

Review [Revisión]



GENOMIC AND TRANSCRIPTOMIC APPROACHES TO STUDY TIMBER TREES: PERSPECTIVES TO STUDY RED CEDAR (*Cedrela odorata* L.) †

[ENFOQUES GENÓMICOS Y TRANSCRIPTÓMICOS PARA ESTUDIAR ÁRBOLES MADERABLES: PERSPECTIVAS PARA EL ESTUDIO DE CEDRO ROJO (*Cedrela odorata* L.)]

Lorena Jacqueline Gómez-Godínez, Carlos Iván Cruz-Cárdenas, Edith Rojas-Anaya, Marco Aurelio Aragón-Magadan and Luis Felipe Guzmán*

Centro Nacional de Recursos Genéticos. Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. Boulevard de la Biodiversidad 400, Rancho las Cruces, C.P. 47600. Tepatitlán de Morelos, Jalisco, México. Email: guzman.luis@inifap.gob.mx Teléfono: 55-38-71-87-00 ext. 84823

**Corresponding author*

SUMMARY

Introduction: The high-throughput genomic and transcriptomic approach has been developed and implemented to address the main challenges that the timber forestry sector faces, such as population growth, climate change, deforestation and the loss of forest ecosystem services. **Objective:** To carry out a bibliographic review focused on the genomes and transcriptomes of timber trees reported in the databases, with special attention to red cedar (*Cedrela odorata* L.), due to its importance as precious wood in Mexico. **Methodology:** A literature review was carried out, directed at studying timber trees with genomic and transcriptomic strategies in different databases such as Pubmed, Scopus, Google Scholar, ScienceDirect, Wiley Online Library, MDPI and Scielo to identify the timber species that have been reported genomes and transcriptomes. The structure of the review was the genomics of timber trees, the transcriptomics of wood, and the potential species for study due to their importance and finally, the databases for consultation. Subsequently, a bibliometric study was carried out with the bibliometrix library in R Studio. **Main results:** The first genome to be assembled at the chromosome level was the black cottonwood. Among the timber trees, the genomes of black cottonwood, desert poplar, eucalyptus and oak with a length of 392, 496, 691 and 789 Mb have been reported. Through study of the transcriptome, it has been possible to identify genes related to formation of the wood in a hybrid poplar (*Populus alba* L. × *P. glandulosa*) and *P. tremula* L. and with drought tolerance in *Pinus massoniana* and *Pinus pinaster* Aiton. In red cedar (*Cedrela odorata* L.), the transcriptome was obtained by sequencing a single leaf, identifying 52,181 gene models. In the NCBI, EMBL-EBI, TreeGenes, PLAZA databases and the hardwood genomics website it is possible to find information related to the genomics and transcriptomics of timber species. **Implications:** More research is required in omics in timber, particularly in red cedar, since the search on these topics yielded little information. **Conclusion:** Through the bibliographic review in databases, the timber trees that have a described genome and transcriptome were identified. This information can be used for the assembly and annotation of new genomes, identification of genes and molecular markers, among other applications. **Key words:** *Cedrela odorata* L.; mass sequencing; genomics; transcriptomics; databases.

RESUMEN

Introducción: El enfoque genómico y transcriptómico de alto rendimiento se ha desarrollado e implementado para tratar los principales desafíos que el sector forestal maderable afronta como, el crecimiento poblacional, cambio climático, deforestación y la pérdida de los servicios ecosistémicos forestales. **Objetivo:** Realizar una revisión bibliográfica enfocada en los genomas y transcriptomas de árboles maderables reportados en las bases de datos, con especial atención en el cedro rojo (*Cedrela odorata* L.), debido a la importancia como madera preciosa en México. **Metodología:** Se realizó una revisión de literatura, dirigida en el estudio de los árboles maderables con estrategias

† Submitted January 8, 2023 – Accepted September 26, 2023. <http://doi.org/10.56369/tsaes.4773>



Copyright © the authors. Work licensed under a CC-BY 4.0 License. <https://creativecommons.org/licenses/by/4.0/>

ISSN: 1870-0462.

ORCID = Lorena Jacqueline Gómez-Godínez: <https://orcid.org/0000-0002-9759-7207>; **Carlos Iván Cruz-Cárdenas:** <https://orcid.org/0000-0002-1952-0191>; **Edith Rojas-Anaya:** <https://orcid.org/0000-0002-4553-3914>; **Marco Aurelio Aragón-Magadan:** <https://orcid.org/0000-0002-8633-3337>; **Luis Felipe Guzmán:** <https://orcid.org/0000-0002-1663-3457>

genómica y transcriptómica, en diferentes bases de datos como Pubmed, Scopus, Google Scholar, ScienceDirect, Wiley Online Library, MDPI y Scielo para identificar las especies maderables que cuentan con genomas y transcriptomas reportados. La estructura de la revisión fue: la genómica de árboles maderables, la transcriptómica de la madera, las especies potenciales de estudio por su importancia y finalmente, las bases de datos para consulta. Posteriormente se realizó un estudio bibliométrico con la librería de bibliometrix en R Studio. **Resultados principales:** El primer genoma forestal en ser ensamblado a nivel de cromosoma fue el álamo negro. Entre los árboles maderables, están reportados los genomas de álamo negro, álamo del desierto, eucalipto y roble con una longitud de 392, 496, 691 y 789 Mb. Con el estudio del transcriptoma, ha sido posible identificar genes relacionados con la formación de la madera en un álamo híbrido (*Populus alba* L. × *P. glandulosa*) y *P. tremula* L. y con la tolerancia a la sequía en *Pinus massoniana* y *Pinus pinaster* Aiton. En cedro rojo (*Cedrela odorata* L.), se obtuvo el transcriptoma con la secuenciación de una sola hoja, identificando 52,181 modelos de genes. En las bases de datos NCBI, EMBL-EBI, TreeGenes, PLAZA y el sitio web de la genómica de la madera dura es posible encontrar información relacionada con la genómica y transcriptómica de las especies maderables. **Implicaciones:** Se requieren más investigaciones en el área de las ómicas en maderables, particularmente en cedro rojo, ya que la búsqueda en estos temas arrojó poca información. **Conclusión:** A través de la revisión bibliográfica en las bases de datos, se identificaron los árboles maderables que cuentan con genoma y transcriptoma descritos. Dicha información puede ser utilizada para el ensamble y anotación de nuevos genomas, identificación de genes y marcadores moleculares, entre otras aplicaciones. **Palabras clave:** *Cedrela odorata* L.; secuenciación-masiva; genómica; transcriptómica; bases de datos.

INTRODUCCIÓN

Importancia de los árboles maderables

Las plantas comprenden más del 99% de materia viva en el planeta y son de gran importancia debido a que son los productores primarios, mantienen la composición, la calidad de la atmósfera, de los suelos y son los componentes principales de diversos sistemas ecológicos (Raven y Wackernagel, 2020). Sin las plantas, la mayoría de los organismos que habitan en la tierra actualmente no estarían presentes (Keddy, 2017; Simpson, 2010).

En el planeta existen de 375,000 a 450,000 especies de plantas distribuidas por todo el mundo (Christenhusz y Byng, 2016) (<http://www.theplantlist.org/>). Las espermatofitas representan un 79% de abundancia dentro del dominio de las plantas, convirtiéndose en el clado más diverso (Vargas *et al.*, 2012). Entre las espermatofitas, se encuentran las angiospermas (grupo de plantas que producen flores y semillas cerradas, dicotiledóneas, a menudo de hoja caduca, incluyendo roble, abedul, haya, fresno) y las gimnospermas (plantas que producen semillas descubiertas como el abeto, el pino, el abeto) las cuales son plantas formadoras de especies maderables (Ramage *et al.*, 2017).

Las plantas maderables, son especies que se cultivan con el propósito de obtener madera, presentan características peculiares, por ejemplo, tallos únicos o múltiples con crecimiento que supera dos metros de altura y la formación extensa de madera producida por un cámbium vascular o secundario, lo que genera una amplia variación en la anatomía de la madera (Groover y Cronk, 2017).

