la reactivación rápida de los programas transcripcionales incluyendo a factores de transcripción y proteínas de señalización (Han *et al.*, 2022), lo cual coincide con el hecho de que las funciones de “transcripción” (K) y “mecanismos de transducción de señales” (T) son los siguientes grupos más dominantes dentro de la clasificación funcional de las proteínas encontradas por eggNOG en semillas de *C. odorata* (Figura 1.13). Así mismo, se ha registrado que estas modificaciones juegan un papel importante en la germinación de semillas de arroz y *Arabidopsis* (Yu *et al.*, 2021). La transición de un estado quiescente en la semilla seca a uno que permite la germinación como consecuencia de la imbibición, representa un cambio profundo que afecta los distintos componentes celulares y macromoléculas (Rajjou *et al.*, 2012) y que puede equipararse a un estado de estrés (Zaynab *et al.*, 2018), donde las modificaciones postraduccionales y la función de chaperonas juegan un papel importante en facilitar la activación, recambio y organización de proteínas asociadas al funcionamiento y homeostasis celular.

Dentro de los procesos biológicos en los términos GO, se encontró que la ubiquitinación de proteínas es el más representativo, además de la autofosforilación de proteínas, aunque en este último, el número de genes identificados con este término fue menor (Figura 1.14). La ubiquitinación es una de las modificaciones postraduccionales más frecuentes en plantas, está involucrada en el recambio de proteínas, la integridad genómica y el proceso de señalización (Yu *et al.*, 2021). Por ejemplo, el gen RGA-like2 (RGL2) codifica para proteínas DELLA (Foley *et al.*, 2013) y se expresa durante la imbibición de las semillas (Yu *et al.*, 2021), este es un regulador transcripcional que reprime la señalización de las giberelinas e impiden la germinación (Foley *et al.*, 2013). Por tanto, para promover la germinación y activar al ácido giberélico (AG) es importante la ubiquitinación de las proteínas DELLA tal como sucede en semillas de *Arabidopsis* (Yu *et al.*, 2021)

Los mecanismos de transducción de señales son clave en la germinación de las semillas y muchas de ellas son coordinadas por las modificaciones postraduccionales. Al respecto en este transcriptoma se detectaron proteínas quinasas y los dominios

WD40 y P450 (Cuadro A3), estos dos últimos han sido relacionados con el metabolismo de los fenilpropanoides (Torre *et al.*, 2014). Los fenilpropanoides son derivados de la fenilalanina y tienen diversas estructuras y funciones. Por un lado, funcionan como señalizadores e inducen genes de defensa y protectores UV. También se relacionan con la prevención de pérdida de agua y la estabilidad de la pared celular (Foley *et al.*, 2013). La lignina que es un polímero derivado de los fenilpropanoides y es indispensable para el soporte mecánico, rigidez e hidrofobicidad a las paredes celulares secundarias en el crecimiento de las plantas y facilita el transporte de agua y nutrientes a distancias largas (Zhao, 2016). Por otro lado, los fenilpropanoides en semillas están vinculados con la profundidad de la latencia primaria de acuerdo con la temperatura obtenida durante su desarrollo. En semillas de *Arabidopsis* se ha observado que las bajas temperaturas aumentan la expresión de este gen, el cual tiene una correlación directa con las concentraciones de procianidinas en la cubierta de las semillas manteniendo una latencia más profunda. Referente a ello se ha reportado que los individuos con menor cantidad de procianidina experimentan una latencia primaria menos profunda (Macgregor *et al.*, 2015). Hasta donde se sabe, no hay reporte de que las semillas de *C. odorata* experimenten algún tipo de latencia, a pesar de ello, es posible que su rol en este transcriptoma esté relacionado con alguna otra función en la pared celular o bien sea parte del ARN almacenado de las etapas de desarrollo y maduración de la testa. Sin embargo, esta afirmación necesita de estudios más profundos para analizar el papel real de los fenilpropanoides en el proceso de germinación de esta especie.

Las fitohormonas también son fundamentales durante el proceso de germinación (Yu *et al.*, 2021). En este transcriptoma se localizaron rutas de señalización activadas por el ácido abscísico (ABA), auxina y etileno (Figura 1.14), y aunque el AG no se encontró en el gráfico de los términos GO los genes involucrados a este proceso sí se encuentran dentro del transcriptoma (Cuadro A3). El AG y ABA regulan de forma antagónica a la germinación. Aquí la fosforilación (otro tipo de modificación postraduccional) también participa de forma importante durante la germinación. Por ejemplo, diversas proteínas quinasas como SnRK2, BIN2, CIPK26, PKS5 y KEG fosforilan al gen que codifica para *ABI5* (Insensible al ácido abscísico 5,

quien codifica un miembro de la familia básica del factor de transcripción de leucina, implicado en la señalización de ABA durante la germinación de la semilla) (Cuadro A3) e inhiben la germinación de las semillas al promover su estabilidad. La fosforilación de distintas isoenzimas se ha localizado durante la germinación en semillas de sorgo y trigo (Yu *et al.*, 2021).

Por otra parte, se encontró una alta actividad en los procesos relacionados con traducción, estructura ribosomal y biogénesis (J) en la base de datos de eggNOG (Figura 1.13). Tanto los procesos de traducción como la participación de los ribosomas son reafirmados por la clasificación funcional de los términos GO (Figura 1.14). Los ribosomas forman el aparato de traducción universal para la decodificación de la información genética en proteínas (Kumar, 2021), por tanto, su biogénesis es indispensables para que se lleve a cabo la traducción *de novo* de proteínas implicadas en distintos procesos celulares como los son el crecimiento y la división celular (Kumar, 2021) mismos que darán paso a la germinación de la semilla.

Además de estas funciones más representativas en el transcriptoma generado, también se encontraron otras con mucha importancia, por ejemplo, existe una comunicación de la célula con su entorno a través de las membranas y mediante vesículas que transportan proteínas y lípidos por la vía exocítica y endocítica (Tokarev *et al.*, 2009), que se refleja en la clasificación funcional tráfico intracelular, secreción y transporte vesicular (U) (Figura 1.13) y en los componentes celulares de los término GO, componente integral de la membrana y plasmodesmos (Figura 1.14). Además de muchas otras funciones relacionados con la conversión y producción de energía (C) (Figura 1.13), donde se incluye a los términos GO, mitocondria y ATP (Figura 1.14). Así mismo, el transporte y metabolismo de los carbohidratos (G) y otros (Figura 1.13). La generación de ATP (adenosín trifosfato) es indispensable durante la germinación temprana para sostener las demandas metabólicas que culminan con a la emergencia de la radícula (Qu *et al.*, 2020). Durante este proceso se ven involucradas un conjunto de reacciones bioquímicas y rutas metabólicas como la respiración, la glucólisis, el ciclo del ácido tricarbóxico y la cadena transportadora de electrones (Zaynab *et al.*, 2018). Aquí, la actividad mitocondrial juega un papel importante y al igual que los procesos

mencionados anteriormente, también depende de las modificaciones postraduccionales de las proteínas asociadas a las funciones metabólicas. La biogénesis mitocondrial es producida poco a poco después de la imbibición de las semillas, lo que es consistente con el robustecimiento de las rutas implicadas en la síntesis de ATP. Además de ello, también está involucrada en la regulación redox para la transcripción y regulación hormonal (El-Maarouf-bouteau, 2022).

Muchas de las funciones obtenidas en el transcriptoma de *Myrciaria Dubia* “camu-camu, donde también se analizan fases del proceso germinativo y crecimiento de plántula (Castro *et al.*, 2020), coinciden con los resultados de este experimento. Además de los que se han mencionado anteriormente, se observa también un mayor número de transcritos relacionados con la respuesta a defensas, responsable de estrés por sales y procesos metabólicos de carbohidratos. En los componentes celulares (CC) también coinciden con el término cloroplastos. En las angiospermas se ha visto que el cloroplasto en semillas está relacionado con la prevención de la hipoxia al mantener la respiración en semillas maduras (Smolikova y Medvedev, 2016). Por otro lado, en la función molecular (FM) coincidieron en los términos unión de ion zinc y actividad de proteína serina/treonina quinasa (Castro *et al.*, 2020). El ion zinc se ha encontrado participando como un señalizador en la transición de información dentro de las células y en la comunicación entre ellas, es decir, participa en la señalización en vías en las que los iones de zinc son estímulos extracelulares o mensajeros intracelulares y dichas vías actúan a través de la señalización de Ca^{2+} , redox y fosforilación (Maret, 2017). Posiblemente esta señalización se esté dando por la fosforilación, motivo por el que el término actividad de proteína serina/treonina quinasa, sea uno de los términos de mayor frecuencia en ambos estudios.

El análisis de similitud contra la base de datos Nr mostró que la mayor cantidad de las anotaciones obtenidas en este transcriptoma coinciden con la de los géneros *Arabidopsis* y *Oryza* (Figura 1.15), posiblemente porque son dos de los géneros más estudiados y con genomas secuenciados. Estos mismos géneros coinciden con una mayor similitud en las anotaciones encontradas en *A. indica*, además de que ésta, también coincidió con anotaciones encontradas en *Vitis vinifera* L. (Krishnan *et al.*,

2012). En *Q. pubescens* también se encontró una mayor cantidad de anotaciones que coincidieron con las de *V. vinifera*, además de *Pronus persica* (L.) Batsch, *Theobroma cacao* L., *Populus Trichocarpa* Torr. & A. Gray y otros (Torre *et al.*, 2014). En *M. oleifera* también se encontró mayor similitud con secuencias de *V. vinifera*, *T. cacao*, *Nelumbo Nucifera* Bercht. & J.Presl, *Setaria italica* (L.) P.Beauv. y *Jatropha curcas* L. (Yang *et al.*, 2018). Aún con menos cantidad de transcritos anotados, en este transcriptoma también se encontraron coincidencias con otros géneros de plantas como *Nicotiana*, *Zea*, *Lycopersicon*, *Soya*, *Solanum*, *Gossypium*, *Populus* y otros, sin embargo, se puede observar que estas anotaciones no son específicas de plantas si no que existe alta coincidencia con otros géneros diferente de plantas como lo son los del géneros de *Homo*, *Mus*, *Schizosaccharomyces*, *Rattus*, *Dictyostelium* y otros.

1.7 CONCLUSIONES

En este estudio se generó de *novo* y anotó el transcriptoma de semillas de *C. odorata* analizando el proceso germinativo de las mismas durante tres etapas: en semilla quiescente (seca), embebida y germinada.

La generación del transcriptoma evidenció los genes y procesos involucrados durante la germinación de las semillas, el cual tiene un gran potencial para los fines de manejo, conservación y propagación de la especie, ya que pueden ser utilizados para posteriores estudios enfocados en analizar en forma directa genes o vías específicas del proceso germinativo.

La generación de este transcriptoma también es importante para posteriores estudios comparativos en el género *Cedrela* y especialmente para la especie *C. odorata*, ya que representa una referencia en la generación de un nuevo ensamble utilizando el mismo o diferentes condiciones y tejidos.

El análisis de términos GO muestra que destacan las modificaciones postraduccionales principalmente relacionadas con ubiquitinación de proteínas y fosforilación. Además, destacan transcritos relacionados con vías de transducción de señales, así como biosíntesis de fenilpropanoides.

Otros procesos importantes en este transcriptoma están relacionados con la traducción y la biogénesis de los ribosomas. Además de otras funciones como la comunicación celular y la producción de energía en forma de ATP, los cuales son fundamentales para dar paso al proceso de germinación de las semillas de *C. odorata*.

Con el transcriptoma generado es posible realizar el análisis de expresión diferencial que permite conocer a los genes más dinámicos durante el proceso germinativo en semillas de *C. odorata*.

CAPÍTULO II. ANÁLISIS DE EXPRESIÓN DIFERENCIAL DE GENES DURANTE LA GERMINACIÓN DE SEMILLAS DE *Cedrela odorata* L.

2.1 RESUMEN

Cedrela odorata L. es una especie forestal maderable importante en el mundo debido a cualidades de su madera, lo que ha ocasionado alta explotación de la especie en su zona de distribución natural y declarado su protección. La conservación de la semilla en bancos de germoplasma es una manera de preservar la especie, pero la pronta pérdida de su viabilidad lo complica. Analizar el proceso fisiológico de la germinación a nivel genético y molecular permite concebir novedosos sistemas de manejo y conservación del germoplasma. Con el transcriptoma generado y anotado, en el capítulo anterior, se propuso un análisis para identificar genes expresados diferencialmente (DEG) asociados a la transición de una semilla seca (S) a un estado embebido pregerminativo (E), culminando con la emergencia de la radícula (G). Para ello se cuantificaron los transcritos con *Salmon*, usando los archivos *fastq* de las muestras. Con los datos se generó una matriz de conteo de transcritos en cada condición y se identificaron los DEG a través de las comparaciones SvsE, EvsG y SvsG. Para el análisis se utilizó el lenguaje de programación R y el paquete *DESeq2*. El umbral para considerar un DEG fue: un $p_{adj} \leq 0.01$ y un valor absoluto $\log_2 \text{FoldChange} \geq 2$. Los DEG se analizaron con los términos de la ontología génica (GO). Finalmente se realizó una anotación individual de los 100 DEG con mayor y menor expresión en cada comparación y un análisis de enriquecimiento GO con REVIGO. Con este análisis se obtuvieron 8,863 DEG. El análisis evidenció dos etapas del proceso germinativo. La primera conduce a la reactivación metabólica resultado de la imbibición donde hay preponderancia de genes de señalización, metabolismo energético y pared celular. La segunda apoya a los procesos de emergencia de la radícula donde los DEG se asocian al metabolismo de la pared celular y alargamiento del embrión. Además, se detectaron genes relacionados con tolerancia a la desecación como las HSP, GGT1 y GOLS2. Este trabajo describe el primer estudio de secuenciación y anotación del transcriptoma de semillas de *C. odorata* importante para posteriores planes de manejo y conservación de semillas no ortodoxas a largo plazo.

Palabras clave: *DESeq2*, gen diferencialmente expresado, imbibición, $\log_2 \text{FoldChange}$, ontología génica.

ANALYSIS OF DIFFERENTIAL EXPRESSION OF GENES DURING SEED GERMINATION OF *Cedrela odorata* L.

2.2 ABSTRACT

Cedrela odorata L. is an important timber forest species in the world due to the qualities of its wood, which has caused high exploitation of the species in its natural distribution area and declared its protection. The conservation of the seed in germplasm banks is a way of preserve the species, but the rapid loss of its viability complicates it. Analyze the physiological process of germination at the genetic and molecular level makes it possible to conceive novel management and conservation systems for germplasm. With the transcriptome generated and annotated, in the previous chapter, an analysis was proposed to identify differentially expressed genes (DEG) associated with the transition from a dry seed (S) to an embedded pregerminative state (E), culminating in the emergence of the radicle (G). For this, the transcripts were quantified with Salmon, using the *fastq* files of the samples. With the data, a transcript count matrix was generated in each condition and the DEG were identified through SvsE, EvsG and SvsG comparisons. For the analysis, the R programming language and the DESeq2 package were used. The threshold to consider a DEG was a $\text{padj} \leq 0.01$ and an absolute value $\text{Log}_2 \text{FoldChange} \geq 2$. DEGs were analyzed with Gene Ontology (GO) terms. Finally, an individual annotation of the 100 DEG with the highest and lowest expression in each comparison and a GO enrichment analysis with REVIGO was performed. With this analysis, 8,863 DEG were obtained. The analysis evidenced two stages of the germination process. The first leads to metabolic reactivation resulting from imbibition where there is a preponderance of signaling genes, energy metabolism and cell wall. The second supports radicle emergence processes where DEGs are associated with cell wall metabolism and embryo elongation. In addition, genes related to desiccation tolerance such as HSP, GGT1 and GOLS2 were detected. This work describes the first sequencing and annotation study of the seed transcriptome of *C. odorata*, important for subsequent management plans and long-term conservation of unorthodox seeds.

Key words: DESeq2, differentially expressed gene, imbibition, $\log_2 \text{FoldChange}$, gene ontology.

2.3 INTRODUCCIÓN

El almacenamiento de semillas en bancos de germoplasma representa una de las formas efectivas para preservar los recursos genéticos y para producir plantas para abastecer plantaciones con fines comerciales o de conservación (Whitehouse *et al.*, 2020). Sin embargo, las semillas de muchas especies tropicales se caracterizan por su sensibilidad a la desecación, lo que ocasiona que tengan una vida de almacén muy corta, afectando su conservación ya que requiere recolectas constantes para mantener la viabilidad del germoplasma (Obroucheva *et al.*, 2016).

Si bien las semillas de *C. odorata* son clasificadas tradicionalmente como ortodoxas, debido a su reducido contenido de agua al momento de ser liberadas, pierden la viabilidad rápidamente en condiciones ambientales (Sautu *et al.*, 2006; García y Abdelnour, 2013); incluso controlando las condiciones de almacenamiento (refrigeración a 5°C, con 7% en contenido de humedad y dentro de bolsas de polietileno), se observa una reducción de la germinación en un 50 % en menos de dos años (García y Abdelnour, 2013). El conocimiento empírico ha mostrado que la viabilidad de las semillas al almacenarse en frío es muy variable dependiendo del origen de la semilla, siendo muy baja en algunos casos. Por ello, desde la perspectiva de los bancos de germoplasma, es considerada como una especie recalcitrante. De esta forma, el estudio de los mecanismos implicados en mantener el vigor y la viabilidad de las semillas es una pieza clave para conservar la riqueza genética de la especie (Zinsmeister *et al.*, 2020).

En consecuencia, es importante profundizar en el estudio de la fisiología y la genética implicada en procesos que conllevan a la germinación vigorosa de las semillas de *C. odorata* y enriquecer el conocimiento enfocado a maximizar el potencial de viabilidad durante su conservación *ex-situ* (Whitehouse *et al.*, 2020). No obstante, el proceso germinativo de las semillas implica una serie de mecanismos fisiológicos complejos. Tras la pérdida de agua y el término de su desarrollo, las semillas entran a un estado de anhidrobiosis (Zinsmeister *et al.*, 2020), en el que se consideran metabólicamente inactivas. A medida que progresa la hidratación, comienza la reanudación de su metabolismo, lo cual requiere de diversos cambios bioquímicos y

estructurales catalizados por enzimas y demás proteínas funcionales que darán paso a la síntesis de proteínas y ARN de *novo*, éstas llevarán a la semilla al proceso germinativo (Bewley *et al.*, 2013). Además, ocurre la traducción, así como la degradación de ARNm que fueron almacenados en las semillas durante el período de maduración; se han reportado más de diez mil transcritos en arroz, cebada y *Arabidopsis* (Zaynab *et al.*, 2018). Los ARNm son muy diversos según su información genética y partes de las semillas, como por ejemplo el endospermo y el embrión, y son modificados de forma específica durante la imbibición (Zaynab *et al.*, 2018). Este cambio en las poblaciones de transcritos, almacenados y sintetizados *de novo*, que ocurre durante la imbibición reflejan las necesidades de la transición metabólica y de los programas genéticos que regulan el desarrollo que ocurren entre la quiescencia y la imbibición, que finalmente conducen a la germinación. El análisis del transcriptoma asociados a la germinación en *C. odorata* puede proveer información valiosa, no solo respecto a la identidad de genes que participan en las diversas rutas metabólicas, sino revelar los mecanismos implicados en mantener la funcionalidad celular asociada a una germinación vigorosa.

Las investigaciones transcriptómicas para especies forestales han sido poco frecuentes, aun mas cuando se trata de procesos fisiológicos enfocados a la germinación de semillas entre estos se encuentran los trabajos de Zhang *et al.* (2019), Gómez-Maqueo *et al.*(2020) y Qu *et al.* (2020). Ejemplo de otros trabajos se han llevado a cabo a través de distintos contextos biológicos como los relacionados al estudio de las rutas de biosíntesis de terpenoides en *Azadirachta indica* A.Juss (Krishnan *et al.*, 2012), la obtención de respuestas genómicas debido al ozono en *Quercus rubra* L. (Soltani *et al.*, 2020), estrategia de respuesta al frío en *Hevea brasiliensis* Muell (Campos *et al.*, 2019), análisis de la biosíntesis del ácido nervónico en semillas oleaginosas de *Malania oleifera* Chun & SK Lee (Yang *et al.*, 2018) por mencionar otros.

Hasta hace poco tiempo la información genética disponible en *C. odorata* era escasa y a la fecha no se ha descrito el genoma nuclear para la especie. Recientemente se publicó el transcriptoma de hoja de un ejemplar del Jardín Botánico

de Nueva York (CEOD-NYBG), el cual fue realizado con fines de su posible utilidad para la identificación taxonómica (Finch *et al.*, 2019). Así mismo, en el capítulo anterior se logró generar el transcriptoma de *novο* del proceso germinativo en semillas de dicha especie. No obstante, los estudios fisiológicos del proceso de germinación de semillas en *C. odorata* son incipientes y pocos son los estudios enfocados a la fisiología de la germinación en especies recalcitrantes e intermedias utilizando secuenciación de última generación (Goh *et al.*, 2019; Silva *et al.*, 2019; Stavrinides *et al.*, 2020). El siguiente capítulo tiene por objetivo estudiar las fases de la germinación de las semillas de *C. odorata* mediante un enfoque transcriptómico analizando la expresión diferencial de los genes partiendo de tres puntos importantes del proceso germinativo: de semillas quiescentes en estado de anhidrobiosis (secas), hidratadas al 95% y germinadas *sensu stricto*. Los resultados proporcionarán información valiosa para los planes de manejo y conservación de las semillas en el almacenamiento de una especie recalcitrante.

2.4 MATERIALES Y MÉTODOS

2.4.1 Cuantificación de transcritos

En el Capítulo I se describe la generación y anotación de un transcriptoma general de semilla de *C. odorata* independientemente de su estado de germinación. Con esta información se realizó la cuantificación de la abundancia de los transcritos en cada condición por triplicado (semilla seca, hidratada al 95 % y germinada). Para ello, se usó el software *Salmon* v.0.14.0 utilizando los archivos *fastq* (R1 y R2) de las nueve muestras. *Salmon* permite estimar expresiones concretas de una forma rápida a través de la inferencia, en el que se utiliza un modelo expresivo con los datos de *RNA-seq*. Tiene la aptitud de corregir los sesgos del contenido de GC que se presentan en los transcritos, característica que hace que aumente la precisión de los estimadores de abundancia, haciendo que los datos de expresión sean confiables (Patro *et al.*, 2017). Los datos de salida de *Salmon* fueron utilizados para generar una matriz de conteo que incluye los estimadores de abundancia y la longitud de los transcritos, los cuales son usados para el análisis de la expresión diferencial. Lo anterior fue realizado con el paquete *Tximport* v.1.22.0 (Soneson *et al.*, 2015).