Los árboles maderables sustentan ecosistemas complejos, con participación en el secuestro de

carbono y producción de oxígeno, además, representan una fuente de energía renovable, funcionan como materia prima en la industria de productos forestales, en la producción de biomateriales y otros bioproductos (Ragauskas *et al.*, 2014; *State of the World's Forests 2016 - World ReliefWeb*, n.d.). En México, los árboles maderables que se comercializan son: Oyameles (*Abies* spp.), pinos (*Pinus* spp.), encinos (*Quercus* spp.) y las especies tropicales caoba (*Swietenia macrophylla* King), ceiba (*Ceiba pentandra* (L.) Gaertn.) y cedro rojo (*Cedrela odorata* L.), las cuales, se consideran maderas preciosas por el alto valor comercial (CONABIO, 2020).

Importancia del cedro rojo como recurso maderable

Entre los árboles maderables más preciados en México se encuentran especies del género *Cedrela* (Rendón y Neyra, 2020). El género *Cedrela*, es uno de los 53 géneros de árboles y arbustos de la familia Meliaceae, contiene 18 especies, las cuales se distribuyen desde el norte de México hasta el noroeste de Argentina.

La familia Meliaceae contiene especies con una gran calidad de madera, lo que hace que sean valorados y también explotados. Esto ha conducido a que las especies *Cedrela fissilis* Vell., *Cedrela angustifolia* DC. y *Cedrela odorata* L. sean consideradas vulnerables e incluidas a la lista roja de especies amenazadas de la Unión Internacional para la Conservación de la Naturaleza (UICN) (Gálvez López *et al.*, 2020).

En el caso del cedro rojo (*Cedrela odorata* L.) es una de las especies de árboles maderables que más demanda tiene en la industria forestal de México (Romo-Lozano *et al.*, 2017). Esta especie se usa para obtener madera aserrada y chapa de madera terciada para la fabricación de muebles finos e instrumentos

musicales, además, se utiliza en la elaboración de productos aromatizantes y medicinales (hoja, raíz, corteza, semilla y tallo) en infusiones para distintos padecimientos (Romo-Lozano *et al.*, 2017). Además, ofrece otro tipo de servicios no maderables como la captura de carbono, que se calcula ronda los 4.4 toneladas/ha/año (Mendizábal-Hernández *et al.*, 2011; Rueda Sánchez *et al.*, 2014).

El cedro es considerado una especie de madera preciosa según el anuario estadístico de producción forestal de la SEMARNAT (2018). En dicho informe se estima que la producción de maderas preciosas alcanzó los 15,931 m³/ha/año, distribuidos principalmente en los estados de Veracruz, Quintana Roo y Campeche. El precio fue de \$2,484.84 por m³, llegando incluso a alcanzar su punto máximo a \$6,107.00 por m³. Estos valores contrastan con los precios del pino, que es ampliamente utilizado en la carpintería y la producción de celulosa, el cual promedió los \$1,991.54, llegando a un máximo de \$2,100.00. Estos datos resaltan la importancia económica de la madera de cedro en comparación con otras especies. Por otro lado, según la publicación del INIFAP sobre Plantaciones forestales con cedro rojo en el trópico mexicano en 2023, se menciona que la producción óptima de este recurso forestal se alcanza entre los 20 y 25 años de edad, con un promedio de 15 m³/ha/año.

Tecnologías de secuenciación masiva para el estudio de árboles maderables

Las tecnologías de secuenciación masiva son técnicas de secuenciación de DNA de múltiples fragmentos en paralelo, en las cuales, se puede llevar a cabo la caracterización genotípica de miles de marcadores SNP's de varias muestras en una sola corrida. La secuenciación masiva ha revolucionado la forma en la que se estudia la biología y la genética de los organismos debido a que permiten obtener grandes cantidades de información lo cual, favorece al entendimiento de las características y funciones de los organismos (Uhlen y Quake, 2023).

El desarrollo de las tecnologías de secuenciación ha sido un proceso largo y complejo. A comienzos del siglo XXI, se desarrollaron métodos basados en la pirosecuenciación, la cual implica el uso de una hebra molde de DNA para sintetizar la cadena complementaria, detectada por una señal quimioluminiscente, la cual será detectada cada vez que se incorpore un nucleótido a la cadena de DNA complementario (Cummings *et al.*, 2013). Si bien esta tecnología fue utilizada por muchos años, actualmente está discontinuada debido a los errores en la secuenciación en regiones de homopolímeros y a los altos costos. Años más tarde surgieron tecnologías de mayor rendimiento estas plataformas fueron

denominadas *Next Generation Sequencing (NGS)*, actualmente es más apropiado hablar de *secuenciación masiva o High-Throughput Sequencing (HTS)*. Dentro de éstas encontramos a las tecnologías de secuenciación por síntesis como Illumina y Ion Torrent. El principio de estas tecnologías se basa en la fragmentación del DNA, en la secuenciación por Illumina los fragmentos de DNA se adhieren a una superficie sólida, estos se replican formando un clúster y posteriormente se incorporan nucleótidos marcados con fluorescencia. En la tecnología de Ion torrent los fragmentos de DNA se adhieren en microesferas las cuales son depositadas en una cámara individual, donde son añadidos uno a uno los nucleótidos, los cuales si son complementarios al DNA fragmentado se unirán generando una liberación del ión hidrógeno, dicha liberación será detectada y permitirá dilucidar el fragmento de DNA (Shendure *et al.*, 2017; Lahens *et al.*, 2017).

Tanto Illumina como IonTorrent son tecnologías de secuenciación de fragmentos cortos. En términos sencillos, en lugar de secuenciar fragmentos largos y completos de ADN como se hacía tradicionalmente, estas técnicas fragmentan el material genético en pequeñas piezas aleatorias que se secuencian de manera independiente. La técnica de shotgun es la más representativa dentro de este grupo en donde se obtiene una gran cantidad de secuencias que luego son ensambladas por herramientas bioinformáticas para reconstruir la secuencia genómica original (Regalado *et al.*, 2020; Weber y Myers, 1997).

Años más tarde, a inicios del siglo XXI surgió la tecnología de secuenciación de PacBio, donde se amplifica una única molécula de DNA en una pequeña esfera ("smrtbell"), posteriormente la esfera es colocada en una celda de secuenciación, donde se añadirán los nucleótidos marcados con fluorescencia, esto permite la observación de DNA en tiempo real (Rhoads y Au, 2015). En la secuenciación de Oxford Nanopore se amplifica DNA, este se hace pasar por un poro, en presencia de una corriente eléctrica. La secuencia de DNA se detecta mediante la medición de cambios de corriente eléctrica que sucede cada vez que el nucleótido pasa por el nanoporo. Esta tecnología es eficiente para la secuenciación de moléculas de DNA gran longitud (Kono y Arakawa, 2019).

La transcriptómica, por otro lado, utiliza técnicas de secuenciación de ARN, conocida como RNA-seq. Estas tecnologías permiten obtener una instantánea de la expresión génica en una muestra biológica en un momento específico. El proceso implica aislar y extraer el ARN total de la muestra, convertirlo en ADN complementario (cADN) y secuenciar los fragmentos de cADN generados (de las Heras-Saldana *et al.*, 2016; Gaete-Eastman *et al.*, 2022; Liu *et al.*, 2014).

Las tecnologías de secuenciación masiva han aumentado su disponibilidad y con esto se han disminuido costos, lo cual ha permitido un gran avance en la forma en la que se estudian diversas áreas, como por ejemplo los árboles maderables. Gracias a estas tecnologías se han logrado secuenciar genomas completos, para conocer los genes presentes en árboles maderables, lo cual ha permitido un entendimiento en su biología y genética, también se han identificado variaciones genéticas responsables de rasgos específicos de la madera como contenido de lignina y densidad, así como genes relacionados con la resistencia a enfermedades, calidad de madera, adaptación de maderables a diferentes climas, se ha podido evaluar también la respuesta de los maderables a diversos factores ambientales, lo cual está relacionado con la selección y mejora de maderables (Slatko *et al.*, 2018; Ouyng., *et al* 2016; Yaya Lancheros *et al.*, 2021; Yasodha *et al.*, 2018).

Estas tecnologías de secuenciación masiva han permitido secuenciar 236 genomas de angiospermas, la mayoría de ellas son plantas de importancia económica o sus parientes silvestres (Qin *et al.*, 2021). Además, se realizan estudios de asociación de todo el genoma (GWAS), mediante RNA-seq, en los cuales, analizan las interacciones entre árboles y ambiente, así como también se puede evidenciar la filogenómica, es decir la historia evolutiva de los árboles maderables a través del genoma (Kulski, 2016).

Debido a las tecnologías de secuenciación masiva es posible abordar cuestionamientos sobre los árboles maderables, por ejemplo, ¿Qué genes están involucrados en el crecimiento secundario, en la latencia estacional, la longevidad y en la adaptación ante diferentes tipos de estrés biótico y abiótico? (Giacomello y Morgante, 2012; Groover y Cronk, 2017; Parent *et al.*, 2015).

El enfoque genómico y transcriptómico significa incluir en los estudios, la información contenida en toda la secuencia de DNA de una especie y los transcritos expresados diferencialmente bajo una condición fisiológica determinada, respectivamente. Estos enfoques pueden utilizarse para acelerar la mejora de características de la madera, así como el crecimiento de las plantas y resistencia al estrés. Además, la predicción del rendimiento híbrido del genoma completo será muy importante para el diseño de la reproducción de árboles, ya que la heterosis se ha explotado con éxito para muchas especies forestales importantes. La selección y las cruces basadas en enfoques de genómica integrada podría utilizarse potencialmente para desarrollar cultivares nuevos y superiores (Du *et al.*, 2018).

El objetivo del presente trabajo fue realizar una revisión bibliográfica enfocada en los genomas y

transcriptomas de árboles maderables reportados en las bases de datos, con especial atención en el cedro rojo (*Cedrela odorata* L.), debido a la importancia como madera preciosa en México.

MATERIALES Y MÉTODOS

La revisión de la literatura fue dirigida para identificar las especies maderables que cuentan con genomas y transcriptomas reportados y la necesidad de realizar este tipo de estudios en el cedro rojo. Para ello se realizó un análisis bibliométrico y búsqueda de información en línea y en artículos, sobre los avances del estudio genómico y transcriptómico de especies maderables. Para el análisis bibliométrico, en una primera búsqueda secuencial se empleó el término "timber". Posteriormente, se añadieron los términos 1) "timber trees", 2) "genomics" y 3) "transcriptomics", con el uso del operador booleano AND. Esta búsqueda secuencial permitió combinar los términos claves y conocer la cantidad de estudios que se enfocan al estudio de la madera (1), árboles maderables (1 y 2) los árboles maderables con un enfoque genómico y (1 y 3) los árboles maderables con un enfoque transcriptómico. Finalmente se añadió el término clave "*Cedrela odorata*" a las últimas búsquedas (1 y 2) y (1 y 3), para encontrar los posibles estudios que incluyan al cedro mexicano con enfoques transcriptómicos o genómicos. Solo se tomaron en cuenta los resultados de publicaciones de 2010 a 2022. Para evaluar los trabajos realizados con cedro mexicano, se realizó una segunda búsqueda secuencial que incluyó primero el término "*Cedrela odorata*" y posteriormente se añadieron los términos "genomics" y "transcriptomics" empleando el operador booleano AND. Las referencias que se obtuvieron de las dos búsquedas secuenciales se depuraron. Las referencias duplicadas se eliminaron manualmente, a partir del archivo .xls obtenido de Scopus, donde se ordenaron por título de publicación y se utilizó la función de Excel de eliminación de duplicados. Las referencias depuradas se analizaron con el programa bibliometrix en R. Por otro lado, se analizaron los artículos científicos en que se describen los genomas de artículos en Pubmed, Scopus, Google Scholar, ScienceDirect, Wiley Online Library, MDPI y Scielo, b) Páginas web de la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), y de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) y c) las bases de datos EMBL-EBI, NCBI, PLAZA, Theplantlist, TreeGenes y el sitio web del proyecto de genómica de la madera dura (hardwoodgenomics). Esta información sirvió para describir algunos ejemplos de los estudios de genómica de árboles maderables, los avances que hay en la identificación de genes relacionados con la producción de madera a través de estudios de transcriptómica, las bases de datos que contienen información sintetizada sobre genómica y transcriptómica de especies maderable y

los pocos estudios y necesidad que se tiene para estudiar la especie *Cedrela odorata* con enfoques genómicos o transcriptómicos.

Las referencias duplicadas se eliminaron manualmente, a partir del archivo .xls obtenido de Scopus, donde se ordenaron por título de publicación y se utilizó la función de Excel de eliminación de duplicados. Por otro lado, se realizó una búsqueda más extensa en otras bases de datos y artículos científicos, que es parte de lo reportado en los diferentes apartados de resultados.

La revisión fue realizada en cuatro pasos: Primero, la genómica de árboles maderables con ejemplos de especies secuenciadas, segundo, la transcriptómica de la madera con la importancia de contar con un genoma de referencia, tercero, las especies potenciales de estudio por su importancia y finalmente, cuarto, las bases de datos disponibles para consulta.

RESULTADOS Y DISCUSIÓN

En la revisión de la literatura en Scopus se identificaron 18,548 publicaciones relacionadas con las especies maderables (utilizando término “timber”), encontrando que China es uno de los países con mayor número de publicaciones, México ocupa el lugar 13 en publicaciones relacionadas con árboles maderables (Figura 1), cuando se realizó la búsqueda con “timber AND trees”, Scopus arrojó 460 publicaciones. Por otra parte, se encontraron los términos “forestry”, “timber”, “forest management”, “trees”, “biodiversity” y “reforestation” (Figura 1), cuando la búsqueda se realizó de especies maderables más genómica y transcriptómica de árboles, el número de publicaciones disminuyó considerablemente a 67. La transcriptómica de las especies maderables está menos estudiada con 15 publicaciones en comparación con 58 publicaciones relacionadas con la genómica de los árboles maderables. En el caso del cedro rojo, cuando se realizó la búsqueda en Scopus con los términos “*Cedrela odorata* L AND transcriptomics AND genomics, databases and RNA-seq”, no se obtienen resultado, por lo cual se buscó en otras bases de datos. Cabe mencionar que *Cedrela odorata* L cuenta con transcriptoma a partir de una hoja, no obstante, aún no está reportado el genoma y esta información fue obtenida de como ya se mencionó de otras bases de datos.

Genómica de árboles maderables

En la mayoría de los proyectos de secuenciación genómica se realiza el análisis con un genoma de referencia, sin embargo, en los árboles maderables esto no había sido así, debido al costo de la secuenciación en el pasado y al tamaño de los genomas, los cuales comparados con algunas bacterias como *Haemophilus influenzae* (1.8 Kb) suelen ser grandes, hasta de 20 Gb como el de *Pinus sylvestris* (Iskander *et al.*, 2017).

La primera secuenciación de una especie maderable fue el álamo negro (*Populus trichocarpa* Torr. y A. Gray ex Hook.) (PRJNA10772) (Tuskan *et al.*, 2006), la cual, se llevó a cabo mediante la técnica shotgun. Este árbol fue elegido por el rápido crecimiento, relativa facilidad de manipulación experimental, por el uso en la generación de biocombustibles y principalmente, debido al tamaño pequeño del genoma. El genoma del álamo negro es el primero en ser ensamblado a nivel de cromosoma (Tuskan *et al.*, 2006). Actualmente, cuatro versiones diferentes del genoma de *Populus trichocarpa* están reportadas en NCBI (<https://www.ncbi.nlm.nih.gov/data-hub/genome/?taxon=3694>) no obstante, la versión *P.trichocarpa_v4.1*, la cual es la más reciente y considerada genoma de referencia, incluye 19 cromosomas, con una longitud de 392.2 Mb dispuestos en 46 andamios y 52,400 transcripciones que codifican proteínas.