2.4.2 Análisis de expresión diferencial

La matriz obtenida con *tximport* que proporciona los datos de conteo, fue utilizada como archivo de datos de entrada para el análisis de expresión diferencial. Para el análisis se debe tener en cuenta una condición control, por lo tanto, se compararon las siguientes condiciones: seca vs. embebida (SvsE), embebida vs. germinada (EvsG) y seca vs. germinada (SvsG). Se utilizó el lenguaje de programación R v.4.0.3, a través de la interfaz *RStudio* v.1.3.1093 y el paquete *DESeq2* v.1.30.0 de Bioconductor. Éste último se basa en el recuento de lecturas normalizadas. *DESeq2* utiliza modelos lineales generalizados de distribución Binomial negativa y así mismo los valores *p* se ajustan utilizando el enfoque de Benjamini Hochberg para controlar la tasa de descubrimiento falso (FDR) (Love *et al.*, 2014). El umbral utilizado para considerar un gen diferencialmente expresado (DEG, por sus siglas en inglés) fue obtener un valor *p* ajustado (*padj*) ≤ 0.01 y un valor absoluto *Log2 FoldChange* (cambio relativo en la

expresión) ≥ 2 . Si el valor *Log2FoldChange* es positivo, aumenta el nivel de expresión de ese gen en particular, respecto a la condición anterior, si el valor es negativo, significa que el nivel de expresión se reduce. Del conjunto de genes expresados diferencialmente se construyeron diagramas de Venn para ver las superposiciones en las distintas comparaciones con el programa VENNY v.2.1 (Oliveros, 2015).

Con la finalidad de analizar con detalle los genes que más aumentaron o disminuyeron de expresión en cada una de las comparaciones se llevó a cabo una selección de 100 genes en cada comparación (50 con el valor *Log2Foldchange* más alto y 50 con el más bajo). El resto de los análisis realizados se hicieron considerando únicamente estos genes.

2.4.3 Clasificación de la ontología génica de los DEG

Después de obtener el análisis de los DEG en cada una de las tres comparaciones, se asignó a cada lista los términos GO correspondientes obtenidos con Trinotate, para analizar los grupos funcionales y términos más representativos. Se utilizó WEGO (Web Gene Ontology Annotation Plot) 2.0 para comparar y trazar los resultados de la anotación GO (Ye *et al.*, 2018).

2.4.4 Mapa de calor y anotación individual de los DEG con mayor y menor expresión

Los perfiles de expresión de los 100 DEG seleccionados en cada comparación fueron visualizados a través de un mapa de calor, el cual muestra la medida de la actividad de la expresión diferencial en cada una de las transiciones de forma simultánea. El paquete utilizado fue *pheatmap* v.1.0.12 del paquete *ggplot2* v.3.3.2. Para precisar los datos obtenidos con la anotación masiva obtenida por *Trinotate*, los DEG seleccionados anteriormente en cada condición fueron anotados (descritos) uno por uno por similitud de secuencia usando *BlastX* sobre la base de datos de TAIR (*Arabidopsis Information Resource*, por sus siglas en inglés) (TAIR, 2005) y la base de datos de plantas de NCBI (*National Center for Biotechnology Information*, por sus siglas en inglés) (NCBI, 2013).

2.4.5 Enriquecimiento GO de los DEG con mayor y menor expresión

Debido a la lista grande y redundante de términos GO, se utilizó REVIGO (Supek *et al.*, 2011) para analizar procesos biológicos (PB) y descubrir subgrupos típicos de los términos utilizando un algoritmo de agrupamiento simple basado en medidas de similitud semántica y utilizando la base de datos de *A. thaliana*. Para ello, se usaron los 100 genes que más aumentaron y disminuyeron sus niveles de expresión en cada lista de las comparaciones: SvsE, EvsG y SvsG. Los resultados se visualizaron a través de mapas de árboles.

2.5 RESULTADOS

2.5.1 Cuantificación de transcritos

Se obtuvieron resultados de cuantificación para cada una de las nueve muestras analizadas. Cada archivo de salida (*quant.sf*) contiene la siguiente información en distintas columnas: *Name*, identificador (id) del transcrito; *Length*, longitud del transcrito (número de nucleótidos); *EffectiveLength*, longitud efectiva del transcrito, toma en cuenta la distribución de la longitud de los fragmentos, el sesgo específico de secuencia y del fragmento de GC. *TPM* (Transcritos por Millón, que indica la estimación de la abundancia relativa del transcrito), *NumReads*, que es la estimación del número de lecturas asignadas a cada transcrito que se cuantificó. Esta información fue sintetizada en un solo archivo que contiene una matriz de datos para cada muestra y con ello, se obtuvo la matriz de conteo para los análisis de expresión diferencial.

2.5.2. Análisis de la expresión diferencial

El análisis de expresión mostró que, de los 52,490 genes encontrados en el transcriptoma de semilla, solo el 16 % (8,863) se expresaron diferencialmente durante las tres etapas analizadas del proceso germinativo (**¡Error! No se encuentra el origen de la referencia.**). La comparación de los extremos, es decir, las semillas secas y germinadas (SvsG) tuvo la mayor cantidad de DEG con 7,667 DEG de los cuales, 3,241 aumentaron sus niveles de expresión y 4,426 disminuyeron (**¡Error! No se encuentra el origen de la referencia.**C; Cuadro A4).

Durante el proceso de imbibición, evidenciada por la comparación entre semilla seca e hidratada (SvsE), se encontraron 5,530 DEG, con una mayor cantidad de genes que disminuyen (3,804) su expresión respecto a los que aumentan (1,726) (Figura 2.1A, Cuadro A5). En contraste, durante el proceso de germinación *sensu stricto* (comparación de semilla hidratada al 95 % vs. germinada (EvsG)) se encontraron 1,420 DEG de los cuales 972 aumentaron y 448 disminuyeron (Figura 2.1B, Cuadro A6).

En la Figura 2.1D se puede visualizar el número total de DEG exclusivos y compartidos en cada comparación. La mayor cantidad de genes únicos (2,445) se observó en la

comparación de SvsG, seguido por la comparación SvsE y EvsG con 874 y 189 respectivamente, mientras que 399 se compartieron durante todo el proceso

Cuando se analizó solamente a los DEG que aumentaron su expresión, se observó que 155 se comparten en las tres comparaciones: 317, 245 y 1,260 fueron únicos en las comparaciones SvsE, EvsG y SvsG respectivamente, mientras tanto 1,254 se compartieron solo entre SvsE y SvsG, y 572 entre EvsG y SvsG (**¡Error! No se encuentra el origen de la referencia.E**). Respecto al conjunto de genes que disminuyen su expresión, 233 se comparten en las tres comparaciones, mientras que 693, 85 y 1,185 son únicos en la comparación SvsE, EvsG y SvsG respectivamente. Finalmente 2,878 se comparten en SvsE y SvsG, y 130 se comparten entre EvsG y SvsG (**¡Error! No se encuentra el origen de la referencia.F**).

2.5.3 Clasificación de la Ontología Génica

Los 8,863 DEG fueron analizados a través de la ontología génica (GO) para la interpretación biológica y descripción de sus funciones y se distribuyeron en tres grupos: componentes celulares (CC), función molecular (FM) y procesos biológicos (PB). En la categoría CC, predominaron los términos: “célula”, “parte de la célula”, “parte intracelular”, “intracelular”, “organelo”, “organelo intracelular” y “orgánulo delimitado por membrana”. En la categoría FM sólo se encontraron cuatro términos de los cuales, la “actividad catalítica” tuvo mayor dominancia, seguido de “actividad oxido reductasa”, “actividad de la molécula estructural”, y “constituyente estructural del ribosoma”. Finalmente, la categoría PB tuvo con mayor representatividad a los “procesos metabólicos primarios”, “regulación biológica”, “regulación del proceso biológico” y “regulación del proceso celular” (Figura 2.2A). Cuando se analizó de forma estadística los términos de los PB se encontró que los más significativos fueron: “desarrollo de estructuras anatómicas” (GO:0048856), “proceso de desarrollo” (GO:0032502), “proceso metabólico de molécula pequeña” (GO:0044281), “división celular” (GO:0051301), “desarrollo post-embionario”(GO:0009791) entre otros (Figura 2.2B)

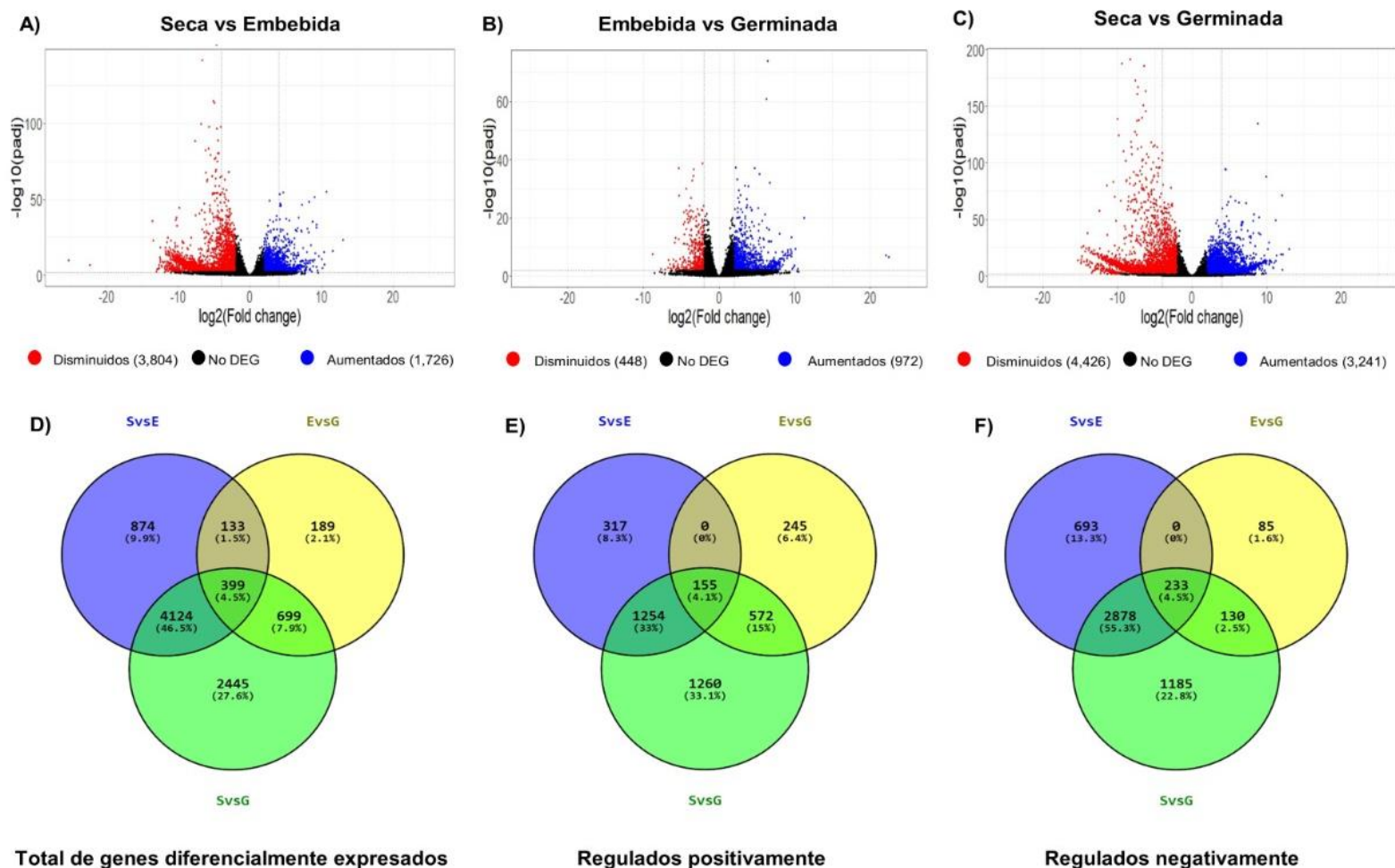


Figura 2.1. Análisis del número de genes expresados diferencialmente (DEG), identificados en tres etapas de germinación de semillas en *C. odorata*. A) Diagrama de volcán que muestra el patrón de expresión entre semillas secas contra embebidas (SvsE), B) embebidas c contra germinadas (EvsG) y C) secas contra germinadas (SvsG). El eje X representa el valor $Log_2FoldChange$ y el eje Y el nivel de significación según $-\log_{10}(padj)$. Los puntos azules y rojos representan los DEG regulados en forma positiva y negativa, respectivamente, los puntos negros representan genes que no se expresaron de manera diferencial. D) Diagrama de Venn que muestra superposiciones de DEG totales; E) superposición de los genes regulados positivamente y F) genes regulados negativamente. Los números entre paréntesis indican el porcentaje de DEG en cada situación de los diferentes grupos.

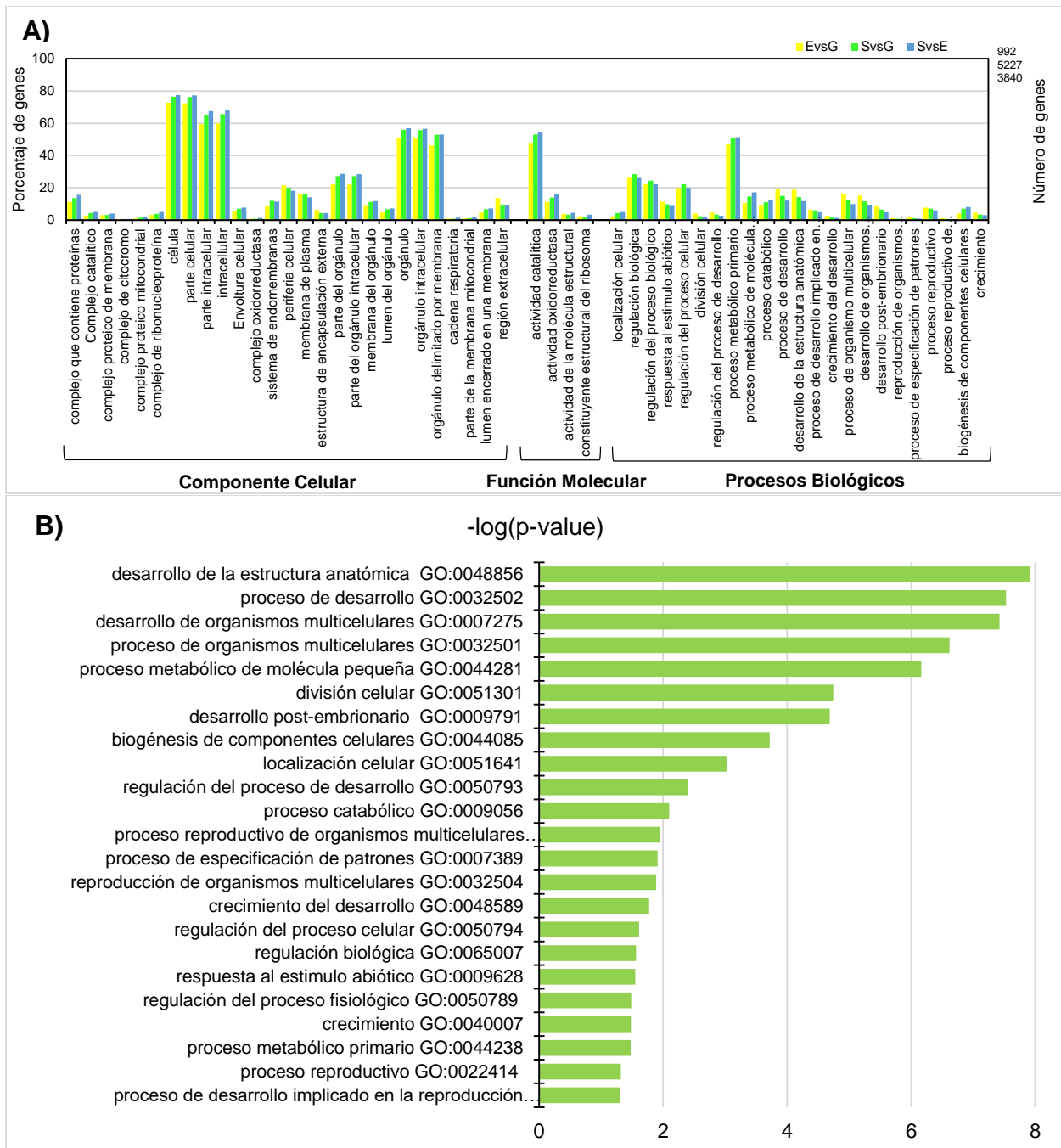


Figura 2.2. Ontología genética (GO) de los genes expresados diferencialmente (DEG) en las comparaciones: seca vs embebida (SvsE), embebida vs germinada (EvsG) y seca vs germinada (SvsG). **A)** Clasificación funcional de los términos GO. El eje X representa los términos organizados en tres categorías: componente celular, función molecular y procesos biológicos. El eje Y, en el lado izquierdo, muestra el porcentaje de genes anotados en cada subgrupo de cada categoría y en el lado derecho, el número de genes anotados en cada comparación. **B)** Diferencias significativas en función del $\log_{10} p\text{-value}$ ($<0,05$) de los procesos biológicos analizados, organizados en orden descendente.

2.5.4 Mapa de calor y anotación individual de los DEG con mayor y menor expresión

De la selección de los 300 genes con mayor y menor expresión (100 para cada comparación (SvsE, EvsG y SvsG)) se encontraron 53 que se compartieron en al menos dos comparaciones, por lo que el número de genes para analizar y anotar de forma individual fue de 247. Estos se visualizaron en un mapa de calor en el que se exponen 12 grupos acorde a su patrón de expresión (Figura 2.3A).

El comportamiento de cada grupo en las tres etapas analizadas se visualizó a través de graficas que muestran el cambio relativo en la expresión. Cada grupo tuvo diferente número de genes, obteniendo un mínimo de dos y un máximo de 44 (Figura 2.3B). De los 247 genes analizados, los grupos 1, 2 y 3 muestran a los 50 que reducen significativamente su expresión al pasar de semilla seca a hidratada, mientras que los grupos 9, 11 y 12 muestran a los que aumentan en esta misma condición. Probablemente estos 100 genes se encuentran asociados a la quiescencia. Por su parte, durante la germinación *sensu stricto*, 50 genes de los grupos 4, 10 y 12 disminuyen su expresión mientras que otros 50 de los grupos 1, 6 y 8 aumentan. Estos genes probablemente estén involucrados en el proceso de germinación *sensu stricto*. Por otra parte, cuando se analiza la comparación entre los extremos del proceso se observan 28 genes que disminuyen gradualmente (grupo 5) mientras que 25 van aumentando (grupo 7).

Con las bases de datos de TAIR y NCBI, de la anotación individual por alineamiento local y similitud (BLAST) de las 247 secuencias se encontró que solo el 47% tiene una función putativa coincidente en ambas bases de datos, mientras que el 4% corresponde a secuencias desconocidas, no localizadas en ninguna de las bases. El resto, (49 %) se encontró como proteína no caracterizada, hipotética, predicha, o no coincidían en la anotación en ambas bases y/o la mayoría mantenía valores-*E* no significativos o un porcentaje de identidad bajo (Cuadro A7).

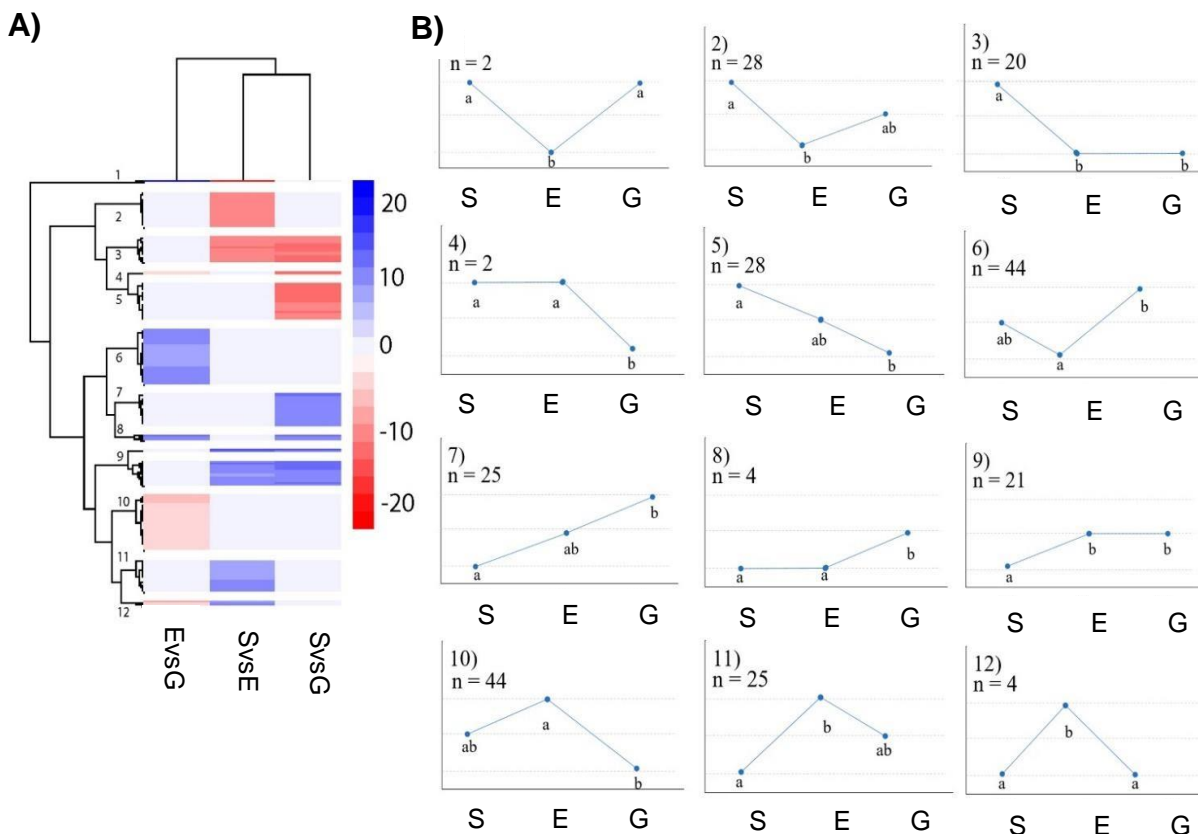


Figura 2.3. A) Mapa de calor que muestra el agrupamiento y patrón de expresión de 247 genes expresados diferencialmente (DEG), según el valor \log_2 FoldChange. Cada fila representa un gen y cada columna la comparación en la que se expresa el gen: seca vs embebida (SvsE), embebida vs germinada (EvsG) y seca vs germinada (SvsG). Los colores rojo y azul indican DEG regulados al alza y a la baja, respectivamente, mientras que el blanco indica que no hubo cambios. Cuanto más fuerte es el color, mayor es la expresión. **B)** Gráficos hipotéticos que representan la tendencia en el comportamiento de los doce grupos representados en el mapa de calor. Las letras S, E y G, indican las condiciones de tratamiento: seca, embebida y germinada respectivamente, n, indica el número de genes en ese grupo.