El genoma del álamo del desierto (*Populus euphratica* Oliv.) (PRJNA178692) fue secuenciado en el año 2013 mediante la estrategia de agrupación de fósmidos (Ma *et al.*, 2013). Durante el ensamble se utilizaron 67.1 Gb de lecturas obtenidas por shotgun de genoma completo, en conjunto con 200 lecturas de alta calidad de 66,240 clones de fósmidos. El ensamble final produjo un genoma con una longitud total de 496.5 Mb lo cual, representa 83.7% del genoma de *P. euphratica*. Este árbol presenta alta tolerancia al estrés salino, además, al comparar con el genoma del álamo *P. trichocarpa*, se identificaron genes específicos involucrados en tolerancia al estrés, transporte de iones, actividad ATPasa y actividad oxidoreductasa (Ma *et al.*, 2013). La versión uno del genoma de *P. euphratica* tiene un tamaño de 495.9 Mb, dispuesto en 9,614 andamios, 32,788 contigs, 49,760 secuencias codificantes y 32% de GC. Además, el genoma del cloroplasto contiene 130 genes conocidos, incluidos 89 genes que codifican proteínas, ocho genes de ARN ribosomal y 37 genes de ARNt, de los cuales, 18 están ubicados en las regiones repetidas invertidas (Ma *et al.*, 2013; Zhang y Gao, 2014).

ensambló a partir de secuencias largas y cortas obtenidas con la tecnología 454 GS FLX (Roche Diagnostic) e Illumina (Bio Scientific). En la cuarta y más reciente versión del genoma del roble (GCF_932294415.1)

(https://www.ncbi.nlm.nih.gov/labs/data-hub/genome/GCF_932294415.1/) se describe un genoma con 12 cromosomas, 789.2 Mb de tamaño, dispuesto en 95 andamios, 214 contigs y 35.5% de GC. Un evento reciente de duplicaciones de genes locales y gran cantidad de genes relacionados con interacciones bióticas y resistencia a enfermedades fueron identificados (Plomion *et al.*, 2018).

La obtención de genomas de especies maderables ha representado un gran desafío, debido a cuatro razones: 1) la longitud de secuencias cortas obtenidas mediante la plataforma Illumina y las familias de genes con secuencias muy similares representan un problema durante el ensamble, 2) la heterocigosidad, 3) las duplicaciones segmentarias y en tándem dentro del genoma y 4) las poliploidías (Tuskan *et al.*, 2006). Con otras estrategias, se ha intentado producir bibliotecas de secuenciación a partir de tejido haploide (1n) o mediante el uso de fósmidos para poder subsanar el problema de las secuencias cortas y ensamblar de forma independiente y fusionar longitudes de lecturas más largas (Alexeyenko *et al.*, 2014; Neale *et al.*, 2017). En el caso de los genomas de árboles maderables disponibles en la base de datos del NCBI, hay un alto grado de fragmentación, representados por dos o más contigs ensamblados. Esto significa grandes desafíos para la utilización posterior de la anotación del genoma correspondiente.

En resumen, la genómica enfocada a árboles maderables va avanzando a paso lento, hasta el momento son pocas especies las que han sido secuenciadas con éxito y en su mayoría son especies que tiene todas las condiciones para ser planta modelo para la genómica de árboles debido a su pequeño tamaño de genoma, rápido crecimiento, facilidad de propagación clonal, regeneración de alto rendimiento y disponibilidad de extensos mapas genéticos. Comprender el genoma de plantas modelo ayuda en gran medida a nuestra comprensión de los procesos de crecimiento y desarrollo en otras especies. Como ejemplo tomamos a *Populus trichocarpa*, la interacción de su genoma ha permitido identificar más de 45,000 genes codificadores, genes duplicados, sustitución de nucleótidos y el reordenamiento cromosómico. Esta información sobrerrepresentada en *Populus* sirve de base para conocer genes asociados con la resistencia a enfermedades, el desarrollo de meristemas, el transporte de metabolitos y la biosíntesis de la pared lignocelulósica en las demás especies del género *Populus* (Tuskan *et al.*, 2006).

Los avances en genómica de árboles maderables muestran que esta tecnología podría mejorar la percepción del valor genético estimado por la importancia de estas especies y dar avance a la revolución genómica. Con la promesa de mayor información de las especies para la toma de decisiones, además del beneficio económico para la industria forestal. La genómica como herramienta en especies forestales ayuda a mejorar rápidamente el conocimiento y desarrollar estrategias para los tomadores de decisiones en gestión y conservación forestal, y de esta manera fortalecer los programas de reproducción, crear mecanismos para enfrentar los desafíos del cambio climático, incluso en especies que no son modelo y no han sido secuenciadas.

Transcriptómica de árboles maderables

Los organismos vivos no siempre tienen la misma expresión de genes, esta expresión está determinada por factores bióticos y abióticos. Con las tecnologías de secuenciación y los experimentos basados en transcriptómica es posible analizar la expresión de genes en diversas condiciones (Kulski, 2016; Lowe *et al.*, 2017).

En estudios de transcriptómica, contar con genoma de referencia es un factor relevante a considerar. El análisis de transcriptoma basado en genoma de referencia conlleva revisión, limpieza y control de calidad de las secuencias obtenidas por secuenciación, además, el alineamiento permite la identificación de empalme y la detección de intrones (Voshall *et al.*, 2021).

Una vez realizado el mapeo, se cuantifican las secuencias alineadas con cada *locus* o contig, lo cual queda plasmado en una matriz de cuentas y posteriormente se analiza a la expresión diferencial de genes. Esta cuantificación de expresión diferencial a nivel de transcripción sigue siendo un desafío significativo (Blumenberg, 2019; Dries *et al.*, 2021; Lowe *et al.*, 2017).

Por otra parte, en ausencia de genoma de referencia, se realiza ensamble *de novo* a partir de los resultados obtenidos en la secuenciación. En este paso existe un gran desafío ya que, en caso de tener secuencias cortas, hay más probabilidad de resultar en un ensamble incorrecto o redundante (*Best Practices for De Novo Transcriptome Assembly with Trinity - Harvard FAS Informatics*, n.d.; Hölzer y Marz, 2019).

El ensamble *de novo* a partir de datos de RNA-Seq es una alternativa para obtener un genoma de referencia, con el inconveniente de ser ineficiente y poco óptimo durante la cuantificación (Voshall *et al.*, 2021). Una opción es la utilización de tecnologías de secuenciación de lecturas largas, como PacBio. Con

esta tecnología se facilita realizar ensamblaje y, por lo tanto, reducción del error y la redundancia en el mapeo (Amarasinghe *et al.*, 2020; Wang *et al.*, 2021).

Finalmente, cuando se obtiene un genoma de referencia ensamblado *de novo*, se evalúan las secuencias, de la misma manera que con el enfoque utilizando un genoma de referencia previamente reportado. Adicionalmente, se lleva a cabo control de calidad para eliminar secuencias de baja calidad y adaptadores, mapeo o comparación con el genoma de referencia, obtención de la matriz de cuenta y análisis de expresión diferencial (Wang *et al.*, 2021).

A través del uso de la transcriptómica, se ha estudiado el proceso de formación de la madera, pero, los mecanismos moleculares para la formación de madera no han sido completamente dilucidados. Por esta razón Kim *et al.* (2019) estudiaron la formación del cámbium vascular del álamo híbrido (*Populus alba* L. × *P. glandulosa* Torr.) mediante RNA-Seq y lograron identificar genes responsables de la biosíntesis de celulosa y lignina y genes involucrados en la formación del cámbium vascular (Kim *et al.*, 2019).

Sundell *et al.* (2017) identificaron genes involucrados en la biosíntesis de xilano y celulosa y reguladores de la diferenciación y lignificación de xilema durante la evaluación de la formación de madera en diferentes tejidos de *Populus tremula* L. (Sundell *et al.*, 2017).

Por otro lado, el efecto del estrés biótico y abiótico en los árboles maderables se han estudiado mediante RNA seq. La sequía es un estrés abiótico que tiene efectos devastadores en la productividad de las plantas y árboles maderables. Las plantas tienen la capacidad de responder y contender ante los diferentes tipos de estrés, regulando mecanismos fisiológicos y moleculares.

En *Pinus massoniana* Lamb. expuesto a diferentes niveles de sequía, se identificaron genes relacionados con el ajuste osmótico, respuesta al ácido abscísico, y eliminación de especies reactivas de oxígeno (Du *et al.*, 2018).

Pinus pinaster Aiton es una de las principales fuentes de madera blanda en Europa y se ha encontrado que tiene un rápido ajuste osmótico en respuesta al déficit hídrico (de María *et al.*, 2020). La transcriptómica se utilizó para identificar genes relacionados con esta contención ante el estrés, encontrando que los genes involucrados con la biosíntesis de pantotenato, CoA y alcaloides de isoquinolina y con la degradación del estireno están sobre expresados, mientras que los genes implicados en el crecimiento, la traducción y la actividad de las moléculas estructurales estaban regulados negativamente (de María *et al.*, 2020).

En resumen, el transcriptoma de árboles maderables se ha abordado con tres principales aspectos: 1) el mecanismo molecular de la formación de madera con varios tipos de tejido en el álamo híbrido (*Populus alba* L. × *P. glandulosa*) (Kim *et al.*, 2019) y *Populus tremula* L. (Sundell *et al.*, 2017), 2) tolerancia a la sequía en el pino de Masson (*Pinus massoniana* Lamb.) (Du *et al.*, 2018) y en la conífera mediterránea *Pinus pinaster* Ait. (de María *et al.*, 2020) y, 3) la identificación de marcadores moleculares (*Cedrela balansae* C. DC.) por ejemplo, microsátélites para estudios de diversidad genética con fines de conservación (Torales *et al.*, 2018). Además, el avance en el estudio del transcriptoma puede ser dirigido en interacciones planta-patógeno y en respuestas a diversas condiciones ambientales como estrés calórico e hídrico o a la exposición a sustancias (Soto y López *et al.*, 2012).

***Cedrela odorata* L.: Una visión desde la genómica y transcriptómica**

Entre los árboles maderables más preciados en México se encuentran especies del género *Cedrela* (Rendón y Neyra, 2020). El género *Cedrela*, es uno de los 53 géneros de árboles y arbustos de la familia Meliaceae, contiene 18 especies, las cuales se distribuyen desde el norte de México hasta el noroeste de Argentina.

La familia Meliaceae contiene especies con una gran calidad de madera, lo que hace que sean valorados y también explotados. Esto ha conducido a que las especies *Cedrela fissilis* Vell., *Cedrela angustifolia* DC. y *Cedrela odorata* L. sean consideradas vulnerables e incluidas a la lista roja de especies amenazadas de la Unión Internacional para la Conservación de la Naturaleza (UICN) (Gálvez López *et al.*, 2020).

En el caso del cedro rojo (*Cedrela odorata* L.) es una de las especies de árboles maderables que más demanda tiene en la industria forestal de México (Romo-Lozano *et al.*, 2017). Esta especie se usa para obtener madera aserrada y chapa de madera terciada para la fabricación de muebles finos e instrumentos musicales, además, se utiliza en la elaboración de productos aromatizantes y medicinales (hoja, raíz, corteza, semilla y tallo) en infusiones para distintos padecimientos (Romo-Lozano *et al.*, 2017).

En el año 1992, se publicó el primer artículo científico que discute la utilización de *C. odorata* L., para la reforestación de un bosque tropical en México y a partir de este momento, existe un alza en publicaciones y reportes encaminados al estudio *C. odorata* L. reportándose 76 artículos en el año 2022 (Figura 2). En la búsqueda de términos relacionados con las diferentes publicaciones de *C. odorata* L. se

obtuvieron algunos términos relacionados con “meliaceae”, “biodiversity”, “plant extract” (Figura 2).

encontraron tres reportes, en los cuales los términos más mencionados fueron “meliaceae”, “phylogeny”, “genomics” (Figura 3).

Al realizar una búsqueda en Scopus con los términos de *C. odorata* L. más genómica y transcriptómica se

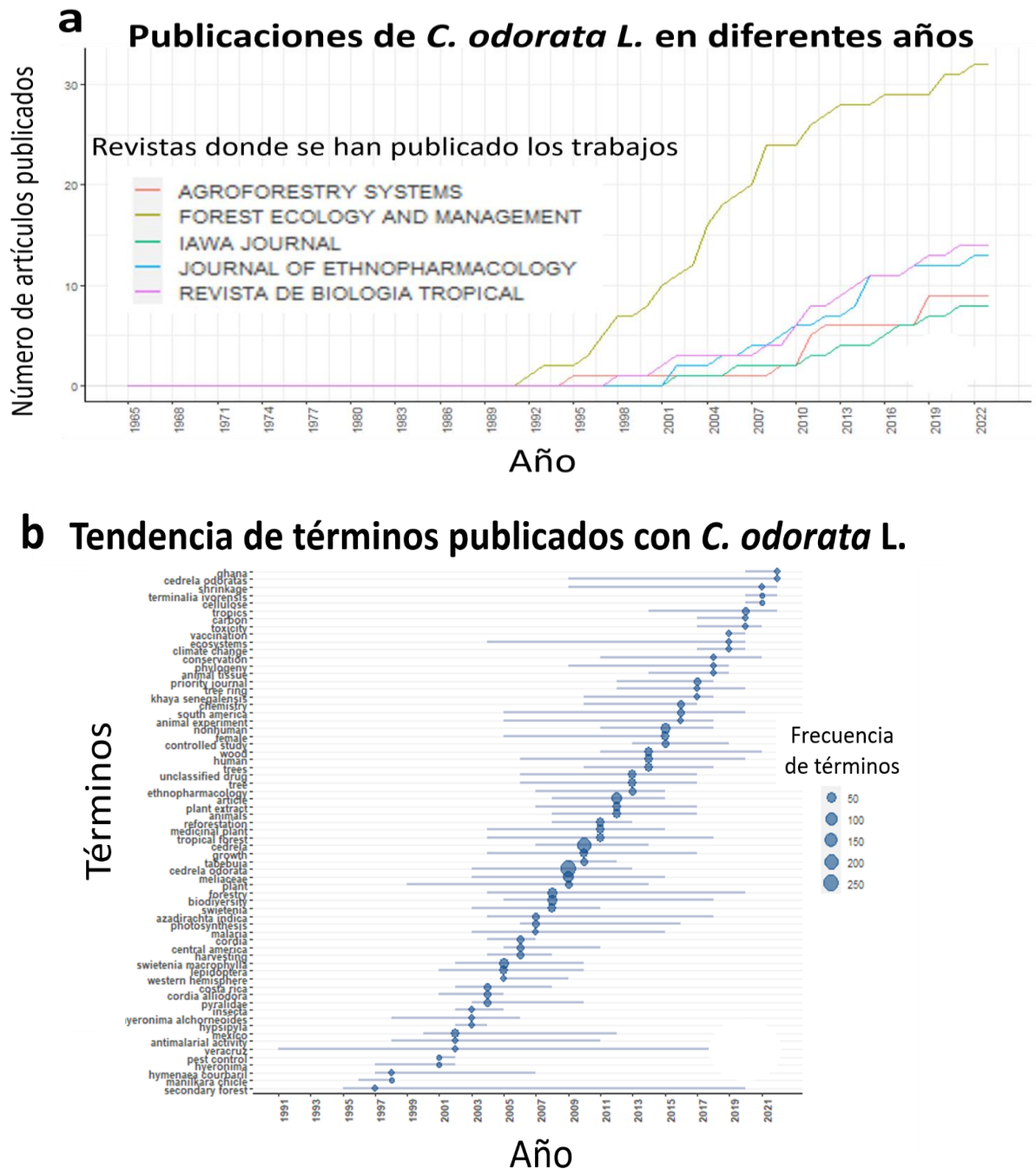


Figura 2. Investigaciones de *C. odorata* L. a) Cantidad de publicaciones de *C. odorata* L. en diferentes años. b) Tendencia de los términos más frecuentes en los diferentes trabajos publicados de *C. odorata* L. c) Red de colaboración mundial trabajando con *C. odorata* L.