Dentro de los genes caracterizados (mostrados en el mapa de calor) que reducen en menor medida su expresión durante la hidratación de las semillas se encontraron genes que codifican para “proteína transportadora de azúcar 7” (*STP7*), “alfa-amilasa 3” (*AMY3*), “fosfoglicerato quinasa 2” (*PGK2*), “superóxido dismutasa 2” (*SOD2*), “proteína de la superfamilia de oxidoreductasa ligada a NAD (P)”, “citocromo P450 familia 705” (*CYP705A22*), “acil-CoA desaturasa” entre otros. Mientras tanto, dentro de los que aumentan su expresión en esta misma condición se encuentran: “acc oxidase1” (*ACO1*), “familia de proteínas regulada por la giberelina” (*GSA6*), genes señalizadores

como el gen “clavata3/esr-relacionado 4” (*CLE4*), distintas peroxidasas, genes relacionados con la patogénesis, genes involucrados en la organización de la pared celular como los xiloglucanos, xilosidasas, endo-beta-mananasas, también se encontró a una familia de proteínas que responden al súper complejo respiratorio, “sulfotransferasa 2A” (*ST2A*), “proteína de la superfamilia de albúmina 2S de almacenamiento de semillas”, “Oxidasa de cisteína vegetal 2” (*PCO2*) que responden al etileno, “HECATE 2” (*HEC2*) que codifica para un factor de transcripción, entre otros (Cuadro A7) .

Por otro lado, dentro de los genes caracterizados que disminuyeron su expresión durante la germinación *sensu stricto* se encuentran nuevamente el gen *CLE4*, “Citocromo oxidasa 1”(COX1), proteínas HSP, proteínas sensibles a auxinas, “proteína de la superfamilia del dominio DnaJ chaperona”, algunos genes relacionados con la patogénesis, “proteínas de la familia del retículo”, “aminopeptidasa M1” (*APM1*), entre otros. Y dentro de los que aumentaron se encuentra una vez más *STP7*, “gamma-glutamyl transpeptidasa 1” (*GGT1*), una “proteína del dominio BURP dirigida a las vacuolas” (*USPL1*), un miembro de la familia de genes “lacasa 7” (*LACCASE 7*), *BZIP61* que codifica un miembro de la familia *BZIP* de factores de transcripción, proteínas de membrana, “pectinas metilesterasa inhibidor 13” (*PMEI13*), genes que codifican para tubulina y actina, “proteína 18 de dedos de Zinc de caja B”, (*BBX18*), “peroxidasa 3” (*PER3*), “subunidad O del fotosistema I” (*PSAO*), y otros (Cuadro A7).

Dentro de los genes caracterizados en la comparación de los extremos (SvsG) los genes significativos que disminuyeron gradualmente durante todo el proceso de geminación fueron: “Proteína de la superfamilia de pliegues de Rossmann que se une a NAD (P)”, “proteína de la familia de las oxidoreductasas de glucosa-metanol-colina” (*GMC*) y “proteína ribosómica L7Ae/L30e/S12e/ proteína de la familia Gadd 45” (*RPL30B*). Mientras que en los que aumentaron gradualmente se encuentran: un gen que codifica a una “proteína de la superfamilia de las peroxidasas”, diversos genes relacionados con la síntesis de la pared celular, “proteína de transferencia de lípidos 1” (*LTP1*), nuevamente “proteína de la superfamilia de albúmina 2S de almacenamiento de semillas”, “cinamil-alcohol deshidrogenasa 7” (*CAD7*), nuevamente *PMEI13* y

“pectina metilesterasa 41” (*PME41*), “fosfatasa acida 3” (*PAP3*), “inducible de etileno y sal 3” (*ESE3*) y otros (Cuadro A7) .

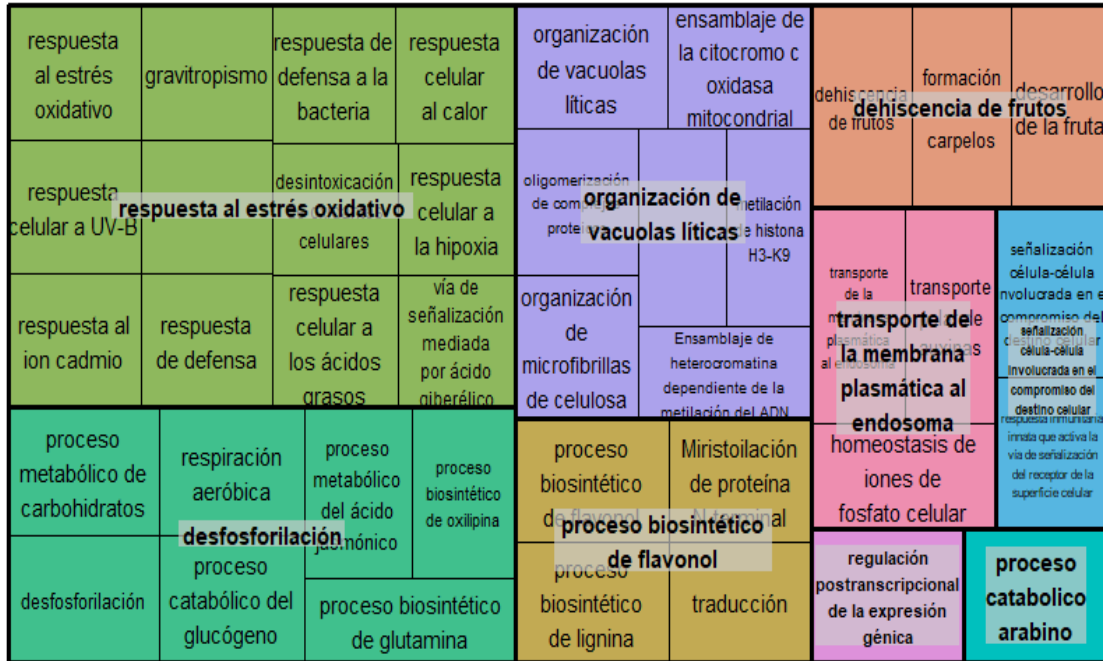
2.5.5 Enriquecimiento de los términos GO de los DEG con mayor y menor expresión

Los procesos biológicos de los 100 genes seleccionados por su mayor y menor expresión diferencial en cada comparación se analizaron a través del enriquecimiento de términos GO usando REVIGO. El proceso de hidratación de las semillas se enriqueció con 40 términos (Cuadro A8) que se organizaron en 9 grupos. El grupo con mayor frecuencia fue el de "respuesta al estrés oxidativo" (GO:0006979), seguido del grupo "desfosforilación" (GO:0016311), y "organización de las vacuolas líticas" (GO:0080171) (Figura 2.4A).

Durante la germinación *sensu stricto* se encontraron 15 términos (Cuadro A9) reunidos en 7 grupos. Aquellos que agruparon el mayor número de genes fueron "respuesta a la auxina" (GO:0009733), "regulación negativa de la actividad catalítica" (GO:0043086), además de otros grupos relacionados con el transporte de energía, fotosíntesis, así como procesos relacionados con el crecimiento de la plántula (Figura 2.4B).

La comparación de los extremos (SvsG) se enriqueció con 34 términos (Figura A1) y 7 grupos, algunos similares a la comparación SvsE, sin embargo, aquí también se encontraron otros términos interesantes como: "respuesta celular al óxido nítrico" (GO:0071732), "vía de señalización activada por etileno" (GO:0009873) y "desintoxicación de oxidantes celulares" (GO:0098869), todos reunidos en el grupo de "respuesta celular a UV-C" (GO:0071494) (Cuadro A10).

A)



B)



Figura 2.4. Análisis REVIGO de los procesos biológicos en 100 genes diferencialmente expresados (DEG), regulados al alza y a la baja en las comparaciones: **A)** Seca vs. embebida (SvsE), y **B)** embebida vs. germinada (EvsG). Cada rectángulo representa un grupo. Los representantes se clasifican en "supergrupos" de términos vagamente relacionados, reflejados por diferentes colores. La frecuencia del término GO se refleja en el tamaño de los rectángulos.

2.6 DISCUSIÓN

El análisis comparativo del transcriptoma de semillas de *C. odorata* en distintos momentos fisiológicos permitió evidenciar dos etapas en la transición de una semilla quiescente hacia la germinación, la primera conduce a la reactivación metabólica como resultado de la imbibición, seguida por los procesos que apoyan la emergencia de la radícula.

De acuerdo con el análisis de ontología génica, en la categoría PB se mostraron cambios significativos en 23 procesos que permiten ver la actividad metabólica reactivada en las semillas posterior al proceso de imbibición, además del crecimiento embrionario (Figura 2.2B). Dentro de la categoría CC se activaron diversos términos relacionados con la actividad celular y en la FM se mostró con mayor frecuencia procesos relacionados con la velocidad de las reacciones químicas como es la catálisis (Figura 2.2A). Estos mismos términos también se encontraron como los más representativos en otros estudios enfocados al análisis de la germinación de semillas en la especie *Punica granatum* L., (FU *et al.*, 2021) y *Calanthe tsoongiana* Tang & F.T.Wang (Jiang *et al.*, 2021). Por lo que podría decirse que estas características son consistentes en el proceso de germinación de semillas de algunas especies desarrolladas en climas tropicales.

2.6.1 Reactivación metabólica durante la hidratación de la semilla

Durante la imbibición de las semillas (SvsE) se observó que de los 5,530 DEG detectados, cerca del 70 % disminuyeron su expresión. Esto puede relacionarse a que las semillas secas conservan una gran cantidad de ARNm almacenado remanente de las etapas tardías del desarrollo embrionario y maduración seminal que corresponden a un programa genético ligado a mantener la quiescencia, por lo que se requiere disminuir su expresión o de su degradación para lograr la transición hacia la reactivación metabólica ligada a la germinación (Bewley *et al.*, 2013; Bai *et al.*, 2020).

Un ejemplo de lo anterior lo constituyen enzimas involucradas en los procesos energéticos y de señalización como los genes que codifican para la familia de “carbohidrato quinasa” (*FGGY*, AT4G30310.2) y *PGK2* (AT1G56190.2). Por ello, el

término GO “desfosforilación” (GO:0016311) se muestra como uno de los de mayor frecuencia (Figura 2.4A). La familia *FGGY* lleva a cabo la fosforilación dependiente de ATP de distintos azúcares que van desde las triosas hasta heptosas (Zhang *et al.*, 2011). En lo que respecta a las PGK, se asocian a diversos procesos como la glucólisis donde convierten el 1,3-bisfosfoglicerato en 3-fosfoglicerato, en la gluconeogénesis y el ciclo de Calvin-Benson (Rosa-Téllez *et al.*, 2018). La isoforma PKG2 participa en el metabolismo fotosintético del carbono (Massange-Sánchez *et al.*, 2020). Así mismo, PKG2 ha sido localizada en semillas de *Arabidopsis* durante la germinación pudiendo ser inducida por estrés salino (Liu *et al.*, 2015) lo que conjuntamente sugiere una posible función en respuesta a la desecación; además se ha identificado su expresión en hojas en la etapa de plántula, involucrada en la glucólisis del plastidio (Rosa-Téllez *et al.*, 2018).

Nuestros resultados no nos permiten saber con precisión cual es la función real de la familia *FGGY* y *PKG2* en las semillas de *C. odorata* y porqué son regulados en forma negativa. Sin embargo, podría tratarse de un ajuste metabólico; estudios con mutantes para las isoformas *PKG1* y *PKG3* en *Arabidopsis* mostraron que ambos disminuyen su expresión, regulando a la baja la actividad fotosintética, lo que podría tratarse de un mecanismo estratégico de la planta cuando la glucólisis está siendo afectada, logrando con ello un ajuste metabólico y optimizar el crecimiento de la planta bajo estrés (Rosa-Téllez *et al.*, 2018).

Es posible que la baja expresión de los genes mencionados anteriormente repercuta de igual manera en otros como los que codifican para el “transportador de azúcar *STP7*” (AT4G02050.2) y la proteína *AMY3* (AT1G69830.1). La función de *STP7* es transportar a las pentosas L-arabinosas y D-xilosas y se ha encontrado en tejidos con alto recambio de la pared celular, por lo que podría relacionarse con la captación y reciclaje de azúcar (Rottmann *et al.*, 2018). Mientras que algunas alfa amilasas como la *AMY3* incentivan el catabolismo del almidón almacenado para nutrir la planta en desarrollo y se ha encontrado regulado en forma negativa por el azúcar en los embriones y positivamente por la giberelina durante la germinación en arroz (Damaris *et al.*, 2019). El descenso en la expresión de un gen que codifica para proteína de la

familia piruvato quinasa, también se observa en semillas de *Arabidopsis* durante la hidratación de las semillas (Bai *et al.*, 2017).

Aunado a lo anterior, dentro del mismo grupo “desfosforilación” (GO:0016311) se encontraron cambios en la expresión de genes asociados al metabolismo energético y transporte reflejados en el término GO “procesos metabólicos de los carbohidratos” (GO:0005975) (Figura 2.4A). La energía es importante para llevar a cabo el proceso de germinación. En particular, se encontró el aumento de expresión de varios genes como: *COX1* (ATMG01360.1); que forma el complejo IV de la cadena de transporte de electrones mitocondrial (Senkler *et al.*, 2017); el “dominio de gen 3 inducido por hipoxia” (*HIGD3*, AT3G05550.1) que se cree codifica subunidades específicas del complejo mitocondrial citocromo c oxidasa en *Arabidopsis* y está relacionado con la respuesta a hipoxia (Hwang *et al.*, 2017); o bien, la “proteína de la superfamilia de pliegues Rossmann” (AT2G47120.2) que se une a NAD (P) y que se ha descrito como un motivo estructural muy conservado. Los NAD (P) también juegan un papel importante en la cadena transportadora de electrones (Hanukoglu, 2015). Dentro de este grupo también se localizó una “terpeno sintasa 02”, (*TPS02*, AT4G16730.2) que está relacionada con la biosíntesis de los terpenoides, quienes están involucrados con el metabolismo primario y secundario de las plantas (Chen *et al.*, 2021).

Por otro lado, entre de los genes que aumentaron su expresión se encontró a *CLE4*. Los genes *CLE* codifican para péptidos cortos de señalización. *CLE3* ha sido mejor caracterizado y consta de 12 a 13 aminoácidos. Los genes *CLE* son indispensables durante el desarrollo y crecimiento vegetal en diversos tejidos meristemáticos (Qin *et al.*, 2021). Por lo tanto, es posible que *CLE4* esté actuando como una molécula señalizadora para dar paso a los procesos implicados en el crecimiento del embrión en *C. odorata* solo durante el proceso de hidratación de las semillas, pues una vez que se pasa a la germinación *sensu stricto* está disminuye su expresión (Cuadro A7).

En esta comparación, la “respuesta al estrés oxidativo” (GO:0006979) juega un papel preponderante al ser el término GO detectado con mayor frecuencia (Figura 2.4A). El estrés oxidativo es generado por la formación y acumulación de especies

reactivas de oxígeno (ROS), principalmente de anión superóxido ($O_2^{\cdot-}$), peróxido de hidrógeno (H_2O_2) y radical hidroxilo ($\cdot OH$) (Bailly, 2019). En semillas el sitio principal de ROS es la mitocondria, sin embargo, durante la germinación se han localizado en el citoplasma, núcleo y pared celular (Bailly, 2019). No obstante, en semilla las ROS juegan un papel dual, por un lado, tienen un efecto negativo al ocasionar la oxidación de los sistemas membranales y macromoléculas (ADN, ARN y proteínas). Sin embargo, también actúan como moléculas de señalización importantes para romper la latencia y dar paso a la germinación (Lariguet *et al.*, 2013; Bailly, 2019; Kurek *et al.*, 2019), regulando la señalización por el ácido giberélico (AG), así como la movilización de las proteínas de almacenamiento (Kurek *et al.*, 2019). Dichas ROS son activadas por peroxidasa tipo III (Lariguet *et al.*, 2013; Renard *et al.*, 2019), tal es el caso de las Peroxidasas 4 (*PER4*, AT1G14540.1) y 52 (*PRX52*, AT5G05340.1) (Cuadro A7) que en este trabajo y durante la etapa de imbibición aumentaron su expresión. Por ello, para contener con su acumulación a fin de mantener una homeostasis REDOX, las células contienen sistemas enzimáticos y no enzimáticos de detoxificación. Cuando estos sistemas son dañados o sobrepasados por los ROS producidos se aceleran los procesos de daños y envejecimiento celular, que en las semillas lleva a una pérdida del vigor y finalmente de la viabilidad de estas (Verdier *et al.*, 2013; Sano *et al.*, 2016).

Al respecto, durante el paso de la semilla seca a hidratada (SvsE) en *C. odorata* se detectó cambio en la expresión de una “superóxido dismutasa cloroplástica de cobre/zinc” (*CSD2*, XP_023886042.1), enzima antioxidante que ayuda a la regulación de la sobreproducción de radicales libres generada por el reinicio del metabolismo (Sano *et al.*, 2016). Sin embargo, este transcrito en particular bajó sus niveles de expresión, lo que indica que podría estar participando en la acumulación de ROS para la liberación de la latencia en las semillas, lo cual sucede cuando los sistemas antioxidantes disminuyen el nivel de expresión (Bailly, 2019). No obstante, es posible que corresponda a un gen cuya transcripción sea importante durante la maduración de la semilla y al dar paso a la reactivación metabólica para la germinación esta tenga que disminuir su expresión o ser degradado. En *G. mangostana* (otra especie cuyas semillas están clasificadas como semilla recalcitrantes) la superóxido dismutasa y *ACC oxidasa* aumentan su nivel de expresión y la peroxidasa disminuye (Goh *et al.*, 2019),

mientras que en este estudio sucede lo contrario y solo coinciden en el aumento de *ACC oxidasa*.

Además, durante la imbibición (SvsE) en semillas de *C. odorata* se detectaron cambios en la expresión de genes relacionados con la liberación de la latencia de las semillas. Particularmente, aumentó la presencia de transcritos de *GSA6* (AT1G74670.1), cuya proteína en *Arabidopsis* (Zhong *et al.*, 2015) participa en la señalización del AG, ácido abscísico (ABA) y glucosa, regulando la germinación a través del alargamiento celular. También se observó un aumento en la expresión de *ACO1* (XP_023885201.1) que agrupa señales para la biosíntesis del etileno y es importante para el desarrollo normal de la raíz (Park *et al.*, 2018). La concentración de etileno en las semillas está relacionada con su capacidad germinativa, por ello, se considera que el etileno está implicado en la regulación de la germinación y latencia de las semillas (Li *et al.*, 2020). También se detectó un aumento de *ST2A* (AT5G07010.1), que participa en el metabolismo del ácido jasmónico (Yamashino *et al.*, 2013), y el jasmonato regula la germinación de semillas a través de la señalización de ABA (Yu *et al.*, 2021).

Los términos GO “organización de vacuolas líticas” (GO:0080171) y el “transporte de la membrana plasmática al endosoma” (GO:0048227) también se presentaron como términos de mayor frecuencia durante esta etapa (Figura 2.4A). Las vacuolas líticas funcionan como degradadoras generales y son muy importantes en la germinación al descomponer a las proteínas de almacenamiento, a fin de movilizar los nutrientes necesarios para sostener el alargamiento del eje embrionario durante el proceso de germinación, por lo que existe un tráfico de proteínas desde la red trans-golgi hacia las vacuolas a través de un grupo de endosomas tardíos (Feraru *et al.*, 2010). Esto es consistente con el aumento observado en la expresión del grupo de “proteína de almacenamiento 2S” (AT2G10940.2) durante la hidratación de las semillas (SvsE). Estas proteínas se caracterizan por ser ricas en arginina, glutamina, cisteína y metionina y se utilizan como fuente de nitrógeno, azufre y carbono durante la germinación de la semilla y hasta el crecimiento de la plántula (Souza, 2020). También hubo un aumento en la expresión de una “Endo-Beta-Manasa 7” (*MAN7*,

AT5G66460.1). Las endo-beta-manasas están relacionadas con la degradación de los galactomananos los cuales son polisacáridos de reserva y están presente en las paredes de algunas semillas, como en la leguminosa *Sesbania virgata* (Cav.) Pers. (Lisboa *et al.*, 2006). En la germinación de tomate, la endo-beta-manasa es considerada un marcador de la germinación (Moles *et al.*, 2019). En *C. arabica* también se localizó antes y después de la emergencia de la radícula (Silva *et al.*, 2019). El homólogo de este gen en *Populus trichocarpa*, PtrMAN7 se ha encontrado expresado en hojas jóvenes y en la punta del brote (Zhao *et al.*, 2013).

Dentro de “procesos catabólicos de arabinosa” (GO:0031222) se localizan genes que se encuentran relacionados a la pared celular, mismos que también aumentaron en la comparación SvsG. Por ejemplo, se incrementó la expresión del gen “Xiloglucano endotransglucosilasa/hidrolasa” (*XTH*). El xiloglucano es una hemicelulosa, que se une de forma covalente a la celulosa y une microfibrillas de celulosa adyacentes constituyendo la red de sostén de la pared (Carrillo *et al.*, 2013) los embriones de algunas dicotiledóneas y monocotiledóneas (Bewley *et al.*, 2013). Los *XTH* rompen y reforman estos enlaces β -(1,4) glicosídicos en las moléculas de xiloglucano, probablemente acomodando la reestructuración de la pared celular asociada con el ablandamiento y la expansión del embrión (Bewley *et al.*, 2013; Carrillo *et al.*, 2013; Cao *et al.*, 2021). En este estudio, se encontró modificación en la expresión de los genes “xiloglucano endotransglucosilasa/hidrolasa 9” (*XTH9*, AT4G03210.1) y “xiloglucano endotransglucosilasa/hidrolasa 16” (*XTH16*, AT3G23730.1). *XTH9* también se localizó en *G. mangostana*, pero en el establecimiento de la plántula durante el alargamiento celular (Goh *et al.*, 2019). Del mismo modo, una “Beta-d-xilosidasa 4” (*BXL4*, AT5G64570.1) aumentó sus niveles de expresión al pasar de semilla seca a hidratada. Las beta-xilosidasas son una familia de genes que también han sido encontradas en *Arabidopsis* y específicamente *BXL4* ha estado asociada al desarrollo de la pared celular (Goujon *et al.*, 2003).

El término “proceso biosintético de flavonoles” (GO:0051555) también estuvo preferentemente representado entre los términos GO de los procesos biológicos (Figura 2.4A). Al respecto, el gen “flavonol sintasa 6” (*FLS6*, AT5G43935.1) cuya

anotación arrojó un *e-value* de 2.6 y un porcentaje de identidad de 32 se encontró presente aumentando en sus niveles de expresión desde la etapa SvsE y SvsG. Su función se asocia a la síntesis de flavonoides, los cuales tienen diversas funciones como la eliminación de radicales libres, protección de células y tejidos ante el estrés oxidativo y respuesta al estrés biótico. Durante la germinación de semillas de *Punica granatum* L. también se detectó la presencia de flavonoides (FU *et al.*, 2021). En el maní, las semillas maduras y secas tuvieron un mayor contenido de *FLS6*, pero con mayor cantidad durante el desarrollo de la semilla (Hou *et al.*, 2020).

2.6.2 Expansión celular y germinación *sensu stricto*

En contraste al comportamiento de una semilla seca a hidratada, a medida que avanza la imbibición y se transita a la germinación (EvsG) se revierte la tendencia/relación con el incremento del número de genes que aumentan su expresión en casi el 70 %. Esto es consistente con lo reportado en el análisis transcriptómico de la germinación en semillas de *Punica granatum* L. en donde en esta misma etapa de preparación para la germinación, aumentan considerablemente la cantidad de genes expresados (FU *et al.*, 2021).

La transición de una semilla embebida a germinada (EvsG) evidenció una segunda etapa en el proceso germinativo, donde se observa que en su mayoría los términos GO encontrados están relacionados con el crecimiento del embrión. Por tanto, el término con mayor frecuencia es el de “respuesta a auxina” (GO:0009733) (Figura 2.4B), que se relaciona con diversos procesos del desarrollo de la planta como la división, crecimiento y diferenciación celular (Blázquez *et al.*, 2020). En el presente estudio se detectó al gen “Ácido indol-3-acético inducible 14” (*IAA14*, AT4G14550.1) aumentando su expresión. El “ácido indol-3-acético” (IAA) es un regulador de crecimiento y desarrollo de las plantas de la clase de las auxinas, asociada a distintos procesos fisiológicos como división celular, dominancia apical, iniciación de raíces laterales y otros. En *Pinus sylvestris* L., el *IAA* se relaciona al proceso de germinación durante la imbibición de las semillas (Ljung *et al.*, 2001). “Ácido indol-3-acético Inducible 9” (*IAA9*) y el “Ácido indol-3-acético inducible 4” (*IAA4*) también se han encontrado participando en la germinación de semillas de *C. arabica* (Silva *et al.*,

2019). Por lo que se sugiere que *IAA14* también incida en la división celular durante la germinación de la semilla en *C. odorata*.

Así mismo, se sobre expresaron dos genes sensibles a auxina, uno que codifican para SAUR (ARN pequeño de auxina), específicamente “ARN regulado al alza de auxina pequeña 46” (*SAUR46*, AT2G37030.1) y otro para el factor de transcripción “teosinte ramificado1/cycloidea/PCF 15” (*TCP15*, AT1G69690.1) que activa directamente a los genes de la subfamilia “ARN regulado al alza de auxina pequeña 63” (*SAUR63*). Se considera que tanto las proteínas como los transcritos de *SAUR* tienen una vida útil corta lo que provoca que su actividad disminuya con rapidez después de la eliminación del estímulo de la auxina permitiendo respuestas dinámicas (Stortenbeker y Bemer, 2019). Esto podría explicar el comportamiento de estos dos genes que tienen un pequeño aumento durante la hidratación de la semilla y posteriormente bajan durante la germinación *sensu stricto*. En *Arabidopsis* se reporta que *SAUR46* aumenta su expresión, pero en niveles bajos en plántulas (Wang *et al.*, 2020a). La sobreexpresión de *SAUR63* en plantas transgénicas de *Arabidopsis* resultó en la obtención de hipocótilos más largos y un mayor transporte de auxina en dichos hipocótilos (Chae *et al.*, 2012). En semillas de *Paeonia lactiflora* Pall, distintos miembros de *SAUR* se expresaron en forma diferencial durante la liberación de la latencia (Li *et al.*, 2020).

En esta etapa también se observan cambios en la expresión de transcritos del dominio DUF581 (AT3G63230.1), el cual a pesar de que ha sido poco estudiado, se ha encontrado en *Arabidopsis* regulado por ABA y asociado al mantenimiento de la latencia (Jamsheer y Laxmi 2014), y en este trabajo y durante el paso de la semilla embebida a germinada disminuyó su expresión posiblemente para dar paso a los procesos que conllevan a la germinación.

De acuerdo con los términos GO, entre los genes con mayor cambio en sus niveles de expresión hubo un descenso de aquellos asociados a las “actividades catalíticas” (GO:0043086); sin embargo, aumentó la expresión de genes de los procesos de requerimiento de energía como el “transporte transmembrana de ATP mitocondrial” (GO:1990544). Así mismo, se expresaron términos GO relacionados con

la expansión celular y la formación de placas celulares como lo es la “organización del citoesqueleto de actina” (GO:0030036) y los procesos basados en microtúbulos (GO:0007017) (Figura 2.4B). Al respecto, algunos genes que aumentaron su expresión en esta etapa (EvsG) fueron Actina-11 (*ACT11*, AT3G12110.1), alfa-tubulina 2,4 (*TUA2*, AT1G50010.1) y una cadena de tubulina beta-2/beta-3 (*TUB3*, AT5G62700.1) (Cuadro A7). Así mismo, como parte de los procesos que dirigen el alargamiento celular se encontró la presencia de un gen que codifica para la “proteína 18 de dedos de Zinc de caja B” (*BBX18*, AT2G21320.1) y en *P. granatum* se cree que esta proteína participa en la regulación de la apertura del cotiledón de la semilla en la germinación (FU *et al.*, 2021).

También se observó la presencia del término GO “recolección de luz en el fotosistema 1” (GO:0009768) (Figura 2.4B), que se relaciona con la expresión del gen *PSAO*, (AT1G08380.1) que aumento su transcripción en esta etapa. La expresión de este gen en el proceso de germinación de las semillas de *C. odorata* podría ser parte de la preparación de la plántula-planta para su vida autótrofa. Sin embargo, en algunas semillas oleosas la presencia de clorofila en semillas maduras es una característica indeseable que provoca reducción del vigor y longevidad de la semilla (Zinsmeister *et al.*, 2020). Por otro lado, en algunas angiospermas, el cloroplasto de semillas en desarrollo *cloroembriófitas* (con presencia de clorofila en el embrión) actúa en la activación de la fotofosforilación cíclica y, por lo tanto, de un aumento de la síntesis de ATP. Durante la última etapa de desarrollo las semillas entran en un estado de latencia donde ocurre deshidratación y disgregación del aparato fotosintético, degradación del cloroplasto y transformación de cloroplastos a plastidios llenos de sustancias nutritivas. Sin embargo, el cloroplasto no se desintegra por completo y ayuda a la prevención de la hipoxia manteniendo la respiración en las semillas maduras (Smolikova y Medvedev, 2016). Por lo tanto, sería importante analizar más a fondo cual es el papel que juega esta ruta en el proceso de germinación de las semillas de *C. odorata* para poder determinar si ayuda al mantenimiento de la latencia o bien participa en la reducción de su longevidad.

Durante la transición de una semilla hidratada hacia la germinación se caracterizó el descenso de varias proteínas que en otros estudios han sido asociadas a la tolerancia a la desecación, tal es el caso de los genes que codifica para “galactinol sintasa 2” (*GOLS2*, AT1G56600.1) (Giarola *et al.*, 2017), la “proteína de choque térmico 70” (*HSP70*, AT3G09440.4, AT3G12580.1), “proteína de choque térmico 70 B” (*HSP70B*, AT1G16030.1) y el gen que codifica para la “proteína de la superfamilia del dominio DnaJ chaperona” (AT4G19570.1). El descenso de genes HSP después de la imbibición de las semillas, también ocurrió en arroz (Wei *et al.*, 2015), sin embargo, fue una “proteína de choque térmico 20” (*HSP20*) y en este estudio dicho gen (AT4G21870.1) se activó en forma positiva durante el paso de la semilla SvsE al igual que ocurrió en semillas de *Arabidopsis* (Bai *et al.*, 2017).

Las HSP han sido identificadas en distintos procesos como la longevidad de la semilla, el plegamiento correcto en proteínas no sintetizadas, plegamiento de las proteínas alteradas durante el almacenamiento de semillas y protección de las semillas contra el daño oxidativo (Sano *et al.*, 2016). Dos de estos genes han sido relacionados con la protección de organismos implicados en el estrés por calor, *HSP70* (Su y Li, 2008; Wang *et al.*, 2020b) y *Dnaj* (Fan *et al.*, 2017), mismos que se identificaron en los embriones de *C. arabica*, aumentando su expresión durante la germinación (Silva *et al.*, 2019). *HSP70B* se ha encontrado en algas verdes durante y después de la inhibición participando en la protección y reparación del fotosistema II (Su y Li, 2008). En general las proteínas HSP junto a las proteínas LEA y los azúcares promueven a que las semillas pasen del estado líquido cristalino al estado vítreo, aumentando sus niveles de expresión a medida que aumenta la maduración y secado de las semillas, por lo que se relacionan con la tolerancia a la desecación (Bewley *et al.*, 2013). Por lo tanto, durante esta etapa en el que la semilla de *C. odorata* se prepara para la germinación, estos genes deben regularse negativamente a fin de llegar a la protrusión de la radícula.

En esta misma etapa también se expresó el gen “gamma-glutamil transpeptidasa 1” (*GGT1*, AT4G39640.2) que está implicado en la mitigación del estrés oxidativo. Las “gamma-glutamil transferasa” (*GGT*) son enzimas encargadas de degradar el glutatión (*GSH*) en espacios extra-citosolicos. La degradación mitiga el estrés oxidativo al

metabolizar GSSG que es la forma oxidada del GSH. En *Arabidopsis* la isoforma GGT1, que es la localizada en este trabajo, se ha encontrado en el apoplasto unida a la pared celular (Giaretta *et al.*, 2017). Con esta información se puede observar la existencia de genes con altos niveles de expresión, relacionados con los mecanismos de protección en las semillas de *C. odorata*

2.6.3 Expresión diferencial de genes en la comparación de semillas secas vs. germinadas

Cuando se comparan las condiciones fisiológicamente más extremas analizadas en este trabajo el término GO detectado con mayor frecuencia fue el de respuesta celular a UV-C (GO:0071494), (Figura A1), dentro del cual se encontraban otros términos. Por ejemplo, aun se localizaba el término de “respuesta a auxina” (GO:0009733) que como se mencionó anteriormente, está relacionado con la expansión celular, así mismo, los términos de “desintoxicación de oxidantes celulares” (GO:0098869), “transducción de señales” (GO:0007165) y “responsable de defensas” (GO:0006952) Por lo que puede en realidad representar un grupo artificial que conjunta diversos elementos asociados al desarrollo y respuesta al estrés.

En el mismo grupo se localizó a la “ruta de señalización activada por etileno” (GO:0009873) y la respuesta celular al óxido nítrico (GO:0071732). Estas vías son muy importantes durante el proceso de germinación de las semillas. Al respecto durante esta transición se ve el aumento gradual del gen (*ESE3*, AT5G25190.1) que codifica para un miembro de la familia B-6 de los ERF (factor de respuesta al etileno) de la familia de factores de transcripción ERF/AP2 y son sensibles al etileno. Dentro de este grupo se encuentra a los ERFVII (factores de respuesta a etileno del grupo VII) los cuales se encargan de regular la respuesta ante el estrés biótico y abiótico y en particular al estrés por bajo nivel de oxígeno (Gibbs *et al.*, 2015). Este gen también está relacionado con la enzima “cisteína oxidasa vegetal” (*PCO*, AT5G39890.1), que se expresó en forma positiva durante la hidratación de la semillas (SvsE), dicha enzima está implicada en el control de la estabilidad de los ERF-VII y catalizan la conducción dependiente de oxígeno que da paso a la degradación proteosomal de dichos ERF-VII a través de la ruta N-terminal (Taylor-Kearney y Flashman, 2021). Los ERF-VII

modulan la señalización de AG y ABA para regular procesos de desarrollo y estrés (Gibbs *et al.*, 2015).

Adicionalmente, dentro de los procesos que regula el óxido nítrico (NO) se encuentra la latencia de las semillas y el crecimiento postgerminación. El aumento rápido del NO celular y la oxigenación de la semilla durante la imbibición estimulan la degradación de los ERFVII, esto hace que haya una baja expresión del gen “Insensible al ácido abscísico 5” (*ABI5*, regulador positivo de la latencia de las semillas) en el endospermo y promueva la germinación (Gibbs *et al.*, 2014). Por tanto, también se expresaron términos GO relacionados con la degradación como es el caso de la “proteólisis” (GO:0006508) y “ubiquitinación de proteínas” (GO:0016567), que se encuentran en el grupo de los términos relacionados a fotosíntesis (Figura A1). La ubiquitinación es importante porque media los cambios precisos y necesarios para el crecimiento, desarrollo y la adaptación a tensiones ambientales (Han *et al.*, 2022).

Entre los factores de transcripción detectados, desde la comparación de los extremos (SvsG) y durante la germinación de las semillas (EvsG) se observó el aumento en la transcripción del gen *HEC2* (AT3G50330.1) que codifica al “factor de transcripción Bhlh” (Schuster *et al.*, 2015) el cual pudiera estar posiblemente asociado a la emergencia de la radícula, del mismo modo que actúa en semillas de *G. mangostana* (Goh *et al.*, 2019). Además, también se detectó al gen que codifica para el “factor de transcripción de la familia BZIP” que están asociado a la respuesta del estrés por desecación (Giarola *et al.*, 2017), localizando específicamente a BZIP61 (AT3G58120.1). Tanto el factor de transcripción *bHLH* y *BZIP* también han sido localizados en el desarrollo de semillas y acumulación de aceite en *M. oleifera* (Yang *et al.*, 2018).

Por otro lado, y dentro de este grupo también se detectaron genes relacionados con la “respuesta a defensa” (GO:0006952) que también se encontraron desde la hidratación de las semillas. Durante el proceso de germinación de la semilla es posible que se liberen azúcares en el transcurso de la degradación del endospermo, estos azúcares funcionan como alimentos de ciertos hongos, y éste, podría ser el motivo por el que se activan genes para contrarrestar dicho efecto (Silva *et al.*, 2019), tal es el

caso de la proteína similar a “receptor 51” (*RLP51*, AT4G18760.1) y la “proteína similar a la taumatina 1” (*ATLP-1*, AT1G18250.1).

Durante esta transición de semilla seca hasta su germinación también se siguió expresando el término “organización de vacuolas líticas” (GO:0080171). Esta movilización de reservas es fundamental en la germinación de las semillas al carecer de fotosíntesis y un sistema de absorción de minerales (FU *et al.*, 2021). Por tanto, se presentó el aumento gradual de genes que codifican para la proteína de la superfamilia de “albúmina 2S de almacenamiento de semillas” (AT1G62510.1, AT4G22460.1). Así mismo, se presentó el aumento de genes asociados al metabolismo de arabinogalctanos (AGP), tal es el caso de los genes que codifican para la “proteína arabinogalactano 14” (*AGP14*, AT5G56540.1), “proteína arabinogalactano 41” (*AGP41*, AT5G24105.1) y dos “AGP tipo fasciclina” (*FLA*), “arabinogalactano similar a la fasciclina 6”, (*FLA6*, AT2G20520.1) y “arabinogalactano similar a la fasciclina 7” (*FLA7*, AT2G04780.2).

Se considera que los AGP también forman parte de las reservas en los cotiledones y se han encontrado con mucha frecuencia en las paredes celulares engrosadas de los cotiledones en semillas del género *Lupinus*, compuestos por residuos de galactosa unidos con cadenas laterales de arabinosa; estas hemicelulosas son degradadas a la vez que se movilizan las reservas almacenadas (proteínas y aceites) (Bewley *et al.*, 2013). En particular *AGP14* se ha localizado en *Arabidopsis* asociada a la expansión radial de las células en la punta de la raíz (Ogawa-Ohnishi y Matsubayashi, 2015) mientras que *AGP41* esta clasificada como una supuesta proteína y la información de su función es limitada. Los genes *FLA* contienen dominios tipo fasciclina y *AGP* típicos y dominios de adhesión celular putativos (Schultz *et al.*, 2004; Costa *et al.*, 2019). Los genes *FLA* ejercen funciones relevantes en el crecimiento y desarrollo de raíces tallos y semillas en *Arabidopsis* (Wang *et al.*, 2015), así como la integridad y elasticidad de la matriz de la pared celular. También están relacionados con aspectos fundamentales de la embriogénesis y desarrollo de semillas en álamo, maíz y alcornoque (Costa *et al.*, 2019). En el genoma de *Populus trichocarpa* Torr. & A. Gray se han identificado 35 genes *FLA* (Zang *et al.*, 2015). *FLA6* se presenta de forma

abundante en las células de la fibra del xilema (Wang *et al.*, 2015) mientras que *FLA7* se presenta tanto en las membranas como en los cloroplastos (Zang *et al.*, 2015).

Otros genes que siguieron su aumento gradual fueron “heme oxigenasa 1” (*HO1*, AT2G26670.1) y “xiloglucano endotransglucosilasa/hidrolasa 8” (*XTH8*, AT1G11545.1). *HO1* es una enzima inducida por la hematina y se ha encontrado promoviendo la germinación en arroz en condiciones de sequía (Wu *et al.*, 2016). El ortólogo de *XTH8* en semillas de *Lycopersicon esculentum* P. Mill también ha sido encontrada en el embrión durante la germinación previa emergencia de la radícula (Bewley *et al.*, 2013).

A manera de resumen la Figura 2.5 muestra un modelo de los genes y procesos fisiológicos más representativos en este estudio durante la germinación de semillas en *C. odorata*.

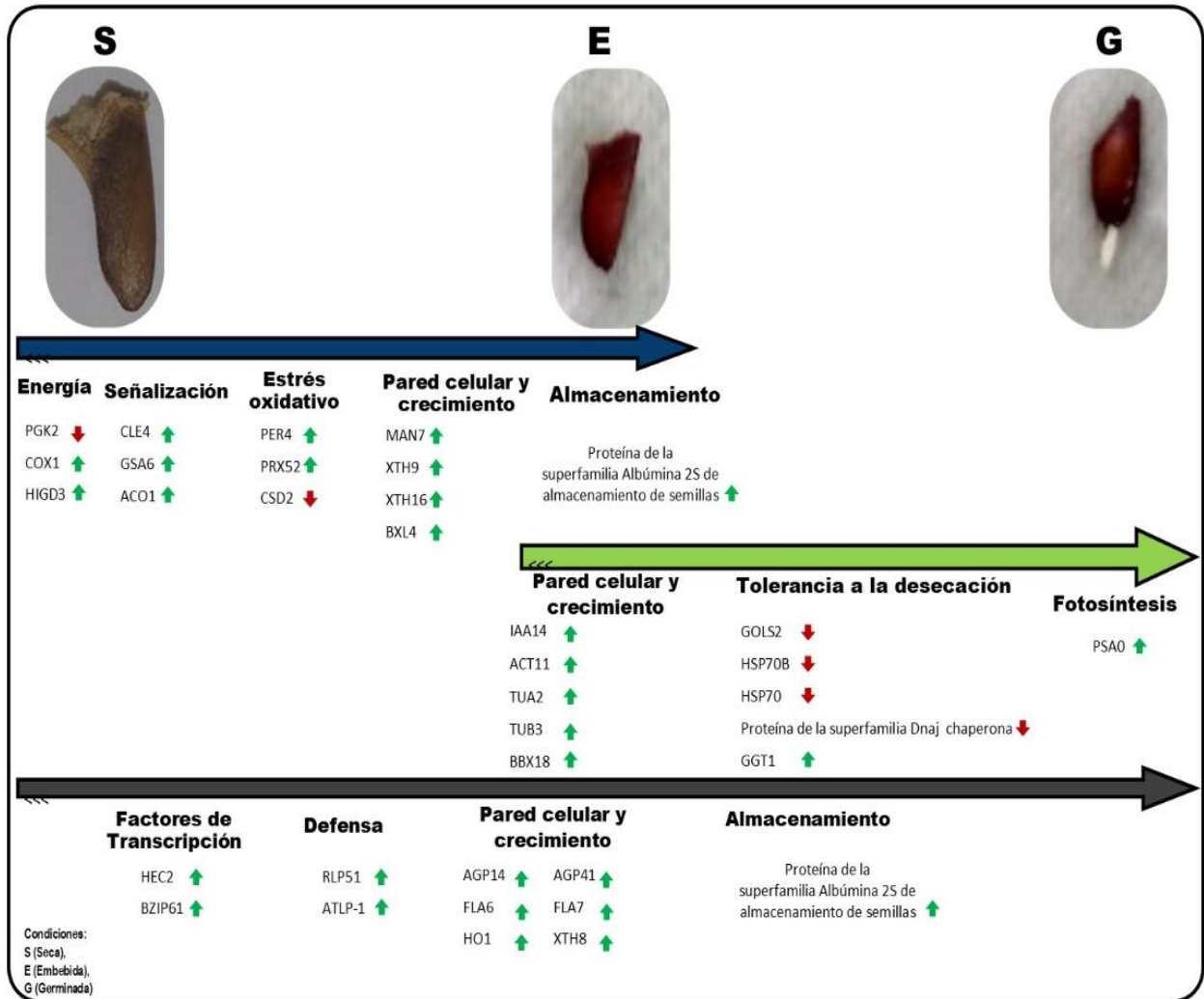


Figura 2.5. Genes diferencialmente expresados durante la germinación de semillas de *C. odorata*. La flecha azul con sus respectivos enunciados indica los procesos y genes que se expresaron durante el paso de la semilla seca a embebida (SvsE), la verde durante el paso de la semilla embebida a germinada (EvsG) y la negra aquellos que se expresaron durante todo el proceso (SvsG). Las flechas verdes y rojas indican que el gen aumentó o disminuyó su expresión respectivamente.