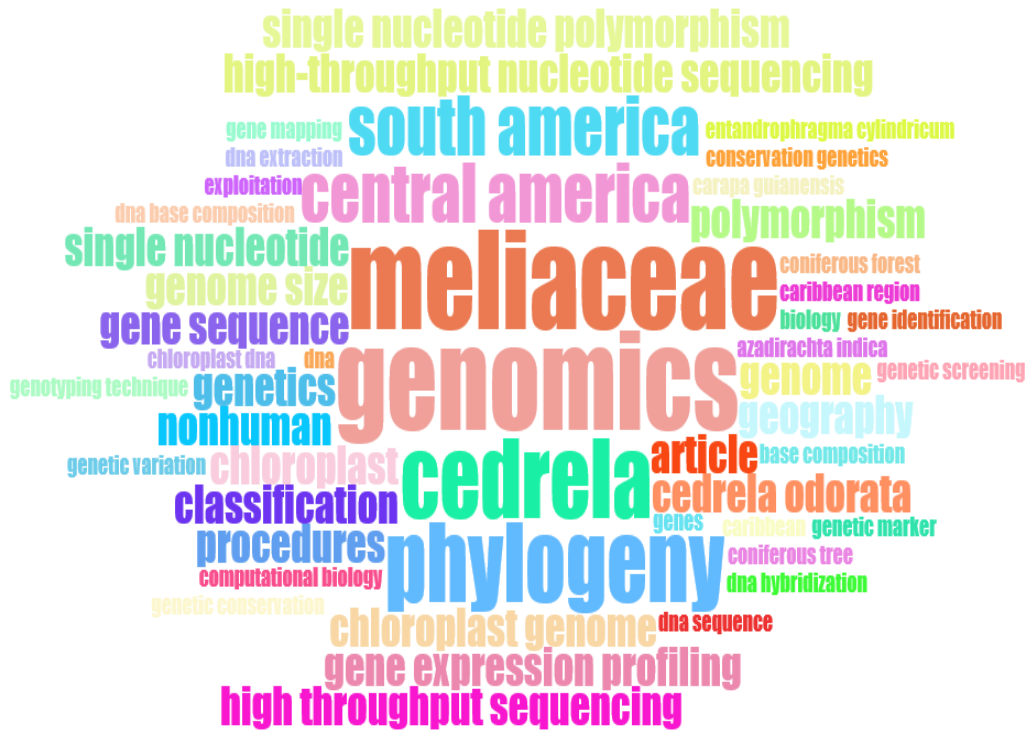


Figura 3. Nube de palabras con las principales palabras clave en investigaciones de *C. odorata* L. relacionadas con genómicas y transcriptómica

El cedro rojo es de suma importancia como maderable, sin embargo, el genoma nuclear no está reportado. Por otro lado, el genoma del cloroplasto de *C. odorata* L. tiene un tamaño de 158,558 pb (Accesión MG724915), en el que las regiones de copia sencilla larga (*Long Single Copy*) y corta (*Short Single Copy*) miden 86,390 pb y 18,380 pb, respectivamente, mientras que, la región invertida (*Inverted Region*) mide 26,894 pb. La lista de genes anotados del plastoma del cedro rojo contiene 112 genes, de los cuales, 78 codifican para proteínas, 30 para ARNt y cuatro para ARNr (Mader *et al.*, 2018).

En relación con el transcriptoma Finch, Jones y Cronn (2019) secuenciaron una hoja de *C. odorata* L. de Oaxaca (México), obtenida del Jardín Botánico de Nueva York (CEOD-NYBG) mediante la plataforma Illumina HiSeq 3000. Con el análisis se identificaron 52,181 modelos de genes con un rango de longitud de 200 a 37,635 pb, de los cuales, entre 65% a 70% presentan homología con los genomas de los árboles *Populus trichocarpa* y *Quercus lobata* Née (Finch *et al.*, 2019).

A pesar de la importancia del cedro rojo por el uso y explotación, los reportes del genoma, transcriptoma y adaptación a diferentes condiciones es escaso, en

comparación con las especies forestales de los géneros *Populus* y *Pinus*. Por lo tanto, la realización de estudios en el cedro rojo que revelen la composición del genoma nuclear e identificación de genes relacionados con la adaptación a diferentes ambientes o tipos de estrés biótico y abiótico es altamente recomendable.

Bases de datos ómicas en árboles maderables

Actualmente se cuenta con mayor acceso a la información y una gran cantidad de recursos en línea, que permiten explotar y analizar datos genómicos y transcriptómicos de plantas, específicamente de árboles maderables. Algunas bases de datos globales son NCBI (National Center for Biotechnology Information) (<https://www.ncbi.nlm.nih.gov/>) y EMBL-EBI (European Molecular Biology Laboratory - European Bioinformatics Institute) (<https://www.ebi.ac.uk/>), donde se encuentra información general sobre los maderables, desde libros, artículos, proyectos de secuenciación, genomas y transcriptomas.

Estas bases de datos ofrecen herramientas y recursos, como GenBank y ENA (European Nucleotide Archive), que contienen secuencias de ácidos

nucleicos de genomas y genes de una amplia gama de organismos, incluidos los árboles maderables. La interacción y complementariedad entre NCBI y EMBL-EBI son esenciales para el avance en la investigación genómica y la aplicación de conocimientos en recursos forestales. El intercambio de datos y colaboración entre estas instituciones permite la actualización constante y el enriquecimiento de sus repositorios, así como la identificación de homologías y la comparación de secuencias entre especies mediante herramientas como BLAST e InterPro.

La importancia de NCBI y EMBL-EBI en el contexto de árboles maderables radica en su capacidad para proporcionar un marco sólido de información y recursos que facilitan el estudio de la diversidad genética, la adaptación, la resistencia a enfermedades y la mejora de la productividad en árboles forestales. Además, al ofrecer acceso abierto y gratuito a sus recursos, estas instituciones fomentan la colaboración científica y el intercambio de conocimientos entre investigadores de todo el mundo, promoviendo el avance en la comprensión y conservación de árboles maderables y sus ecosistemas.

También existen otras bases de datos con integración únicamente de información de diferentes especies de árboles. Los árboles forestales están clasificados en 16 órdenes y contienen más de 1,790 especies. TreeGenes (<https://treegenesdb.org/>) es una base de datos genómicos y genéticos de árboles forestales que proporciona herramientas para la gestión de información obtenida por secuenciación, flujo de datos automatizados y visualización de toda la información (Falk *et al.*, 2018).

El proyecto TreeGenes es especialmente relevante para investigadores que buscan información sobre poblaciones naturales y mejoradas de árboles maderables, ya que incluye datos de genotipado por secuenciación (GBS), secuenciación de ARN (RNA-Seq), marcadores moleculares y genotipos de referencia. Además, la base de datos también ofrece información sobre la diversidad genética, la estructura de las poblaciones y la localización geográfica de las muestras, lo que puede ser útil para estudiar la adaptación y resistencia de las especies forestales a diferentes condiciones ambientales y climáticas. TreeGenes también proporciona herramientas de análisis bioinformático, como búsqueda de homologías y anotación funcional, lo que permite a los investigadores identificar genes de interés, estudiar la conservación de secuencias y funciones entre diferentes especies y examinar las vías metabólicas y de señalización involucradas en procesos biológicos clave.

Una característica importante de TreeGenes es su enfoque en la colaboración y el intercambio de datos, lo que permite a los usuarios contribuir con sus propios datos y aprovechar los conocimientos generados por otros investigadores en la comunidad científica. Además, la plataforma ofrece la posibilidad de integrar datos de diferentes fuentes, como bases de datos externas y proyectos de investigación colaborativos, lo que amplía aún más su utilidad para la investigación genómica y genética de árboles forestales.

Además, en la base de datos PLAZA (<https://bioinformatics.psb.ugent.be/plaza/>) están contenidos datos genómicos comparativos de las plantas a través de una interfaz web de fácil uso (Van Bel *et al.*, 2018). Mediante la plataforma es posible transferir el conocimiento de organismos modelo a organismos no modelo, conocer la evolución de genes específicos, vías metabólicas y de señalización completas, detectar genes de importancia para la adaptación de plantas específicas de nicho e identificar eventos genómicos como duplicaciones del genoma completo (Proost *et al.*, 2015).

Más allá de las funciones mencionadas, PLAZA proporciona herramientas de análisis funcional y estructural de genes, lo que permite a los investigadores estudiar familias, identificar ortólogos y parálogos y examinar la conservación de secuencias y estructuras de proteínas en diferentes especies. La plataforma también incluye datos sobre microARNs, lo que puede contribuir al estudio de la regulación génica en árboles maderables.

Uno de los aspectos destacados de PLAZA es su capacidad para analizar vías metabólicas y de señalización completas, lo que permite identificar genes clave y procesos biológicos involucrados en la adaptación y supervivencia de especies de nicho. Estos conocimientos pueden ser útiles en la conservación y mejora de árboles maderables, así como en la identificación de nuevos objetivos para el mejoramiento genético y la biotecnología forestal.