2.7 CONCLUSIONES

En el estudio de expresión diferencial se localizaron a 8,863 DEG de los cuales 243 corresponden a los genes con mayores cambios en sus niveles de transcripción en las tres condiciones evaluadas. Esto permitió evidenciar dos fases importantes en la germinación, la primera, que implica el paso de una semilla seca quiescente y su posterior hidratación, en el que se reinicia la activación metabólica a través de distintas cascadas de señalización que dan paso a la actividad de genes implicados en el metabolismo energético, respuesta al estrés oxidativo, organización de vacuolas líticas y recambio de la pared celular.

Durante la segunda etapa, que abarca el paso de la semilla embebida a germinada, se modifica la expresión de genes involucrados en el alargamiento del embrión, la ruptura de la testa y protrusión de la radícula, es decir a la germinación *sensu stricto*, para posteriormente dar lugar a una plántula con vida autótrofa. En particular, se evidenció cambios en la expresión de genes de la respuesta a auxina, organización del citoesqueleto, pared celular, además de otros procesos relacionados con la fotosíntesis.

A pesar de que las semillas de *C. odorata* no sufren una desecación alta como lo hacen las semillas ortodoxas, se pudo observar la presencia de distintos genes que están involucrados en la tolerancia a la desecación, tal es el caso de las HSP que, junto a otras proteínas, hacen que el citoplasma pase del estado líquido al estado vítreo, lo que disminuye las actividades metabólicas. Entre estas se localizaron a: *HSP70*, *HSP70B*, *HSP40* y *HSP20*, además de otros genes como *GGT1* y *GOLS2*.

CONCLUSIONES GENERALES

En este estudio se generó el primer transcriptoma de *novο* y anotación funcional para el proceso germinativo de semillas de *C. odorata*. Dicho transcriptoma se puede utilizar como un transcriptoma de referencia para posteriores investigaciones en la misma especie u otra relacionada o de interés particular.

Los datos obtenidos brindan información importante que permiten analizar la clasificación funcional de proteínas ortólogas en el proceso germinativo de *C. odorata*, lo que permitió su clasificación dentro de 22 grupos funcionales, destacando los procesos relacionados a las modificaciones postraduccionales, transcripción y traducción de señales.

Del transcriptoma total se expresaron diferencialmente 8,863 genes. La selección de genes con mayor expresión en cada comparación permitió evidenciar dos fases importantes del proceso germinativo, en el primero se activan genes relacionados a los procesos energéticos, señalización, pared celular, y movimiento de proteínas de reserva, además de genes relacionados con la respuesta al estrés oxidativo generado por las ROS que podrían estar participando como moléculas de señalización para dar paso a la germinación o bien como oxidantes de sistemas membranales y macromoléculas. En la segunda se activan procesos relacionados con el crecimiento del embrión entre los que se encuentran: respuesta a auxina, organización del citoesqueleto y pared celular, además de la presencia de genes relacionados con la fotosíntesis.

Como parte de los DEG que más aumentaron o disminuyeron sus niveles de expresión en todo el transcriptoma, se localizaron 11 secuencias que no han sido caracterizadas en ninguna de las bases de datos utilizadas, por lo que se necesita profundizar en el estudio de dichas secuencias y determinar si podría tratarse de nuevos genes expresados en la germinación de las semillas de *C. odorata*.

La información generada en esta investigación es de gran importancia para futuros estudios enfocados en los planes de manejo y conservación *ex-situ* de semillas de *C. odorata* en los bancos de germoplasma, pues permite conocer genes y

mecanismos moleculares implicados en el proceso germinativo. Esta información podrá ser utilizada para seleccionar genes de interés y hacer estudios más profundos enfocados en el análisis de la longevidad de las semillas en el almacenamiento. Además de ello, con los datos obtenidos, se da pauta a investigaciones más profundas de interés particular en la especie.

LITERATURA CITADA

- Aguilar-Bultet, L. y L. Falquet. 2015. Secuenciación y ensamblaje de *novu* de genomas bacterianos: una alternativa para el estudio de nuevos patógenos. *Revista de Salud Animal* 37: 125–132.
- Alía, R., D. Angúndez, N. Alba, S.C. González-Martínez, y A. Soto. 2003. Variabilidad genética y gestión forestal. *Ecosistemas* 12(3): 1–7.
- Alonso-Sánchez, A., E. Villafan, C.A. Pérez-Torres y E. Ibarra-Laclette. 2017. Transcriptómica: aplicaciones en biotecnología vegetal. *In: Fundamentos de Biotecnología Vegetal*. Rodríguez, A., G. Acevedo y O. Castellanos (comps.). Guadalajara, Jalisco. Ediciones. pp: 295–320.
- Argout, X., J. Salse, J.M. Aury, M.J. Guiltinan, G. Droc, J. Gouzy, M. Allegre, C. Chaparro, T. Legavre, S.N. Maximova, M. Abrouk, F. Murat, O. Fouet, J. Poulain, M. Ruiz, Y. Roguet, M. Rodier-Goud, J.F. Barbosa-Neto, F. Sabot, D. Kudrna, J.S.S. Ammiraju, S. C. Schuster, J. E. Carlson, E. Sallet, T. Schiex, A. Dievart, M. Kramer, L. Gelley, Z. Shi, A. Bérard, C. Viot, M. Boccara, A.M. Risterucci, V. Guignon, X. Sabau, M.J. Axtell, Z. Ma, Y. Zhang, S. Brown, M. Bourge, W. Golser, X. Song, D. Clemen, R. Rivallan, M. Tahí, J.M. Akaza, B. Pitollat, K. Gramacho, A. D'Hont, D. Brunel, D. Infante, I. Kebe, P. Costet, R. Wing, W.R. McCombie, E. Guiderdoni, F. Quetier, O. Panaud, P. Wincker, S. Bocs and C. Lanaud. 2011. The genome of *Theobroma cacao*. *Nature Genetics* 43: 101–108.
- Babraham Institute. 2020. FastQC. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> Accessed: 10 de Septiembre 2022.
- Bailly, C. 2019. The signalling role of ROS in the regulation of seed germination and dormancy. *Biochemical Journal* 476: 3019–3032.
- Bai, B., S. Van Der Horst, J.H.G. Cordewener, T.A.H.P. America, J. Hanson and L. Bentsinka. 2020. Seed-stored mRNAs that are specifically associated to monosomes are translationally regulated during germination. *Plant Physiology* 182: 378–392.
- Bai, B., A. Peviani, S. van der Horst, M. Gamm, B. Snel, L. Bentsink and J. Hanson. 2017. Extensive translational regulation during seed germination revealed by polysomal profiling. *New Phytologist* 214: 233–244.
- Barone, J.D. y C. Luna. 2016. Determinación de la eficacia de métodos de evaluación de calidad de semillas de especies forestales nativas de la Selva Atlántica. Quebracho - *Revista de Ciencias Forestales* 24: 70–80.
- Bewley, J.D., K.J. Bradford, H.W.M. Hilhorst, and H. Nonogaki. 2013. *Seeds Physiology of development, Germination and dormancy*. 3ra ed. Springer. New York. 392 p.
- Blázquez, M.A., D.C. Nelson and D. Weijers. 2020. Evolution of Plant Hormone

- Response Pathways. Annual Review of Plant Biology 71: 327–353.
- Caicedo-Montoya, C., L. Pinilla, L.F. Toro, J. Yepes-García and R. Ríos-Esteva. 2019. Comparative analysis of strategies for de novo transcriptome assembly in prokaryotes: *Streptomyces clavuligerus* as a case study. High-Throughput 8: 1-18.
- Campos M., C., L. Boatwright., C.C. Da Silva., E.J. Scaloppi J., P. de Souza G., W. Brad B. and A. Pereira de S. 2019. Deep expression analysis reveals distinct cold-response strategies in rubber tree (*Hevea brasiliensis*). BMC Genomics 20: 1–20.
- Cao, X., M. Li, J. Li, Y. Song, X. Zhang, D. Yang, M. Li and J. Wei. 2021. Co-expression of hydrolase genes improves seed germination of *Sinopodophyllum hexandrum*. Industrial Crops and Products 164: 113414.
- Carrillo, I., J.P. Elissetche, S. Valenzuela y R. Teixeira. 2013. Formación de elementos anatómicos en maderas duras: una revisión desde una perspectiva genómica. Maderas. Ciencia y Tecnología 15: 93–104.
- Castro, J.C., J.D. Maddox, H.N. Rodríguez, C.G. Castro, S.A. Imán-Correa, M. Cobos, J.D. Paredes, J.L. Marapara, J. Braga and P.M. Adrianzén. 2020. Dataset of de novo assembly and functional annotation of the transcriptome during germination and initial growth of seedlings of *Myrciaria Dubia* “camu-camu.” Data in Brief 31: 105834.
- Cavers, S., C. Navarro and A.J. Lowe. 2003a. A combination of molecular markers identifies evolutionarily significant units in *Cedrela odorata* L. (Meliaceae) in Costa Rica. Conservation Genetics 4: 571–580.
- Cavers, S., C. Navarro and A.J. Lowe. 2003b. Chloroplast DNA phylogeography reveals colonization history of a neotropical tree, *Cedrela odorata* L., in Mesoamerica. Molecular Ecology 12: 1451–1460.
- Cavers, S., A. Telford, F. Arenal C., A.J. Pérez C., R. Valencia, C. Navarro, A. Buonamici, A.J. Lowe and G.G. Vendramin. 2013. Cryptic species and phylogeographical structure in the tree *Cedrela odorata* L. throughout the Neotropics. Journal of Biogeography 40: 732–746.
- Chae, K., C.G. Isaacs, P.H. Reeves, G.S. Maloney, G.K. Muday, P. Nagpal and J.W. Reed. 2012. *Arabidopsis* *SMALL AUXIN UP RNA63* promotes hypocotyl and stamen filament elongation. The Plant Journal 71: 684–697.
- Chen, Z., X. Qi, X. Yu, Y. Zheng, Z. Liu, H. Fang, L. Li, Y. Bai, C. Liang, W. Li and B.M. Lange. 2021. Genome-wide analysis of terpene synthase gene family in *Mentha longifolia* and catalytic activity analysis of a single terpene synthase. Genes 12: 1–18.
- Cintrón, B.B. 1990. *Cedrela odorata* L. In: Silvics of North America. Burns, R.M. and

- B.H.Honkala. (eds). Washington, DC. Department of Agriculture Forest Service. pp: 128–134.
- CITES (Convención sobre el Comercio Internacional de Especies Amenazadas de Flora y Fauna Silvestres). 2017. Apéndices I, II y III. Châtelaine, Ginebra, Suiza. 76 p.
- CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad). 1983. *Cedrela odorata* L. In: Sistema Nature. Edition 10:940. Disponible en línea en: http://www.conabio.gob.mx/conocimiento/info_especies/arboles/doctos/36-melia2m.pdf Acceso: 08 de Septiembre 2022.
- Cortés, A. 2001. El cedro y la caoba en Yucatán, México. *Revista Forestal Centroamericana* 36: 55–57.
- Cortese, I.J., M.L. Castrillo, P.D. Zapata y M.E. Laczeski. 2021. Efecto del filtrado de secuencias en el ensamblado del genoma de *Bacillus altitudinis* 19RS3 aislado de *Ilex paraguariensis*. *Acta Biologica Colombiana* 26: 170–177.
- Costa, M., A.M. Pereira, S.C. Pinto, J. Silva, L.G. Pereira and S. Coimbra. 2019. In silico and expression analyses of fasciclin-like arabinogalactan proteins reveal functional conservation during embryo and seed development. *Plant Reproduction* 32: 353–370.
- Damaris, R.N., Z. Lin, P. Yang and D. He. 2019. The rice alpha-amylase, conserved regulator of seed maturation and germination. *International Journal of Molecular Sciences* 20: 1–17.
- Del Fabro, C., S. Scalabrin, M. Morgante and F. Giorgi. 2013. An extensive evaluation of read trimming effects on illumina NGS data analysis. *PLoS ONE* 8: 1-13.
- de la Torre, A. 2013. Estructura poblacional, diversidad y flujo genético en poblaciones ribereñas de cedro (*Cedrela odorata* L.). In: Reporte Manu 2013: Pasión por la investigación en la Amazonia Peruana. Groenendijk, J. y A.Tovar (comps.). Lima, Peru. SERNAP. pp: 89–97.
- de la Torre, A., C. López, E. Yglesias, and J.P. Cornelius. 2008. Genetic (AFLP) diversity of nine *Cedrela odorata* populations in Madre de Dios, southern Peruvian Amazon. *Forest Ecology and Management* 255: 334–339.
- Deorowicz, S. and S. Grabowski. 2011. Compression of DNA sequence reads in FASTQ format. *Bioinformatics* 27: 860–862.
- Dickmann, D.I., M.A. Gold and J.A. Flore. 1994. The ideotype concept and the genetic improvement of the tree crops. In: *Plant Breeding Review*. J. Janick (ed.). USA. John Wiley & Sons Inc. pp: 521–534.
- Doria, J. 2010. Generalidades sobre las semillas: su producción, conservación y almacenamiento. *Cultivos tropicales* 31: 74–85.

- Dussert, S., J. Serret, A. Bastos-Siqueira, F. Morcillo, E. Déchamp, V. Rofidal, P. Lashermes, H. Etiene and T. Joët. 2018. Integrative analysis of the late maturation programme and desiccation tolerance mechanisms in intermediate coffee seeds. *Journal of Experimental Botany* 69: 1583–1597.
- El-Maarouf-bouteau, H. 2022. The Seed and the Metabolism Regulation. *Biology* 11: 168.
- Fan, F.F., X. Yang, Y. Cheng, Y. Kang and X. Chai. 2017. The DnaJ gene family in pepper (*Capsicum annuum* L.): Comprehensive identification, characterization and expression profiles. *Frontiers in Plant Science* 8: 1–11.
- Feraru, E., T. Paciorek, M.I. Feraru, M. Zwiewka, R. de Groot, R. De Rycke, J. Kleine-Vehn and J. Friml. 2010. The AP-3 β adaptin mediates the biogenesis and function of lytic vacuoles in *Arabidopsis*. *Plant Cell* 22: 2812–2824.
- Finch, K.N., F.A. Jones and R.C. Cronn. 2019. Genomic resources for the Neotropical tree genus *Cedrela* (Meliaceae) and its relatives. *BMC Genomics* 20: 1–17.
- Foley, M.E., W.S. Chao, D.P. Horvath, M. Dođramaci and J. V. Anderson. 2013. The transcriptomes of dormant leafy spurge seeds under alternating temperature are differentially affected by a germination-enhancing pretreatment. *Journal of Plant Physiology* 170: 539–547.
- Foxx, J., S.W. Tighe, C.M. Nicolet, J.M. Zook, M. Byrska-Bishop, W.E. Clarke, M.M. Khayat, M. Mahmoud, P.K. Laaguiby, Z.T. Herbert, D. Warner, G.S. Grills, J. Jen, S. Levy, J. Xiang, A. Alonso, X. Zhao, W. Zhang, F. Teng, Y. Zhao, H. Lu, G.P. Schroth, G. Narzis, W. Farmerie, F.J. Sedlazeck, B.A. Baldwin and C.E. Mason. 2021. Performance assessment of DNA sequencing platforms in the ABRF Next-Generation Sequencing Study. *Nature Biotechnology* 39: 1129–1140.
- FU, F., P. Ying-shu, W. Gui-bin, Y.A. EL-KASSABY and C. Fu-liang 2021. Integrative analysis of the metabolome and transcriptome reveals seed germination mechanism in *Punica granatum* L. *Journal of Integrative Agriculture* 20: 132–146.
- Fuchs, J., A. Brandes and I. Schubert. 1995. Telomere sequence localization and karyotype evolution in higher plants. *Plant Systematics and Evolution* 196: 227–241.
- Galván-Hernández, D.M., M.A. Macedo-Villareal, F.F. Núñez de Cáceres-González, A. Sánchez-González and P. Octavio-Aguilar. 2018. Morphological variation of *Cedrela odorata* (Meliaceae): contrast between natural and managed populations. *Acta Botanica Mexicana* 125: 157–171.
- García, J.M. 2018. Desarrollo y optimización de un pipeline para el análisis dual del transcriptoma planta/patógeno : identificando los mecanismos de patogenicidad y defensa. Catalunya, España. Tesis de Maestría. Universitat Oberta de Catalunya. Disponible en línea en: <https://1library.co/document/wq2jd96y-desarrollo-optimizacion-analisis-transcriptoma-patogeno-identificando-mecanismos->

- patogenicidad.html Acceso: 10 de septiembre de 2022.
- García, T., y A. Abdelnour. 2013. Crioconservación de ápices y semillas de cedro (*Cedrela odorata* L.) mediante las técnicas de vitrificación y deshidratación. *Agronomía Costarricense* 37: 113–126.
- García C., F., V. Sánchez M., V. Jacob C., A. Rosete R., A. Valdes B., E. Hernández F. y A. Martínez R. 2015. Herramientas para el manejo, análisis y almacenamiento de semillas de cedro rojo (*Cedrela odorata* L.) Folleto Técnico Num.17. Cenic-Comef, INIFAP. México, D.F. 80p.
- GBIF (Global Biodiversity Information Facility). 2018. Clasificación de especies. Disponible en línea en: <https://www.gbif.org/species/3190511> Acceso: 17 de enero de 2019.
- Giaretta, S., D. Prasad, I. Forieri, T. Vamerali, A.R. Trentin, M. Wirtz, R. Hell and A. Masi. 2017. Apoplastic gamma-glutamyl transferase activity encoded by GGT1 and GGT2 is important for vegetative and generative development. *Plant Physiology and Biochemistry* 115: 44–56.
- Giarola, V., Q. Hou and D. Bartels. 2017. Angiosperm plant desiccation tolerance: hints from transcriptomics and genome sequencing. *Trends in Plant Science* 22: 705–717.
- Gibbs, D.J., N. Mdlsa, M. Movahedi, J. Lozano-Juste, G.M. Mendiondo, S. Berckhan, N. Marián-de la Rosa, J.C. Conde, C.S. Correia, S.P. Pearce, G.W. Bassel, B. Hamali, P. Talloji, D.F.A. Tome, A. Coego, J. Beynon, D. Alabadí, A. Bachmair, J. León, J.E. Gray, F.L. Theodoulou and M.J. Holdsworth. 2014. Nitric oxide sensing in plants is mediated by proteolytic control of Group VII ERF Transcription Factors. *Molecular Cell* 53: 369–379.
- Gibbs, D.J., J.V. Conde, S. Berckhan, G. Prasad, G.M. Mendiondo and M.J. Holdsworth. 2015. Group VII ethylene response factors coordinate oxygen and nitric oxide signal transduction and stress responses in plants. *Plant Physiology* 169: 23–31.
- Gillies, A.C.M., J.P. Cornelius, A.C. Newton, C. Navarro, M. Hernández and J. Wilson. 1997. Genetic variation in Costa Rican populations of the tropical timber species *Cedrela odorata* L., assessed using RAPDs. *Molecular Ecology* 6: 1133–1145.
- Goh, Hoe-Han., S. Abu B., N.D. Kamal A., Z. Zainal and N. Mohd N. 2019. Transcriptional reprogramming during *Garcinia*-type recalcitrant seed germination of *Garcinia mangostana*. *Scientia Horticulturae* 257: 108727.
- Gómez-Maqueo, X., D. Soriano, D. N. Velázquez-Rosas, S. Alvarado-López, K. Jiménez-Durán, M. D. M. Garciadiego and A. Gamboa-deBuen. 2020. The seed water content as a time-independent physiological trait during germination in wild tree species such as *Ceiba aesculifolia*. *Scientific reports* 10:1-14.

- Goujon, T., Z. Minic, A. El Amrani, O. Lerouxel, E. Aletti, C. Lapierre, J. Joseleau and L. Jouanin. 2003. *AtBXL1*, a novel higher plant (*Arabidopsis thaliana*) putative beta-xylosidase gene, is involved in secondary cell wall metabolism and plant development. *Plant Journal* 33: 677–690.
- Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F.D. Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.
- Grattapaglia, D. 2017. Status and perspectives of genomic selection in forest tree breeding. *In: Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement*. Varshney, R.K., M. Roorkiwal, and M.E. Sorrells (eds.). Brasilia, Brasil. Springer, Cham. pp: 199-249.
- Han, D., Z. Yu, J. Lai and C. Yang. 2022. Post-translational modification: a strategic response to high temperature in plants. *aBIOTECH* 3: 49–64.
- Hanukoglu, I. 2015. Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochemistry and Molecular Biology Education* 43: 206–209.
- Hay, F.R. and R.J. Probert. 2013. Advances in seed conservation of wild plant species: a review of recent research. *Conservation Physiology* 1: 1–11.
- Hernández-Máximo, E., J. López-Upton, V. Sánchez-Monsalvo, J.J. Vargas-Hernández and J.G. Salazar-García. 2016. Early performance and genetic gain of *Cedrela odorata* families from wide-ranging sites in Mexico. *Journal of Tropical Forest Science* 28: 446–456.
- Hernández, M., N.M. Quijada, D. Rodríguez-Lázaro y J.M. Eiros. 2020. Aplicación de la secuenciación masiva y la bioinformática al diagnóstico microbiológico clínico. *Revista Argentina de Microbiología* 52: 150–161.
- Hölzer, M. and M. Marz. 2019. De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *GigaScience* 8: 1–16.
- Hou, M., Y. Zhang, G. Mu, S. Cui, X. Yang and L. Liu. 2020. Molecular cloning and expression characterization of flavonol synthase genes in peanut (*Arachis hypogaea*). *Scientific Reports* 10: 1–11.
- Hwang, S.T., H. Li, H. Alavilli, B. Lee and D. Choi. 2017. Molecular and physiological characterization of *AtHIGD1* in *Arabidopsis*. *Biochemical and Biophysical Research Communications* 487: 881–886.
- ISTA (International Seed Testing Association). 2016. Reglas internacionales para el análisis de las semillas. 187 p.
- ITTO (International Tropical Timber Organization). 2019. Tropical timber market report.

- Japón. 24p. Disponible en línea en: <http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=2009675429&site=ehost-live> Acceso: 15 de enero de 2019.
- IUCN (International Union for Conservation of Nature). 2017. *Cedrela odorata*, Spanish Cedar. Available online at: <https://www.iucnredlist.org/species/32292/68080590> Accessed: 07 de Septiembre 2022.
- Jaillon, O., J.M. Aury, B. Noel, A. Policriti, C. Clepet, A. Cassagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467.
- Jamsheer, M. and A. Laxmi. 2014. DUF581 is plant specific FCS-like zinc finger involved in protein-protein interaction. *PLoS ONE* 9: e99074.
- Jiang, Y., M. Tian, C. Wang and Y. Zhang. 2021. Transcriptome sequencing and differential gene expression analysis reveal the mechanisms involved in seed germination and protocorm development of *Calanthe tsoongiana*. *Gene* 772: 145355.
- Jiménez, P. y C. Collada. 2000. Técnicas para la evaluación de la diversidad genética y su uso en los programas de conservación. *Investigación Agraria. Sistemas y Recursos Forestales* 2: 237–248.
- Jump, A.S., R. Marchant and J. Peñuelas. 2008. Environmental change and the option value of genetic diversity. *Trends in Plant Science* 14: 51–58.
- Khosla, P.K. and B.T. Styles. 1975. Karyological studies and chromosomal evolution in Meliaceae. *Silvae Genetica* 24: 73–83.
- Koecke, A.V., A.N. Muellner-Riehl, T.D. Pennington, G. Schorr and J. Schnitzler. 2013. Niche evolution through time and across continents: The story of Neotropical *Cedrela* (Meliaceae). *American Journal of Botany* 100: 1800–1810.
- Krishnan, N.M., S. Pattnaik, P. Jain, P. Gaur, R. Choudhary, S. Vaidyanathan, S. Deepak, A. K. Hariharan, P.G. Bharath K, J. Nair, L.Varghese, N.K Valivarthi, K.Dass, K. Ramaswamy and B. Panda. 2012. A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. *BMC Genomics* 13: 464.
- Kumar, V. 2021. Ribosomal biogenesis in eukaryotes. *In: Emerging concepts in ribosome structure, biogenesis, and function*. Kumar, V. (ed.). Nueva Delhi, India. Academic Pres. pp. 129–150.
- Kurek, K., B. Plitta-Michalak and E. Ratajczak. 2019. Reactive oxygen species as potential drivers of the seed aging process. *Plants* 8: 1–13.
- Lariguet, P., P. Ranocha, M. de Meyer, O. Barbier, C. Penel and C. Dunand. 2013. Identification of a hydrogen peroxide signalling pathway in the control of light-

- dependent germination in *Arabidopsis*. *Planta* 238: 381–395.
- Lenz, P.R.N., S. Nadeau, M.J. Mottet, M. Perron, N. Isabel, J. Beaulieu and J. Bousquet. 2020. Multi-trait genomic selection for weevil resistance, growth, and wood quality in Norway spruce. *Evolutionary Applications* 13: 76–94.
- Leshner-Gordillo, J.M., J.L. Martínez-Sánchez, P. Orduña-Rodas y A.I. Hernández-Carreta. 2018. Diversidad genética del cedro rojo (*Cedrela odorata*) en el estado de Tabasco, México. *Bosque* 39: 411–417.
- Li, X., R. Fei, Z. Chen, C. Fan and X. Sun. 2020. Plant hormonal changes and differential expression profiling reveal seed dormancy removal process in double dormant plant-herbaceous peony. *PLoS ONE* 15: 1–19.
- Lisboa, C.G.S., P.P. Tonini, M.A.S. Tiné and M.S. Buckeridge. 2006. Endo- β -mannanase from the endosperm of seeds of *Sesbania virgata* (Cav.) Pers. (Leguminosae): purification, characterisation and its dual role in germination and early seedling growth. *Brazilian Journal of Plant Physiology* 18: 269–280.
- Liu, D., W. Li, J. Cheng and L. Hou. 2015. *AtPGK2*, a member of PGKs gene family in *Arabidopsis*, has a positive role in salt stress tolerance. *Plant Cell, Tissue and Organ Culture* 120: 251–262.
- Ljung, K., A. Östin, L. Lioussanne and S. Göran. 2001. Developmental regulation of indole-3-acetic acid turnover in Scots pine seedlings. *Plant Physiology* 125: 464–475.
- López-Ochoa, L.A., M.M. Apolinar-Hernández and Y.J. Peña-Ramírez. 2015. Characterization of chloroplast region *rrn16-rrn23S* from the tropical timber tree *Cedrela odorata* L. And de novo construction of a transplastomic expression vector suitable for meliaceae trees and other economically important crops. *Genetics and Molecular Research* 14: 1469–1478.
- Love, M.I., W. Huber and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 1–21.
- Macgregor, D.R., S.L. Kendall, H. Florance, F. Fedi, K. Moore, K. Paszkiewicz, N. Smirnov and S. Penfiel. 2015. Seed production temperature regulation of primary dormancy occurs through control of seed coat phenylpropanoid metabolism. *New Phytologist* 205: 642–652.
- Mader, M., B. Pakull, C. Blanc-Jolivet, M. Paulini-Drewes, Z. Bouda, B. Degen, I. Small and B. Kersten. 2018. Complete Chloroplast Genome Sequences of Four Meliaceae Species and Comparative Analyses. *International Journal of Molecular Sciences* 19: 701.
- Magnitskiy, S. V. y G.A. Plaza. 2007. Fisiología de semillas recalcitrantes de árboles tropicales. *Agrochimica Colombiana* 25: 96–103.
- Maret, W. 2017. Zinc in cellular regulation: The nature and significance of “zinc

- signals". International journal of molecular sciences 18: 2285.
- Marquetti, J.R. 1990. Híbridos de *Cedrela* resistentes a *Hypsipyla grandella*. Revista Forestal Baracoa 20: 97–101.
- Massange-Sánchez, J.A., L.E. Casados-Vázquez, S. Juárez-Colunga, R.J.H. Sawers and A. Tiessen. 2020. The phosphoglycerate kinase (PGK) gene family of maize (*Zea mays* var. B73). Plants 9: 1–20.
- Matilla, A.J. 2008. Desarrollo y germinación de las semillas. In: Fundamentos de fisiología vegetal. Azcon-Bieto., J y M.Talon (comps.). Madrid España. McGraw-Hill Interamericana. Pp: 537–558.
- Moles, T.M., L. Guglielminetti and T. H. Reyes. 2019. Differential effects of sodium chloride on germination and post-germination stages of two tomato genotypes. Scientia Horticulturae 257: 1–6.
- Muellner, A.N., T.D. Pennington, A. Valerie, and S.S. Renner. 2010. Biogeography of *Cedrela* (Meliaceae, Sapindales) in Central and South America. American Journal of Botany 97: 511–518.
- Murillo-Brito, Y., M. Domínguez-Domínguez, P. Martínez-Zurimendi, L. del C. Lagunes-Espinoza y A. Aldrete. 2017. Índice de sitio en plantaciones de *Cedrela odorata* en el trópico húmedo de México. Revista de la Facultad de Ciencias Agrarias 49: 15–30.
- Navarro, C., S. Cavers, A. Pappinen, P. Tigerstedt, A. Lowe and J. Merilla. 2005. Contrasting quantitative traits and neutral genetic markers for genetic resource assessment of Mesoamerican *Cedrela odorata*. Silvae Genetica 54: 281–292.
- Navarro, C., F. Montagnini and G. Hernández. 2004. Genetic variability of *Cedrela odorata* Linnaeus: Results of early performance of provenances and families from Mesoamerica grown in association with coffee. Forest Ecology and Management 192: 217–227.
- NCBI (National Center for Biotechnology Information). 2013. Available online at: https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome Accessed: 11 de septiembre de 2022.
- Nehra, N.S., M.R. Becwar, W.H. Rottmann, L. Pearson, K. Chowdhury, S. Chang, H.D. Wilde, R.J. Kodzyc, C. Zhang, K.C. Gause, D.W. Parks and M.A. Hinchee. 2005. Forest biotechnology: Innovative methods, emerging opportunities. In Vitro Cellular and Developmental Biology-Plant 41: 701–717.
- Nkongolo, K.K. 1996. Chromosome analysis and DNA homology in three *Picea* species, *P. mariana*, *P. rubens*, and *P. glauca* (Pinaceae). Plant Systematics and Evolution 203: 27–40.
- Obroucheva, N., I. Sinkevich, and S. Lityagina. 2016. Physiological aspects of seed recalcitrance: a case study on the tree *Aesculus hippocastanum*. Tree Physiology

36: 1127–1150.

- Ogawa-Ohnishi, M. and Y. Matsubayashi. 2015. Identification of three potent hydroxyproline O-galactosyltransferases in *Arabidopsis*. *Plant Journal* 81: 736–746.
- Oliveros, J.C. 2015. Una herramienta interactiva para comparar listas con diagramas de Venn. Disponible en línea en: <https://bioinfogp.cnb.csic.es/tools/venny/index.html> Acceso: 11 de septiembre de 2022.
- Orantes-García, C., M.A. Pérez-Farrera, T.M. Rioja-Paradela, y E.R. Garrido-Ramírez. 2013. Viabilidad y germinación de semillas de tres especies arbóreas nativas de la selva tropical, Chiapas, México. *Polibotánica* 36: 117–127.
- Park, Y.S., J. Beaulieu, and J. Bousquet. 2016. Multi-varietal forestry integrating genomic selection and somatic embryogenesis. *Vegetative Propagation of Forest Trees*: 302–322.
- Park, C.H., J. Roh, J.H. Youn, S.H. Son, J.H. Park, S.Y. Kim, T. Kim and S. Kim. 2018. *Arabidopsis* ACC oxidase 1 coordinated by multiple signals mediates ethylene biosynthesis and is involved in root development. *Molecules and Cells* 41: 923–932.
- Patiño, F. 1997. Diversidad genética en *Swietenia* y *Cedrela*. In: Recursos genéticos de *Swietenia* y *Cedrela* en los neotropicos: Propuestas para acciones coordinadas. FAO (comp), Roma, Italia. Disponible en línea en: <https://www.fao.org/3/ad111s/ad111s00.htm> Acceso: 08 de Septiembre 2022.
- Patro, R., G. Duggal, M.I. Love, R.A. Irizarry and C. Kingsford. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 14: 417–419.
- Pavy, N., B. Pelgas, S. Beauseigle, S. Blais, F. Gagnon, I. Gosselin, M. Lamothe, N. Isabel and J. Bousquet. 2008. Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: Application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics* 9: 1–17.
- Pennington, T.D. y K.J. Sarukhán. 2005. Árboles tropicales de México: manual para la identificación de las principales especies. UNAM-Fondo Cultura Económica. México. 523 p.
- Qin, N., Y. Gao, X. Cheng, Y. Yang, J. Wu, S. Li and G. Xing. 2021. Genome-wide identification of CLE gene family and their potential roles in bolting and fruit bearing in cucumber (*Cucumis sativus* L.). *BMC Plant Biology* 21: 1–18.
- Qu, C., S. Zhang, H. Zhao, J. Chen, Z. Zuo, X. Sun, Y. Cheng, Z. Xu and G. Liu. 2020. Analysis of the energy source at the early stage of poplar seed germination: verification of Perl's pathway. *3 Biotech* 10: 1–9.

- Rajjou, L., M. Duval, K. Gallardo, J. Catusse, J. Bally, C. Job and D. Job. 2012. Seed germination and vigor. *Annual review of plant biology* 63: 507.
- Renard, J., I. Martínez-Almonacid, A. Sonntag, I. Molina, J. Moya-Cuevas, G. Bissoli, J. Muñoz-Bertomeu, I. Faus, R. Niñoles, J. Shigeto, Y. Tsutsumi, J. Gadea, R. Serrano and E. Bueso. 2019. PRX2 and PRX25, peroxidases regulated by COG1, are involved in seed longevity in *Arabidopsis*. *Plant Cell and Environment* 43: 1–12.
- Rodríguez, L. 2011. La biotecnología aplicada al mejoramiento genético del cedro (*Cedrela odorata* L.). *Cuba Tabaco* 12: 63–70.
- Rodríguez C., A.E., L. Perlaza J. y A.J. Bernal G. 2014. Analizando datos de RNA-Seq en procariontes: una revisión para no expertos. *Acta Biológica Colombiana* 19: 131–142.
- Romo-Lozano, J.L., J.J. Vargas-Hernández, J. López-Upton, y M.L. Ávila-Angulo. 2017. Estimación del valor financiero de las existencias maderables de cedro rojo (*Cedrela odorata* L.) en México. *Madera Bosques* 23: 111–120.
- Rosa-Téllez, S., A.D. Anoman, M. Flores-Tornero, W. Toujani, S. Alseek, A.R. Fernie, S.G. Nebauer, J. Muñoz-Bertomeu, J. Segura and R. Ros. 2018. Phosphoglycerate kinases are co-regulated to adjust metabolism and to optimize growth. *Plant Physiology* 176: 1182–1198.
- Rottmann, T., F. Klebl, S. Schneider, D. Kischka, D. Rüscher, N. Sauer and R. Stadler. 2018. Sugar transporter STP7 specificity for L-arabinose and D-xylose contrasts with the typical hexose transporters STP8 and STP12. *Plant Physiology* 176: 2330–2350.
- Royal Botanic Gardens Kew. (2022) Seed Information Database (SID). Version 7.1. Available at : <http://data.kew.org/sid/> Accessed: 08 de Septiembre de 2022.
- Rubio, S., R.A. Pacheco-Orozco, A.M. Gómez, S. Perdomo and R. García-Robles. 2020. Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica. *Universitas Médica* 61: 49-63.
- Ruiz-Jiménez, C.A., H.M. De los Santos-Posadas, J.F. Parraguirre-Lezama y F.D. Saavedra-Millán. 2018. Evaluación de la categoría de riesgo de extinción del cedro rojo (*Cedrela odorata*) en México. *Revista Mexicana de Biodiversidad* 89: 938–949.
- Sampayo-Maldonado, S., J. López-Upton, V. Sánchez-Monsalvo and M. Jiménez-Casas. 2019. Genetic parameters of growth, and resistance to the shoot borer, in young clones of the tree *Cedrela odorata* (Meliaceae). *Revista de Biología Tropical* 67: 554–561.
- Sánchez-Coronado, M.E., R. Coates, L. Castro-Colina, A.G. de Buen, J. Paez-Valencia, B.L. Barradas, P. Huante and A. Orozco-Segovia. 2007. Improving seed

- germination and seedling growth of *Omphalea oleifera* (Euphorbiaceae) for restoration projects in tropical rain forests. *Forest Ecology and Management* 243: 144–155.
- Sánchez M., V., J.G. Salazar G., J.J. Vargas H., J. López U. y J. Jasso M. 2003. Parámetros genéticos y respuesta a la selección en características del crecimiento de *Cedrela odorata* L. *Revista Fitotecnia Mexicana* 26: 19–27.
- Sánchez S., S del C. 2015. Análisis de datos de RNA-Seq: comparación de métodos para el estudio de expresión génica diferencial. Sevilla, España. Trabajo de fin de grado. Universidad de Sevilla. Disponible en línea en: <https://idus.us.es/xmlui/handle/11441/40809> Acceso: 10 de septiembre de 2022.
- Sano, N., L. Rajjou, H.M. North, I. Debeaujon, A. Marion-Poll and M. Seo. 2016. Staying alive: Molecular aspects of seed longevity. *Plant and Cell Physiology* 57: 660–674.
- Santos M., A., H.E. Barrios V. and O.G. López Ch. 2015. Bioecología de *Hypsipyla grandella* (Zeller) (Lepidoptera: Pyralidae) e incidencia en *Cedrela odorata* L. (Meliaceae), Panamá. *Revista Científica Centros* 4: 96–115.
- Sautu, A., J.M. Baskin, C.C. Baskin and R. Condit. 2006. Studies on the seed biology of 100 native species of trees in a seasonal moist tropical forest, Panama, Central America. *Forest Ecology and Management* 234: 245–263.
- Schultz, C.J., K.L. Ferguson, J. Lahnstein and A. Bacic. 2004. Post-translational modifications of arabinogalactan-peptides of *Arabidopsis thaliana*. *Journal of Biological Chemistry* 279: 45503–45511.
- Schuster, C., C. Gaillochot and J.U. Lohmann. 2015. *Arabidopsis* HECATE genes function in phytohormone control during gynoecium development. *Development* 142: 3343–3350.
- SEMARNAT (Secretaría de Medio Ambiente y Recursos Naturales). 2010. Norma Oficial Mexicana NOM-059-SEMARNAT-2010, Protección ambiental-Especies nativas de México de flora y fauna silvestres-Categorías de riesgo y especificaciones para su inclusión, exclusión o cambio-Lista de especies en riesgo. Diario Oficial de la Federación. 30 de diciembre de 2010, Segunda Sección, México, D. F. 78 p.Mexico.
- Senkler, J., M. Senkler, H. Eubel, T. Hildebrandt, C. Lengwenus, P. Schertl, M. Schwarzlander, S. Wagner, I. Wittig and H. Braun. 2017. The mitochondrial complexome of *Arabidopsis thaliana*. *Plant Journal* 89: 1079–1092.
- Silva, E.A.A., da, M.L. Acencio, L.A. Bovolenta, N. Lemke, A. de M. Varani, J.P. Bravo, A.A. Hoshino-Bezerra and E.G. Macedo L. 2019. Gene expression during the germination of coffee seed. *Journal of Seed Science* 41: 168–179.
- Smolikova, G.N. and S.S. Medvedev. 2016. Photosynthesis in the seeds of

- chloroembryophytes. *Russian Journal of Plant Physiology* 63: 1–12.
- Solís-Guillén, I., M. Chaires-Pacheco, J. Juárez-Gómez, A. O'Connor-Sánchez and Y.J. Peña-Ramírez. 2017. Development of an ideotype-based selection tool for native tropical tree breeding by smallholder planters in Mexico's maya forest. *Small-scale Forestry* 16: 521–534.
- Soltani, N., T. Best, D. Grace, C. Nelms, K. Shumaker, J. Romero-Severson, D. Moses, S. Schuste, M. Staton, J. Carlson and K. Gwinn. 2020. Transcriptome profiles of *Quercus rubra* responding to increased O₃ stress. *BMC Genomics* 21: 1–18.
- Soneson, C., M.I. Love and M.D. Robinson. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; peer review: 2 approved]. *F1000Research* 4: 1-23.
- Soto S., J.C. y C.E. López C. 2012. RNA-seq: herramienta transcriptómica útil para el estudio de interacciones planta-patógeno. *Fitosanidad* 16: 101–113.
- Souza, P.F.N. 2020. The forgotten 2S albumin proteins: Importance, structure, and biotechnological application in agriculture and human health. *International Journal of Biological Macromolecules* 164: 4638–4649.
- Sreenivasulu, N. and U. Wobus. 2013. Seed-development programs: a systems biology-based comparison between dicots and monocots. *Annual Review of Plant Biology* 64: 189–217.
- Stavrínides, A.K., S. Dussert, M.C. Combes, I. Fock-Bastide, D. Severac, J. Minier, A. Bastos-Siqueira, V. Demolombe, S. Hem, P. Lashermes and T. Joët. 2020. Seed comparative genomics in three coffee species identify desiccation tolerance mechanisms in intermediate seeds. *Journal of Experimental Botany* 71: 1418–1433.
- Stortenbeker, N. and M. Bemer. 2019. The SAUR gene family: The plant's toolbox for adaptation of growth and development. *Journal of Experimental Botany* 70: 17–27.
- Styles, B.T. and C.G. Vosa. 1971. Chromosome Numbers in the Meliaceae. *Taxon* 20: 485–500.
- Su, P.H. and H.M. Li. 2008. Arabidopsis stromal 70-kD heat shock proteins are essential for plant development and important for thermotolerance of germinating seeds. *Plant Physiology* 146: 1231–1241.
- Supek, F., M. Bošnjak, N. Škunca and T. Šmuc. 2011. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 6: 1–9.
- TAIR (The Arabidopsis Information Resource). 2005. Available online at: <https://www.arabidopsis.org/Blast/index.jsp> Accessed: 11 de septiembre de 2022.
- Taylor-Kearney, L.J. and E. Flashman. 2021. Targeting plant cysteine oxidase activity

- for improved submergence tolerance. *The Plant Journal*: 1–10. doi: 10.1111/tpj.15605.
- The plant list. 2013. A working list of all plants species. Disponible en línea en: <http://www.theplantlist.org/1.1/browse/A/Meliaceae/Cedrela/>. Acceso: 12 de junio de 2019.
- Tokarev A., A., A. Alfonso and N. Segev. 2009. Overview of Intracellular Compartments and Trafficking Pathways. *In: Trafficking Inside Cells*. A. Alfonso, G. Paine and N. Segev (eds.). New York, USA. Springer. pp. 3–14.
- Torre, S., M. Tattini, C. Brunetti, S. Fineschi, A. Fini, A. Ferrini and F. Sebastiani. 2014. RNA-seq analysis of *Quercus pubescens* Leaves: De novo transcriptome assembly, annotation and functional markers development. *PLoS ONE* 9: e112487.
- Verdier, J., D. Lalanne, S. Pelletier, I. Torres-Jerez, K. Righetti, K. Bandyopadhyay, O. Leprince, E. Chatelain, B. Ly Vu, J. Gouzy, P. Gamas, M. Udvardi and J. Buitink. 2013. A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. *Plant Physiology* 163: 757–774.
- Wang, Z., M. Gerstein and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10: 57–63.
- Wang, H., C. Jiang, C. Wang, Y. Yang, L. Yang, X. Gao and H. Zhang. 2015. Antisense expression of the fasciclin-like arabinogalactan protein *FLA6* gene in *Populus* inhibits expression of its homologous genes and alters stem biomechanics and cell wall composition in transgenic trees. *Journal of Experimental Botany* 66: 1291–1302.
- Wang, J., N. Sun, F. Zhang, R. Yu, H. Chen, X. W. Deg and N. Wei. 2020a. SAUR17 and SAUR50 differentially regulate PP2C-D1 during apical hook development and cotyledon opening in Arabidopsis. *The Plant Cell* 32: 3792–3811.
- Wang, X., W. Tian and Y. Li. 2008. Development of an efficient protocol of RNA isolation from recalcitrant tree tissues. *Molecular Biotechnology* 38: 57–64.
- Wang, X., H. Zhang, Q. Xie, Y. Liu, H. Lv, R. Bai, R. Ma, X. Li, X. Zhang, Y.D. Guo and N. Zhang. 2020b. SISNAT interacts with HSP40, a molecular chaperone, to regulate melatonin biosynthesis and promote thermotolerance in Tomato. *Plant and Cell Physiology* 61: 909–921.
- Wei, T., Z. He, X. Tan, X. Liu, X. Yuan, Y. Luo and S. Hu. 2015. An integrated RNA-Seq and network study reveals a complex regulation process of rice embryo during seed germination. *Biochemical and Biophysical Research Communications* 464: 176–181.
- Werden, L.K., N.C. Sugii, L. Weisenberger, M.J. Keir, G. Koob and R.A. Zahawi. 2020.