El sitio web del proyecto de la genómica de la madera dura (<https://hardwoodgenomics.org/>) proporciona herramientas bioinformáticas para la navegación, búsqueda, visualización y análisis de datos genómicos, además, contiene enlaces a los datos del genoma y del transcriptoma de varias especies de árboles de madera dura.

Además de las herramientas de navegación, búsqueda, visualización y análisis de datos genómicos, el sitio web del proyecto de genómica de la madera dura ofrece acceso a datos experimentales, como perfiles de expresión génica, información sobre mutantes y fenotipos asociados, y datos de interacción proteína-proteína y proteína-DNA. Estos recursos permiten a

los investigadores explorar la función y regulación de genes específicos y estudiar su relación con características de interés, como resistencia a enfermedades, calidad de la madera y adaptación a diferentes condiciones ambientales. También, proporciona tutoriales y recursos educativos para la capacitación y el desarrollo de habilidades en bioinformática y genómica, lo que facilita el acceso y uso de la plataforma por parte de investigadores y estudiantes con diferentes niveles de experiencia.

La plataforma promueve la colaboración y el intercambio de datos entre investigadores al permitir la contribución por parte de usuarios, así como la integración con otras bases de datos y recursos en línea relevantes para la investigación. Esto incluye enlaces a bases de datos como NCBI, EMBL-EBI, PLAZA y TreeGenes, entre otros, lo que facilita el acceso a una amplia gama de información genómica y transcriptómica de múltiples especies de árboles de madera dura.

El almacenamiento de información en bases de datos alcanzado un desarrollo considerable, no obstante, se presenta el desafío de la implementación de fácil procesamiento, captura automatizada de metadatos, asignación generalizada de términos genéticos, fenotípicos y ambientales e inclusión de la información del microbioma y metatranscriptoma. Además, es necesario invertir más recursos y esfuerzos en acrecentar la representación de árboles en las bases de datos existentes. Por ejemplo, de los 1,863,191 genomas de orgánulos secuenciados para plantas reportados a marzo de 2023 en el NCBI, solo 45,518 pertenecen a árboles, apenas un 2.4 % del total. Esta discrepancia plantea preguntas sobre las posibles causas y la necesidad de aumentar la investigación en la genómica de árboles.

Una posible explicación para esta subrepresentación es la priorización histórica de organismos modelo y plantas de importancia agrícola o económica. Muchos esfuerzos de secuenciación se han centrado en plantas como *Arabidopsis thaliana*, *Oryza sativa* (arroz) y *Zea mays* (maíz), dejando a un lado la diversidad de especies arbóreas que desempeñan roles ecológicos cruciales. Aunado a esto, la secuenciación de genomas de orgánulos en árboles presenta desafíos técnicos y logísticos adicionales.

Los árboles tienen ciclos de vida más largos y tasas de crecimiento más lentas en comparación con otras plantas, lo que dificulta la obtención de material biológico y la realización de experimentos. También es probable que la complejidad genética y la diversidad de especies arbóreas hayan desalentado a los investigadores de abordar estos genomas en comparación con organismos más simples y de rápido crecimiento. A pesar de estos desafíos, es fundamental

aumentar la cantidad de secuenciaciones de orgánulos en árboles. Estos desempeñan un papel crucial en los ecosistemas, la captura de carbono y la mitigación del cambio climático. La información genómica puede ayudar a comprender mejor la adaptación, la resistencia a enfermedades y el cambio ambiental, así como a informar sobre estrategias de conservación y gestión de recursos forestales.

CONCLUSIONES

A pesar del reto para la obtención de genomas de especies maderables, se cuenta con los genomas de álamo, eucalipto y roble, los cuales, pueden servir como referencia para nuevos estudios. También, obtener genoma de referencia a partir de datos de RNA-Seq es una alternativa. No obstante, en ausencia de genoma de referencia es recomendable el empleo de tecnologías de secuenciación de lectura larga para evitar el ensamble incorrecto. En los estudios de transcriptómica, es posible la identificación de genes con potencial inclusión en programas de mejoramiento genético relacionados con la tolerancia a estrés biótico y abiótico.

Agradecimientos

Los autores agradecemos al proyecto de investigación titulado “Identificación de genes diferencialmente regulados en plantas de cedro (*Cedrela odorata* L.) con No. 947834844, financiado con recursos fiscales del Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias.

Funding. Fiscal resources of the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, financed this research. Number Project 947834844

Competing interests. The authors declare no competing interests.

Compliance with ethical standards. This work does not require approval by a bioethical committee.

Data availability. Data is available with the corresponding author upon reasonable request.

Author contribution statement (CRediT).

LJ Gómez-Godínez: Writing – original draft and Visualization, Writing – review y editing, **CI Cruz-Cárdenas:** Writing – review y editing, **E Rojas-Anaya:** Writing – review and editing, **MA Aragón-Magadan:** Writing – review y editing, **LF Guzmán:** Writing – review y editing.

REFERENCIAS

Alexeyenko, A., Nystedt, B., Vezzi, F., Sherwood, E., Ye, R., Knudsen, B., Simonsen, M., Turner, B., de Jong, P., Wu, C. C. and Lundéberg, J.,

2014. Efficient de novo assembly of large and complex genomes by massively parallel sequencing of Fosmid pools. *BMC Genomics*, 15(1), p. 439. <https://doi.org/10.1186/1471-2164-15-439>
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E. and Gouil, Q., 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, 21(1), pp. 1–16. <https://doi.org/10.1186/S13059-020-1935-5>
- Best Practices for De Novo Transcriptome Assembly with Trinity - Harvard FAS Informatics*. (n.d.). Retrieved March 2, 2022, from <https://informatics.fas.harvard.edu/best-practices-for-de-novo-transcriptome-assembly-with-trinity.html>
- Blumenberg, M., 2019. Introductory Chapter: Transcriptome Analysis. *Transcriptome Analysis*. <https://doi.org/10.5772/INTECHOPEN.85980>
- Christenhusz, M. J. M. and Byng, J. W., 2016. The number of known plant species in the world and its annual increase. *Phytotaxa*, 261(3), pp. 201–217. <https://doi.org/10.11646/PHYTOTAXA.261.3.1>
- CONABIO. 2020. Maderable Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Cd. de México. México. Contenido: Alejandro Rendón y Lucila Neyra. <https://www.biodiversidad.gob.mx/diversidad/maderable>.
- Cummings, P. J., Ahmed, R., Durocher, J. A., Jessen, A., Vardi, T. and Obom, K. M., 2013. Pyrosequencing for Microbial Identification and Characterization. *Journal of Visualized Experiments*, 78, p. e50405. <https://doi.org/10.3791/50405>
- de las Heras-Saldana, S., Al-Mamun, H. A., Ferdosi, M. H., Khansefid, M. and Gondro, C., 2016. RNA Sequencing Applied to Livestock Production. En H. N. Kadarmideen (Ed.), *Systems Biology in Animal Production and Health*, Vol. 1 (pp. 63–94). Springer International Publishing. https://doi.org/10.1007/978-3-319-43335-6_4
- de María, N., Guevara, M. Á., Perdiguero, P., Vélez, M. D., Cabezas, J. A., López-Hinojosa, M., Li, Z., Díaz, L. M., Pizarro, A., Mancha, J. A., Sterck, L., Sánchez-Gómez, D., Miguel, C., Collada, C., Díaz-Sala, M. C. and Cervera, M. T., 2020. Molecular study of drought response in the Mediterranean conifer *Pinus pinaster* Ait.: Differential transcriptomic profiling reveals constitutive water deficit-independent drought tolerance mechanisms. *Ecology and Evolution*, 10(18), pp. 9788–9807. <https://doi.org/10.1002/ECE3.6613>
- Dries, R., Chen, J., Del Rossi, N., Khan, M. M., Sistig, A. and Yuan, G. C., 2021. Advances in spatial transcriptomic data analysis. *Genome Research*, 31(10), p. 1706. <https://doi.org/10.1101/GR.275224.121>
- Du, M., Ding, G. and Cai, Q., 2018. The Transcriptomic Responses of *Pinus massoniana* to Drought Stress. *Forests*, 9, p. 326. <https://doi.org/10.3390/F9060326>
- Du, Q., Lu, W., Quan, M., Xiao, L., Song, F., Li, P., Zhou, D., Xie, J., Wang, L. and Zhang, D., 2018. Genome-wide association studies to improve wood properties: Challenges and prospects. *Frontiers in Plant Science*, 871, p. 1912. <https://doi.org/10.3389/FPLS.2018.01912>
- Falk, T., Herndon, N., Grau, E., Buehler, S., Richter, P., Zaman, S., Baker, E. M., Ramnath, R., Ficklin, S., Staton, M., Feltus, F. A., Jung, S., Main, D. and Wegrzyn, J. L., 2018. Growing and cultivating the forest genomics database, TreeGenes. *Database: The Journal of Biological Databases and Curation*, 2018, pp. bay084. <https://doi.org/10.1093/DATABASE/BAY084>
- Finch, K. N., Jones, F. A. and Cronn, R. C., 2019. Genomic resources for the Neotropical tree genus *Cedrela* (Meliaceae) and its relatives. *BMC Genomics*, 20(1), p. 58. <https://doi.org/10.1186/S12864-018-5382-6>
- Gaete-Eastman, C., Stappung, Y., Molinett, S., Urbina, D., Moya-Leon, M. A. and Herrera, R., 2022. RNAseq, transcriptome analysis and identification of DEGs involved in development and ripening of *Fragaria chiloensis* fruit. *Frontiers in Plant Science*, 13, p. 976901. <http://doi.org/10.3389/fpls.2022.976901>
- Gálvez López, L., Vallejo Reyna, M. Á., Méndez Espinoza, C., López Upton, J., Gálvez López, L., Vallejo Reyna, M. Á., Méndez Espinoza,