- Ex situ* conservation of threatened plant species in island biodiversity hotspots: A case study from Hawai'i. *Biological Conservation* 243: 108435.
- White, T.L., W.T. Adams and D.B. Neale. 2007. *Forest Genetics*. CABI Publishing. Cambridge, MA, USA. 682 p.
- Whitehouse, K.J., F.R. Hay and C. Lusty. 2020. Why seed physiology is important for genebanking. *Plants* 9: 1–16.
- Wightman, K., B. Rodríguez, W. Sheila y J. Cornelius. 2005. Domesticación de cedro y caoba en la Península de Yucatán, México. *Experiencias en el mejoramiento del germoplasma forestal. Recursos Naturales y Ambiente* 44: 119–128.
- Wu, H., Y. Zheng, J. Liu, H. Zhang and H. Chen. 2016. Heme oxygenase-1 delays gibberellin-induced programmed cell death of rice aleurone layers subjected to drought stress by interacting with nitric oxide. *Frontiers in Plant Science* 6: 1–13.
- Xu, Q., L.L. Chen, X. Ruan, D. Chen, A. Zhu, C. Chen, D. Bertrand, W.B. Jiao, B.H. Jao, M.P. Lyon, J. Chen, C. Gao, F. Xing, H. Lan, J.W. Chang, X. Ge, Y. Lei, Q. Hu, Y. Miao, I. Wang, S. Xiao, M.K. Biswas, W. Zeng, F. Guo, H. Cao, X. Yag, X.W. Xu, Y.J. Cheng, J. Xu, J.H. Liu, O.J. Luo, T. de Zhonghui, W.W. Guo, H. Kuang, H.Y. Zhang, M.L. Rose, N. Nagarajan, X.X. Deng y and Y. Ruan. 2013. The draft genome of sweet orange (*Citrus sinensis*). *Nature Genetics* 45: 59–66.
- Yamashino, T., M. Kitayama and T. Mizuno. 2013. Transcription of ST2A encoding a sulfotransferase family protein that is involved in jasmonic acid metabolism is controlled according to the circadian clock- and PIF4/PIF5-mediated external coincidence mechanism in *Arabidopsis thaliana*. *Bioscience, Biotechnology and Biochemistry* 77: 2454–2460.
- Yang, T., Q. Yu, W. Xu, D. zhu Li, F. Chen and A. Liu. 2018. Transcriptome analysis reveals crucial genes involved in the biosynthesis of nervonic acid in woody *Malania oleifera* oilseeds. *BMC Plant Biology* 18: 1–14.
- Ye, J., Y. Zhang, H. Cui, J. Liu, Y. Wu, Y. Cheng, X. Xu, X. Huang, S. Li, A. Zhou, X. Zhang, L. Bolund, Q. Chen, J. Wang, H. Yang, L. Fang and C. Shi. 2018. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Research* 46: W71–W75.
- Yu, F., M. Li, D. He and P. Yang. 2021. Advances on Post-translational Modifications Involved in Seed Germination. *Frontiers in Plant Science* 12: 642979.
- Zang, L., T. Zheng, Y. Chu, C. Ding, W. Zhang, Q. Huang and X. Su. 2015. Genome-wide analysis of the fasciclin-like arabinogalactan protein gene family reveals differential expression patterns, localization, and salt stress response in *Populus*. *Frontiers in Plant Science* 6: 1–16.
- Zaynab, M., D. Pan, A. Noman, M. Fatima, S. Abbas, M. Umair, Y. Sharif, S. Chen and W. Chen. 2018. Transcriptome approach to address low seed germination in

- Cyclobalanopsis gilva* to save forest ecology. *Biochemical Systematics and Ecology* 81: 62–69.
- Zhang, C., W. Luo, Y. Li, X. Zhang, X. Bai, Z. Niu, X. Zhang, Z. Li and D. Wag. 2019. Transcriptomic analysis of seed germination under salt stress in two desert sister species (*Populus euphratica* and *P. pruinosa*). *Frontiers in genetics* 10: 231.
- Zhang, Y., O. Zagnitko, I. Rodionova, A. Osterman and A. Godzik. 2011. The FGGY carbohydrate kinase family: Insights into the evolution of functional specificities. *PLoS Computational Biology* 7(12): e1002318.
- Zhao, Q. 2016. Lignification: Flexibility, Biosynthesis and Regulation. *Trends in Plant Science* 21: 713–721.
- Zhao, Y., D. Song, J. Sun and L. Li. 2013. *Populus* endo-beta-mannanase PtrMAN6 plays a role in coordinating cell wall remodeling with suppression of secondary wall thickening through generation of oligosaccharide signals. *Plant Journal* 74: 473–485.
- Zhong, C., H. Xu, S. Ye, S. Wang, L. Li, S. Zhang and X. Wang. 2015. Gibberellic acid-stimulated *Arabidopsis6* serves as an integrator of gibberellin, abscisic acid, and glucose signaling during seed germination in *Arabidopsis*. *Plant Physiology* 169: 2288–2303.
- Zinsmeister, J., O. Leprince, and J. Buitink. 2020. Molecular and environmental factors regulating seed longevity. *Biochemical Journal* 477: 305–323.
- Zobel, B.J. y T.J. Talbert. 1988. Técnicas de mejoramiento genético de árboles forestales Editorial Limusa. México, D. F. 545 p.

ANEXOS

Cuadro A1. Resumen de los análisis de calidad de los 18 archivos *fastq* correspondientes a las tres condiciones evaluadas: seca, embebida al 95% y germinada, cada una con sus tres repeticiones y cada repetición con su respectivo *Read 1* y *Read 2* en semillas de *Cedrela odorata* L.

Archivo	Estadísticas básicas	Calidad por secuencia de base	Calidad por secuencia de mosaicos	Puntuaciones de calidad por secuencia	Contenido de secuencia por base	Contenido de GC por secuencia	Porcentaje de contenido de N base	Distribución de longitud de secuencia	Niveles de duplicación de secuencia	Secuencias sobrerrepresentadas	Contenido del adaptador
FC1_R1_fastqc											
FC1_R2_fastqc											
FC2_R1_fastqc											
FC2_R2_fastqc											
FC3_R1_fastqc											
FC3_R2_fastqc											
FC4_R1_fastqc											
FC4_R2_fastqc											
FC5_R1_fastqc											
FC5_R2_fastqc											
FC6_R1_fastqc											
FC6_R2_fastqc											
FC7_R1_fastqc											
FC7_R2_fastqc											
FC8_R1_fastqc											
FC8_R2_fastqc											
FC9_R1_fastqc											
FC9_R2_fastqc											

Nota: De FC1 a FC3 corresponde a la muestra semilla seca, de FC4 a FC6 a la muestra semilla embebida al 95% y de FC7 a FC9 a la muestra semilla germinada. R1 corresponde a la muestra *Read 1* y R2 a la muestra *Read 2*.

Cuadro A2. Resumen de la estadística básica de los 18 archivos fastq correspondientes a las tres condiciones evaluadas: seca, embebida al 95% y germinada, cada una con sus tres repeticiones y cada repetición con su respectivo *Read 1* y *Read 2* en semillas de *Cedrela odorata* L.

Condición evaluada	Nombre del archivo	Secuencias totales	Secuencias marcadas como de mala calidad	Longitud de la secuencia	%GC
Seca_1	FC1_R1.fastq	12952485	0	75	49
Seca_1	FC1_R2.fastq	12952485	0	75	50
Seca_2	FC2_R1.fastq	12434959	0	75	50
Seca_2	FC2_R2.fastq	12434959	0	75	51
Seca_3	FC3_R1.fastq	14260286	0	75	49
Seca_3	FC3_R2.fastq	14260286	0	75	49
Embebida_1	FC4_R1.fastq	8762203	0	75	46
Embebida_1	FC4_R2.fastq	8762203	0	75	47
Embebida_2	FC5_R1.fastq	10822913	0	75	45
Embebida_2	FC5_R2.fastq	10822913	0	75	47
Embebida_3	FC6_R1.fastq	8567163	0	75	47
Embebida_3	FC6_R2.fastq	8567163	0	75	49
Germinada_1	FC7_R1.fastq	10090716	0	75	45
Germinada_1	FC7_R2.fastq	10090716	0	75	46
Germinada_2	FC8_R1.fastq	11264817	0	75	45
Germinada_2	FC8_R2.fastq	11264817	0	75	46
Germinada_3	FC9_R1.fastq	10296252	0	75	44
Germinada_3	FC9_R2.fastq	10296252	0	75	46

Cuadro A3. Anotación funcional con *Trinotate* del transcriptoma completo de semillas de *Cedrela odorata* L. en las condiciones seca, embebida al 95% y germinada.

Debido a la complejidad de este cuadro que consta de 95,739 datos, para su visualización, vaya al siguiente enlace: <https://osf.io/kfn8q/>

Los significados de las columnas en cada encabezado del cuadro son los siguientes:

- **#gene_id:** Identificación del gen.
- **transcript_id:** Identificación de la transcripción.
- **sprot_Top_BLASTX_hit:** Mejor resultado de aciertos de Blastx de las búsquedas de homología en la base de datos del NCBI. El alineamiento usa como entrada una secuencia de nucleótidos y compara estas secuencias traducidas contra una base de datos de proteínas.
- **prot_id:** Identificación del ORF (Marco de lectura abierto) más largo.
- **prot_coords:** Ubicación del ORF más largo.
- **sprot_Top_BLASTP_hit:** Mejor resultado de aciertos de *Blastp* de las búsquedas de homología en la base de datos del NCBI. El alineamiento compara una secuencia de aminoácidos contra una base de datos del mismo tipo.
- **Pfam:** Resultado de la búsqueda de identificación de familia de proteínas.
- **Eggnog:** Resultado de la búsqueda de la base de datos de genes ortólogos
- **Kegg:** Resultado de la búsqueda de identificación basada en la enciclopedia de genes y genomas de Kioto que es una colección de bases de datos en línea de genomas, rutas enzimáticas y químicos biológicos.
- **gene_ontology_BLASTX:** Resultado de la búsqueda de ontología génica basada en *Blastx*.
- **gene_ontology_BLASTP:** Resultado de la búsqueda de ontología génica basada en *Blastp*.
- **gene_ontology_Pfam:** Resultado de la búsqueda de ontología génica basada en *Pfam*.

Cuadro A4. Análisis de genes diferencialmente expresados en semillas de *Cedrela odorata* L. en la condición SvsG.

ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
TRINITY_DN1781_c0_g1	1648.74908	15.30622237	1.88863807	8.10437034	5.30E-16	1.13E-14
TRINITY_DN1491_c0_g1	1260.13463	14.91894827	2.08763579	7.1463367	8.91E-13	1.33E-11
TRINITY_DN532_c0_g2	1221.74885	14.87436374	1.8004824	8.26132139	1.44E-16	3.26E-15
TRINITY_DN1063_c0_g1	2345.86002	14.85302641	1.94499807	7.63652502	2.23E-14	3.99E-13
TRINITY_DN2650_c0_g1	1181.6376	14.82509964	1.3633074	10.8743631	1.53E-27	8.89E-26
TRINITY_DN2602_c0_g1	1144.96584	14.78034733	1.64245369	8.99894312	2.28E-19	6.72E-18
TRINITY_DN1151_c0_g1	1118.52146	14.74590358	1.65674642	8.90051936	5.56E-19	1.59E-17
TRINITY_DN533_c0_g1	983.888145	14.56063629	1.31260173	11.0929583	1.36E-28	8.29E-27
TRINITY_DN4427_c0_g1	950.400874	14.51210762	1.8641555	7.78481603	6.98E-15	1.32E-13
TRINITY_DN2591_c0_g1	844.167345	14.33665003	1.42833346	10.037327	1.04E-23	4.43E-22
TRINITY_DN561_c0_g1	759.694099	14.18835052	2.12456369	6.67824201	2.42E-11	2.98E-10
TRINITY_DN3006_c0_g1	750.590916	14.17088363	1.33116256	10.6454944	1.83E-26	9.80E-25
TRINITY_DN610_c0_g1	732.390309	14.13534011	1.84524946	7.6603952	1.85E-14	3.34E-13
TRINITY_DN853_c0_g1	728.670507	14.12819211	1.82074308	7.75957481	8.52E-15	1.59E-13
TRINITY_DN9346_c0_g1	713.645698	14.09805628	1.71875933	8.20246093	2.36E-16	5.22E-15
TRINITY_DN532_c0_g1	3709.88936	14.02914123	1.05659999	13.2776276	3.12E-40	3.31E-38
TRINITY_DN4517_c0_g1	624.472134	13.90553306	2.0810969	6.68182873	2.36E-11	2.91E-10
TRINITY_DN1449_c0_g1	607.004421	13.86481685	1.28398685	10.7982545	3.51E-27	1.97E-25
TRINITY_DN3408_c0_g1	605.720886	13.86176504	1.34189035	10.3300281	5.15E-25	2.45E-23
TRINITY_DN10265_c0_g1	603.918723	13.85735153	1.96655586	7.04650796	1.83E-12	2.63E-11

Cuadro A4. Análisis de genes diferencialmente expresados en semillas de *Cedrela odorata* L. en la condición SvsG.

ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
TRINITY_DN7220_c0_g1	3380.58272	-13.06063157	1.2337005	-10.5865496	3.44E-26	1.81E-24
TRINITY_DN21_c0_g2	479.654787	-12.18421919	1.28849884	-9.45613515	3.20E-21	1.09E-19
TRINITY_DN3352_c0_g1	540.324675	-12.15267117	2.48175186	-4.89681155	9.74E-07	5.90E-06
TRINITY_DN1547_c0_g1	8001.93165	-12.11074916	0.66419156	-18.2338197	2.78E-74	9.99E-72
TRINITY_DN6100_c1_g1	445.347883	-11.92485207	1.26838738	-9.40158524	5.37E-21	1.81E-19
TRINITY_DN3463_c0_g2	358.06883	-11.37506295	1.24456901	-9.13976073	6.26E-20	1.93E-18
TRINITY_DN539_c0_g1	3297.67339	-11.22344159	0.73891457	-15.1890923	4.18E-52	6.98E-50
TRINITY_DN1434_c0_g2	691.077992	-11.22186415	1.41628089	-7.92347352	2.31E-15	4.59E-14
TRINITY_DN8106_c0_g1	222.659881	-11.02537397	1.48744484	-7.41229095	1.24E-13	2.03E-12
TRINITY_DN5420_c0_g1	437.439725	-10.9241166	1.90731606	-5.72748104	1.02E-08	8.59E-08
TRINITY_DN10415_c0_g1	663.193323	-10.89809876	1.81856652	-5.99268634	2.06E-09	1.93E-08
TRINITY_DN1085_c0_g2	167.516576	-10.65137861	1.41008968	-7.55368877	4.23E-14	7.28E-13
TRINITY_DN1243_c0_g3	203.477514	-10.57751157	1.35292115	-7.81827645	5.36E-15	1.02E-13
TRINITY_DN35793_c0_g1	145.402229	-10.46126136	1.30893138	-7.99221527	1.33E-15	2.71E-14
TRINITY_DN2053_c0_g1	143.988128	-10.39575298	1.43285726	-7.2552607	4.01E-13	6.19E-12
TRINITY_DN1434_c0_g1	889.455544	-10.37366116	1.32037037	-7.85662976	3.95E-15	7.63E-14
TRINITY_DN7800_c0_g1	919.942369	-10.36026723	0.88200118	-11.746319	7.38E-32	5.44E-30
TRINITY_DN39594_c0_g1	136.738638	-10.35824071	1.51537331	-6.83543829	8.18E-12	1.07E-10
TRINITY_DN9137_c0_g1	653.265118	-10.20081863	0.97082613	-10.507359	7.99E-26	4.08E-24
TRINITY_DN1061_c0_g3	121.559931	-10.01741522	1.27256154	-7.87185134	3.49E-15	6.80E-14

Nota :Esta tabla muestra los 40 genes con mayor nivel de expresión (log2FoldChange), 20 aumentados y 20 disminuidos. La tabla original cuenta con 7667 datos que pueden visualizarse en el siguiente enlace: <https://osf.io/kfn8q/>.

Cuadro A5. Análisis de genes diferencialmente expresados en semillas de *Cedrela odorata* L. en la condición SvsE.

ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
TRINITY_DN8755_c0_g1	155.4272035	25.28813471	3.6701881	6.89014678	5.57E-12	1.25E-10
TRINITY_DN25351_c0_g1	33.30296631	22.30833578	3.9100627	5.70536523	1.16E-08	1.54E-07
TRINITY_DN532_c0_g1	3709.889363	13.59730323	1.04230849	13.0453732	6.75E-39	1.68E-36
TRINITY_DN1449_c0_g1	607.0044214	13.47337556	1.28398685	10.4933906	9.26E-26	8.82E-24
TRINITY_DN41615_c0_g1	448.5516055	13.03734985	3.90686774	3.33703384	0.00084678	0.00351387
TRINITY_DN7381_c0_g1	417.3422393	12.93345324	3.00499306	4.30398773	1.68E-05	0.00011119
TRINITY_DN12634_c0_g1	372.0099563	12.76751099	2.81420713	4.53680571	5.71E-06	4.27E-05
TRINITY_DN2856_c0_g1	340.6967786	12.64064884	2.64752299	4.77451901	1.80E-06	1.52E-05
TRINITY_DN1151_c0_g2	339.9498694	12.63758061	2.21962608	5.69356285	1.24E-08	1.64E-07
TRINITY_DN532_c0_g2	1221.748847	12.55195974	1.68613197	7.44423329	9.75E-14	2.80E-12
TRINITY_DN3408_c0_g1	605.7208855	12.50737033	1.33850752	9.34426605	9.25E-21	5.88E-19
TRINITY_DN4427_c0_g1	950.400874	12.49714811	1.80643964	6.91811	4.58E-12	1.04E-10
TRINITY_DN2618_c0_g2	306.6665074	12.4880501	1.93534768	6.45261325	1.10E-10	2.05E-09
TRINITY_DN5285_c0_g1	271.2605612	12.3118176	2.27554896	5.41048241	6.29E-08	7.26E-07
TRINITY_DN4154_c0_g1	272.1254305	12.27168794	2.34136426	5.24125535	1.59E-07	1.70E-06
TRINITY_DN1993_c0_g1	258.2549294	12.24110566	2.35437369	5.19930448	2.00E-07	2.08E-06
TRINITY_DN5119_c0_g1	467.0243966	12.1363563	1.84382641	6.58215775	4.64E-11	9.15E-10
TRINITY_DN1491_c0_g1	1260.13463	12.03258258	1.91003262	6.29967388	2.98E-10	5.19E-09
TRINITY_DN39403_c0_g1	203.2205685	11.89514336	3.90700701	3.04456668	0.00233016	0.00841736
TRINITY_DN4322_c0_g1	201.5871695	11.88362369	2.48774465	4.77686635	1.78E-06	1.50E-05

Cuadro A5. Análisis de genes diferencialmente expresados en semillas de *Cedrela odorata* L. en la condición SvsE.

ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
TRINITY_DN7220_c0_g1	3380.582717	-13.01946506	1.2337147	-10.5530599	4.92E-26	4.80E-24
TRINITY_DN1434_c0_g1	889.4555439	-11.60018605	1.32027822	-8.78616782	1.55E-18	7.83E-17
TRINITY_DN1547_c0_g1	8001.931646	-10.71969852	0.66424234	-16.1382343	1.37E-58	1.27E-55
TRINITY_DN10415_c0_g1	663.1933226	-10.44717892	1.81865998	-5.74443767	9.22E-09	1.24E-07
TRINITY_DN2808_c0_g1	372.4487101	-10.39174374	1.21439348	-8.55714719	1.16E-17	5.37E-16
TRINITY_DN1525_c0_g1	206.3282107	-10.16435147	2.09919326	-4.84202748	1.29E-06	1.12E-05
TRINITY_DN1434_c0_g2	691.0779923	-10.04278787	1.41651388	-7.08979138	1.34E-12	3.29E-11
TRINITY_DN3463_c0_g2	358.0688295	-9.681016724	1.24532762	-7.77387136	7.61E-15	2.52E-13
TRINITY_DN3352_c0_g1	540.3246749	-9.433716138	2.4822726	-3.80043519	0.00014444	0.00075713
TRINITY_DN1423_c0_g2	2080.738221	-9.334388169	0.76287898	-12.2357391	2.00E-34	4.06E-32
TRINITY_DN37666_c0_g1	66.53702492	-9.310539752	1.83772807	-5.06633157	4.06E-07	3.97E-06
TRINITY_DN539_c0_g1	3297.673391	-9.296546176	0.73907927	-12.5785509	2.77E-36	6.14E-34
TRINITY_DN2060_c0_g1	643.2863234	-9.169607381	1.10569164	-8.29309642	1.10E-16	4.62E-15
TRINITY_DN12162_c0_g1	2757.053368	-9.050611908	0.57888442	-15.6345751	4.23E-55	2.70E-52
TRINITY_DN12069_c0_g1	306.3849491	-9.034669687	1.30034416	-6.94790655	3.71E-12	8.54E-11
TRINITY_DN41835_c0_g1	100.4370778	-8.842154754	1.69094714	-5.22911364	1.70E-07	1.80E-06
TRINITY_DN1243_c0_g3	203.477514	-8.803695776	1.35422421	-6.50091449	7.98E-11	1.52E-09
TRINITY_DN6100_c1_g1	445.347883	-8.766490889	1.27005441	-6.902453	5.11E-12	1.16E-10
TRINITY_DN772_c0_g1	1842.271821	-8.760211068	0.67543504	-12.9697314	1.82E-38	4.44E-36
TRINITY_DN2314_c0_g1	80.19367253	-8.664018555	2.59741115	-3.33563616	0.00085104	0.00352871

Nota: Esta tabla muestra los 40 genes con mayor nivel de expresión (log2FoldChange), 20 aumentados y 20 disminuidos. La tabla original cuenta con 5,530 datos que pueden visualizarse en el siguiente enlace: <https://osf.io/kfn8q/>

Cuadro A6. Análisis de genes diferencialmente expresados en semillas de *Cedrela odorata* L. en la condición EvsG.

ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
TRINITY_DN15588_c0_g1	18.0902303	8.795020559	1.43452789	6.13095126	8.74E-10	3.58E-08
TRINITY_DN10762_c0_g2	18.0266892	7.68613469	2.20848956	3.48026761	0.00050091	0.00391572
TRINITY_DN18720_c0_g2	6.36701858	7.153228658	2.15051935	3.3262796	0.00088014	0.00624984
TRINITY_DN27107_c0_g2	5.0729079	6.75395573	1.78537287	3.78293849	0.00015499	0.00147255
TRINITY_DN4104_c0_g1	80.3197691	6.655282904	1.25657697	5.29635912	1.18E-07	2.89E-06
TRINITY_DN45204_c0_g1	10.4857931	6.485735029	1.4271432	4.54455799	5.51E-06	8.32E-05
TRINITY_DN39528_c0_g1	35.7382298	6.339921839	1.75706668	3.60824203	0.00030828	0.00261249
TRINITY_DN3860_c0_g2	25.9119838	6.29259033	1.80839883	3.47964742	0.00050207	0.00392364
TRINITY_DN7710_c0_g1	255.536004	6.074412469	1.13123999	5.36969387	7.89E-08	2.02E-06
TRINITY_DN13150_c0_g1	8.17872014	6.073203252	1.58758296	3.82543992	0.00013054	0.00126566
TRINITY_DN2591_c0_g1	844.167345	6.042434411	1.45472858	4.15365071	3.27E-05	0.0003915
TRINITY_DN21981_c0_g1	42.5321756	6.020416893	1.33035536	4.52542009	6.03E-06	8.99E-05
TRINITY_DN37666_c0_g1	66.5370249	5.907506541	1.45962224	4.04728456	5.18E-05	0.00057662
TRINITY_DN125_c0_g1	1665.76546	5.835749054	0.76540999	7.62434397	2.45E-14	2.68E-12
TRINITY_DN14449_c0_g1	32.7528982	5.717416299	1.65642014	3.45167036	0.00055713	0.00427078
TRINITY_DN16248_c0_g1	222.820806	5.71009604	0.70244335	8.12890606	4.33E-16	6.56E-14
TRINITY_DN4518_c0_g1	180.924355	5.680515418	1.11198977	5.10842416	3.25E-07	7.04E-06
TRINITY_DN45135_c0_g1	33.3903853	5.648502437	1.75942926	3.21041747	0.00132542	0.00869982
TRINITY_DN26175_c0_g1	5.35961858	5.636761302	1.67786473	3.35948495	0.00078088	0.00565098
TRINITY_DN19282_c0_g1	9.88843293	5.610352893	1.5633591	3.58865273	0.00033239	0.00277688

Cuadro A6. Análisis de genes diferencialmente expresados en semillas de *Cedrela odorata* L. en la condición EvsG.

ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
TRINITY_DN25351_c0_g1	33.3029663	-22.48676688	3.90791734	-5.75415622	8.71E-09	2.84E-07
TRINITY_DN8755_c0_g1	155.427204	-22.09506309	3.67120371	-6.01847919	1.76E-09	6.61E-08
TRINITY_DN21_c0_g2	479.654787	-11.27144125	1.13323213	-9.94627751	2.62E-23	1.16E-20
TRINITY_DN10163_c0_g3	73.8834888	-10.43572206	2.93725381	-3.55288399	0.00038103	0.00311645
TRINITY_DN5865_c0_g1	58.8945879	-10.09775915	1.4698681	-6.86984034	6.43E-12	4.20E-10
TRINITY_DN34_c0_g2	167.136883	-10.02233687	2.31248125	-4.33401865	1.46E-05	0.00019435
TRINITY_DN8938_c0_g1	48.8504634	-9.834055059	2.45626418	-4.00366343	6.24E-05	0.00067591
TRINITY_DN4463_c0_g2	44.107151	-9.682507582	2.71394888	-3.56768238	0.00036015	0.00296857
TRINITY_DN5653_c0_g1	41.7724391	-9.612893493	1.48295992	-6.48223419	9.04E-11	4.62E-09
TRINITY_DN35793_c0_g1	145.402229	-9.611456385	1.17622366	-8.17145301	3.05E-16	4.78E-14
TRINITY_DN12967_c0_g1	41.4168918	-9.595516673	1.5401645	-6.2301895	4.66E-10	2.03E-08
TRINITY_DN22040_c0_g1	41.6087277	-9.572696151	1.65160721	-5.79598835	6.79E-09	2.28E-07
TRINITY_DN21743_c0_g1	38.2150323	-9.484497742	1.39414728	-6.80308162	1.02E-11	6.36E-10
TRINITY_DN7150_c1_g1	36.0578301	-9.400829586	1.56648263	-6.00123449	1.96E-09	7.31E-08
TRINITY_DN36489_c0_g1	34.6190875	-9.341903724	1.53365791	-6.09125649	1.12E-09	4.40E-08
TRINITY_DN13342_c0_g1	30.4644378	-9.136654002	1.35734395	-6.73127398	1.68E-11	1.00E-09
TRINITY_DN5355_c0_g3	29.6566538	-9.119003319	1.51330858	-6.02587167	1.68E-09	6.34E-08
TRINITY_DN7121_c0_g1	29.5745843	-9.114609357	1.42814763	-6.38211987	1.75E-10	8.45E-09
TRINITY_DN8136_c1_g1	28.0976432	-9.015666124	1.62642299	-5.54324808	2.97E-08	8.48E-07
TRINITY_DN3380_c0_g1	27.939804	-9.00810966	1.53122312	-5.8829504	4.03E-09	1.41E-07

Nota: Esta tabla muestra los 40 genes con mayor nivel de expresión (log2FoldChange), 20 aumentados y 20 disminuidos. La tabla original cuenta con 1420 datos que pueden visualizarse en el siguiente enlace: <https://osf.io/kfn8g/>

Cuadro A7. Anotación funcional manual apoyados de las bases de datos de NCBI y TAIR en semillas de *Cedrelela odorata* L.

El cuadro muestra los 50 genes que más aumentaron y 50 que más disminuyeron sus niveles de expresión en cada comparación: SvsE, EvsG y SvsG haciendo un total de 247 genes. Para su visualización, vaya al siguiente enlace: <https://osf.io/kfn8q/>

Los significados de las columnas en cada encabezado del cuadro son los siguientes:

- **ID:** Identificación del gen.
- **Grupo en mapa de calor:** Grupo al que pertenece en el mapa de calor de la Figura 2.3.
- **log2FoldChange SvsE:** Cambio relativo en la expresión de la comparación Seca vs Embebida.
- **log2FoldChange EvsG:** Cambio relativo en la expresión de la comparación Embebida vs Germinada.
- **log2FoldChange SvG:** Cambio relativo en la expresión de la comparación Seca vs Germinada.
- **Identificación de Arabidopsis:** Clave de identificación en la base de datos de TAIR.
- **E-value:** Valor E de la anotación dada en la base de datos de TAIR.
- **Porcentaje de identidad:** Porcentaje de identidad de la anotación dada en la base de datos de TAIR.
- **Otros nombres:** Otros nombres con los que se puede identificar al mismo gen.
- **Descripciones de genes:** Breve descripción del gen por la base de datos de TAIR.
- **GO PB:** Ontología génica de los procesos biológicos dados por la base de datos de TAIR.
- **GO CC:** Ontología génica de los componentes celulares dado por la base de datos de TAIR.
- **GO FM:** Ontología génica de la función molecular dado por la base de datos de TAIR.
- **Identificación NCBI;** Clave de identificación en la base de datos de NCBI.
- **E-value:** Valor E de la anotación dada en la base de datos de NCBI.
- **Porcentaje de identidad:** Porcentaje de identidad de la anotación dada en la base de datos de NCBI.
- **Descripciones de genes:** Breve descripción del gen por la base de datos de NCBI.

Cuadro A8. Resultados de enriquecimiento GO de los PB de los 100 genes (50 aumentados y 50 disminuidos) con mayor nivel de expresión (*log2FoldChange*) en la comparación SvsE en semillas de *Cedrela odorata* L.

Término ID	Nombre	Frecuencia	Unicidad	Dispensabilidad	Representante
GO:0006979	"respuesta al estrés oxidativo"	1.979	0.849	0	nulo
GO:0071493	"respuesta celular a UV-B"	0.032	0.848	0.156	"respuesta al estrés oxidativo"
GO:0046686	"respuesta al ion cadmio"	1.316	0.846	0.226	"respuesta al estrés oxidativo"
GO:0009630	"gravitropismo"	0.433	0.843	0.32	"respuesta al estrés oxidativo"
GO:0009741	"respuesta a brasinoesteroide"	0.491	0.837	0.361	"respuesta al estrés oxidativo"
GO:0098869	"desintoxicación de oxidantes celulares"	0.775	0.791	0.379	"respuesta al estrés oxidativo"
GO:0034605	"respuesta celular al calor"	0.298	0.797	0.383	"respuesta al estrés oxidativo"
GO:0006952	"respuesta a defensa"	7.096	0.829	0.456	"respuesta al estrés oxidativo"
GO:0042742	"respuesta de defensa a la bacteria"	1.821	0.807	0.487	"respuesta al estrés oxidativo"
GO:0071456	"respuesta celular a la hipoxia"	1.037	0.742	0.532	"respuesta al estrés oxidativo"
GO:0071398	"respuesta celular a los ácidos grasos"	0.334	0.784	0.561	"respuesta al estrés oxidativo"
GO:0009740	"vía de señalización mediada por ácido giberélico"	0.347	0.741	0.67	"respuesta al estrés oxidativo"
GO:0010047	"dehiscencia de frutos"	0.05	0.966	0	nulo
GO:0048462	"formación de carpelos"	0.018	0.953	0.291	"dehiscencia de frutos"
GO:0010154	"desarrollo de frutos"	2.727	0.942	0.467	"dehiscencia de frutos"
GO:0010608	"regulación postranscripcional de la expresión génica"	2.011	0.955	0	nulo
GO:0031222	"proceso catabólico de arabinosa"	0.036	0.919	0	nulo
GO:0048227	"transporte de la membrana plasmática al endosoma"	0.188	0.951	0	nulo
GO:0009926	"transporte polar de auxina"	0.302	0.94	0.311	"Transporte de la membrana plasmática al endosoma"
GO:0030643	"homeostasis de iones de fosfato celular"	0.014	0.935	0.408	"Transporte de la membrana plasmática al endosoma"
GO:0080171	"organización vacuola lítica"	0.014	0.894	0.022	nulo

Cuadro A8. Resultados de enriquecimiento GO de los PB de los 100 genes (50 aumentados y 50 disminuidos) con mayor nivel de expresión ($\log_2\text{FoldChange}$) en la comparación SvsE en semillas de *Cedrela odorata* L.

Término ID	Nombre	Frecuencia	Unicidad	Dispensabilidad	Representante
GO:0033617	"ensamblaje de citocromo c oxidasa mitocondrial"	0.063	0.853	0.285	"organización de vacuolas líticas"
GO:0051567	"metilación de histona H3-K9"	0.036	0.812	0.304	"organización de vacuolas líticas"
GO:0071555	"organización de la pared celular"	2.371	0.829	0.325	"organización de vacuolas líticas"
GO:0051259	"oligomerización de complejos proteicos"	0.261	0.846	0.533	"organización de vacuolas líticas"
GO:0006346	"Ensamblaje de heterocromatina dependiente de la metilación del ADN"	0.032	0.77	0.58	"organización de vacuolas líticas"
GO:0010215	"organización de microfibrillas de celulosa"	0.077	0.84	0.653	"organización de vacuolas líticas"
GO:0045168	"señalización célula-célula involucrada en el compromiso del destino celular"	0.167	0.91	0.027	nulo
GO:0002220	"Respuesta inmune innata que activa la vía de señalización del receptor de la superficie celular"	0.009	0.815	0.292	"señalización célula-célula involucrada en el compromiso del destino celular"
GO:0051555	"proceso biosintético de flavonoles"	0.081	0.902	0.068	nulo
GO:0006499	"Miristoilación de proteína N-terminal"	0.018	0.84	0.218	"proceso biosintético de flavonoles"
GO:0009809	"proceso biosintético de lignina"	0.18	0.873	0.256	"proceso biosintético de flavonoles"
GO:0006412	"traducción"	2.831	0.809	0.414	"proceso biosintético de flavonoles"
GO:0016311	"desfosforilación"	1.578	0.899	0.075	nulo
GO:0009694	"proceso metabólico del ácido jasmónico"	0.194	0.872	0.108	"desfosforilación"
GO:0009060	"respiración aeróbica"	0.582	0.892	0.12	"desfosforilación"
GO:0005975	"proceso metabólico de carbohidratos"	5.058	0.926	0.127	"desfosforilación"
GO:0031408	"proceso biosintético de oxilipina"	0.104	0.849	0.462	"desfosforilación"
GO:0006542	"proceso biosintético de glutamina"	0.032	0.841	0.502	"desfosforilación"
GO:0005980	"proceso catabólico del glucógeno"	0.009	0.872	0.668	"desfosforilación"

Cuadro A9. Resultados de enriquecimiento GO de los PB de los 100 genes (50 aumentados y 50 disminuidos) con mayor nivel de expresión ($\log_2\text{FoldChange}$) en la comparación EvsG en semillas de *Cedreia odorata* L.

Término ID	Nombre	Frecuencia	Unicidad	Dispensabilidad	Representante
GO:0009733	" respuesta a la auxina "	1.785	0.85	0	nulo
GO:0071483	"respuesta celular a la luz azul"	0.122	0.823	0.174	"respuesta a la auxina"
GO:0009651	"respuesta al estrés salino"	1.961	0.798	0.406	"respuesta a la auxina"
GO:0009738	" vía de señalización activada por ácido abscísico "	0.947	0.741	0.579	"respuesta a la auxina"
GO:0009768	" fotosíntesis, recolección de luz en el fotosistema I "	0.104	0.974	0	nulo
GO:0043086	" regulación negativa de la actividad catalítica "	1.109	0.932	0	nulo
GO:0045905	" regulación positiva de la terminación de la traducción "	0.014	0.89	0.127	" regulación negativa de la actividad catalítica "
GO:0045893	" regulación positiva de la transcripción, plantilla de ADN "	2.493	0.868	0.557	"regulación negativa de la actividad catalítica"
GO:0048316	" desarrollo de semillas "	2.574	0.952	0	null
GO:0042335	" desarrollo de la cutícula "	0.104	0.952	0.364	" desarrollo de semillas "
GO:1990544	" transporte transmembrana de ATP mitocondrial "	0.023	0.937	0.022	nulo
GO:0055085	" transporte transmembrana "	5.973	0.924	0.335	" transporte transmembrana de ATP mitocondrial "
GO:0007017	" proceso basado en microtúbulos "	1.01	0.968	0.03	nulo
GO:0030036	" organización del citoesqueleto de actina "	0.442	0.931	0.034	nulo
GO:0042546	" biogénesis de la pared celular "	1.055	0.929	0.335	" organización del citoesqueleto de actina "

Cuadro A10. Resultados de enriquecimiento GO de los PB de los 100 genes (50 aumentados y 50 disminuidos) con mayor nivel de expresión (*log2FoldChange*) en la comparación SvsG en semillas de *Cedrela odorata* L.

Término ID	Nombre	Frecuencia	Unicidad	Dispensabilidad	Representante
GO:0051555	"proceso biosintético de flavonol"	0.081	0.966	0	nulo
GO:0080126	" desarrollo del septum ovárico "	0.005	0.957	0	nulo
GO:0010074	mantenimiento de la identidad del meristemo "	0.334	0.941	0.228	" desarrollo del septum ovárico "
GO:0048767	" alargamiento del pelo de raíz "	0.311	0.899	0.609	" desarrollo del septum ovárico "
GO:0080165	" depósito de calosa en la placa de tamiz del floema "	0.014	0.975	0	nulo
GO:0045037	" importación de proteínas en el estroma del cloroplasto "	0.117	0.944	0.352	" depósito de calosa en la placa de tamiz del floema "
GO:0080171	" organización de vacuolas líticas "	0.014	0.922	0	nulo
GO:0033617	" ensamblaje del citocromo c oxidasa mitocondrial "	0.063	0.905	0.285	" organización de vacuolas líticas "
GO:0051567	" metilación de histona H3-K9 "	0.036	0.861	0.304	" organización de vacuolas líticas "
GO:0042546	" biogénesis de la pared celular "	1.055	0.903	0.434	" organización de vacuolas líticas "
GO:1902290	" regulación positiva de la respuesta de defensa a los oomicetos "	0.045	0.911	0	nulo
GO:0010608	" regulación postranscripcional de la expresión génica "	2.011	0.905	0.168	" regulación positiva de la respuesta de defensa a los oomicetos "
GO:1901430	" regulación positiva del proceso biosintético de lignina siringal "	0.009	0.908	0.305	" regulación positiva de la respuesta de defensa a los oomicetos "
GO:1900370	" regulación positiva del ARN interferencia "	0.018	0.902	0.318	" regulación positiva de la respuesta de defensa a los oomicetos "
GO:0071494	" respuesta celular a UV-C "	0.018	0.833	0.019	nulo

Cuadro A10. Resultados de enriquecimiento GO de los PB de los 100 genes (50 aumentados y 50 disminuidos) con mayor nivel de expresión ($\log_2\text{FoldChange}$) en la comparación SvsG en semillas de *Cedrela odorata* L.

Término ID	Nombre	Frecuencia	Unicidad	Dispensabilidad	Representante
GO:0071732	" respuesta celular al óxido nítrico "	0.018	0.82	0.198	" respuesta celular a UV-C "
GO:0048015	" señalización mediada por fosfatidilinositol "	0.068	0.772	0.215	" respuesta celular a UV-C "
GO:0009733	" respuesta a la auxina "	1.785	0.813	0.276	" respuesta celular a UV-C "
GO:0009651	" respuesta al estrés salino "	1.961	0.825	0.344	" respuesta celular a UV-C "
GO:0098869	" desintoxicación de oxidantes celulares "	0.775	0.762	0.392	" respuesta celular a UV-C "
GO:0006952	" respuesta de defensa "	7.096	0.827	0.455	" respuesta celular a UV-C "
GO:0071456	" respuesta celular a la hipoxia "	1.037	0.724	0.521	" respuesta celular a UV-C "
GO:0009873	" vía de señalización activada por etileno "	0.875	0.678	0.581	" respuesta celular a UV-C "
GO:0007165	" transducción de señales "	8.038	0.682	0.632	" respuesta celular a UV-C "
GO:0009768	" fotosíntesis, recolección de luz en el fotosistema I "	0.104	0.941	0.061	nulo
GO:0016311	" desfosforilación "	1.578	0.926	0.102	" fotosíntesis, recolección de luz en el fotosistema I "
GO:0006508	" proteólisis "	4.909	0.915	0.119	" fotosíntesis, recolección de luz en el fotosistema I "
GO:0009116	" proceso metabólico de nucleósidos "	0.261	0.913	0.241	" fotosíntesis, recolección de luz en el fotosistema I "
GO:0016567	" ubiquitinación de proteínas "	3.688	0.877	0.424	fotosíntesis recolección de luz en el fotosistema I "

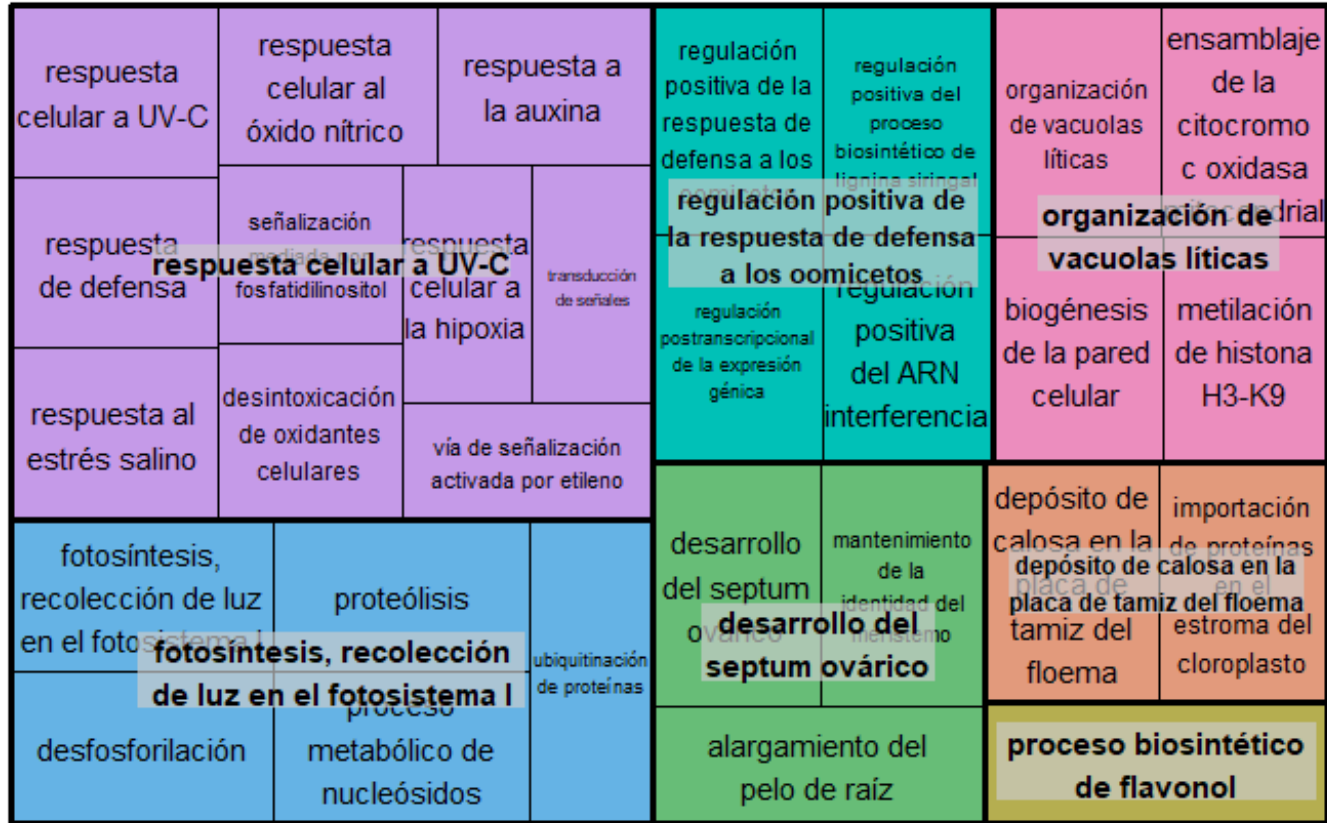


Figura A1. Análisis REVIGO de los procesos biológicos en 100 genes expresados diferencialmente (DEG), regulados al alza y a la baja en la comparación Seca vs Germinada (SvsG). Cada rectángulo representa un grupo. Los representantes se clasifican en "supergrupos" de términos vagamente relacionados, reflejados por diferentes colores. La frecuencia del término GO se refleja en el tamaño de los rectángulos.