- C. and López Upton, J., 2020., *Cedrela odorata* L.: oportunidades para su conservación y mejoramiento genético. *Revista Mexicana de Ciencias Forestales*, 11(58), pp. 4–25. <https://doi.org/10.29298/RMCF.V11I58.622>
- Giacomello, S. and Morgante, M., 2012. *Next-generation sequencing in Populus nigra: de novo assembly, genome-wide SNP map and comparative genomic analysis Candidate*. Tesis doctorado. Università degli Studi di Udine, Italia. https://air.uniud.it/retrieve/e27ce0c5-7534-055e-e053-6605fe0a7873/10990_41_Stefania_Giacomello_tesi_dottorato_pdfa.pdf
- Groover, A. and Cronk, Q., 2017., *Comparative and Evolutionary Genomics of Angiosperm Trees*. Springer Cham. <https://doi.org/10.1007/978-3-319-49329-9>
- Hölzer, M. and Marz, M., 2019. De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *GigaScience*, 8(5), p. giz039.. <https://doi.org/10.1093/GIGASCIENCE/GIZ039>
- Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, 2023. *Plantaciones forestales con cedro rojo en el trópico mexicano*. <http://www.gob.mx/inifap/articulos/plantaciones-forestales-con-cedro-rojo-en-el-tropico-mexicano>
- Iskander, M., Hayden, K., Van Domselaar, G. and Tsang, R., 2017. First Complete Genome Sequence of *Haemophilus influenzae* Serotype a. *Genome Announcements*, 5(3), p. e01506-16. <https://doi.org/10.1128/GENOMEA.01506-16>
- Keddy, P. A. 2017. *Plant Ecology: Origins, Processes, Consequences*. <https://doi.org/10.1017/9781316321270>
- Kim, M. H., Cho, J. S., Jeon, H. W., Sangsawang, K., Shim, D., Choi, Y. I., Park, E. J., Lee, H. and Ko, J. H., 2019. Wood Transcriptome Profiling Identifies Critical Pathway Genes of Secondary Wall Biosynthesis and Novel Regulators for Vascular Cambium Development in *Populus*. *Genes*, 10(9), p. 690. <https://doi.org/10.3390/GENES10090690>
- Kono, N. and Arakawa, K., 2019. Nanopore sequencing: Review of potential applications in functional genomics. *Development, Growth and Differentiation*, 61(5), pp. 316–326. <https://doi.org/10.1111/dgd.12608>
- Kulski, J. K. 2016. Next-Generation Sequencing — An Overview of the History, Tools, and “Omic” Applications. In: *Next Generation Sequencing*. J.K. Kulski, ed. InTechOpen. <https://doi.org/10.5772/61964>
- Lahens, N. F., Ricciotti, E., Smirnova, O., Toorens, E., Kim, E. J., Baruzzo, G. and Grant, G. R., 2017. A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. *BMC Genomics*, 18, p. 602. <https://doi.org/10.1186/s12864-017-4011-0>
- Liu, Y., Zhou, J. and White, K.P., 2014. RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics*, 30(3), p. 301–304. <https://doi.org/10.1093/BIOINFORMATICS/BTT688>
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S. and Shafee, T. 2017. Transcriptomics technologies. *PLOS Computational Biology*, 13(5), p. e1005457. <https://doi.org/10.1371/JOURNAL.PCBI.1005457>
- Mader, M., Pakull, B., Blanc-Jolivet, C., Paulini-Drewes, M., Bouda, Z. H. N., Degen, B., Small, I. and Kersten, B. 2018. Complete Chloroplast Genome Sequences of Four Meliaceae Species and Comparative Analyses. *International Journal of Molecular Sciences*, 19(3), p. 701. <https://doi.org/10.3390/IJMS19030701>
- Ma, T., Wang, J., Zhou, G., Yue, Z., Hu, Q., Chen, Y., Liu, B., Qiu, Q., Wang, Z., Zhang, J., Wang, K., Jiang, D., Gou, C., Yu, L., Zhan, D., Zhou, R., Luo, W., Ma, H. and Jianquan, L. 2013. Genomic insights into salt adaptation in a desert poplar. *Nature Communications*, 4(1), pp. 1–9. <https://doi.org/10.1038/ncomms3797>
- Maderable: Biodiversidad Mexicana*. (n.d.). Retrieved March 2, 2022, from <https://www.biodiversidad.gob.mx/diversidad/maderable>

- Mendizábal-Hernández L. C., Alba-Landa, J., Márquez Ramírez, J., Cruz-Jiménez, H. and Ramírez-García, E. O. 2011. Captura de carbono por *Cedrela odorata* l. en una prueba genética. *Revista Mexicana de Ciencias Forestales*, 2(4), pp. 105–111
- Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., Goodstein, D. M., Dubchak, I., Poliakov, A., Mizrachi, E., Kullán, A. R. K., Hussey, S. G., Pinard, D., Van Der Merwe, K., Singh, P. and Schmutz, J. 2014. The genome of *Eucalyptus grandis*. *Nature*, 510, pp. 356–362. <https://doi.org/10.1038/nature13308>
- Neale, D. B., McGuire, P. E., Wheeler, N. C., Stevens, K. A., Crepeau, M. W., Cardeno, C., Zimin, A. V., Puiu, D., Perteau, G. M., Sezen, U. U., Casola, C., Koralewski, T. E., Paul, R., Gonzalez-Ibeas, D., Zaman, S., Cronn, R., Yandell, M., Holt, C., Langley, C. H., and Wegrzyn, J. L. 2017. The Douglas-Fir genome sequence reveals specialization of the photosynthetic apparatus in Pinaceae. *G3: Genes, Genomes, Genetics*, 7(9), pp. 3157–3167. <https://doi.org/10.1534/G3.117.300078>
- Ouyang, K., Li, J., Zhao, X., Que, Q., Li, P., Huang, H. and Chen, X. 2016. Transcriptomic analysis of multipurpose timber yielding tree *Neolamarckia cadamba* during xylogenesis using RNA-seq. *PLoS One*, 11(7), e0159407. <http://doi.org/10.1371/journal.pone.0159407>
- Paiva, J. A. P., Prat, E., Vautrin, S., Santos, M. D., San-Clemente, H., Brommonschenkel, S., Fonseca, P. G. S., Grattapaglia, D., Song, X., Ammiraju, J. S. S., Kudrna, D., Wing, R. A., Freitas, A. T., Bergès, H. and Grima-Pettenati, J. 2011. Advancing Eucalyptus genomics: Identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. *BMC Genomics*, 12(1), p. 137. <https://doi.org/10.1186/1471-2164-12-137>
- Parent, G. J., Raheison, E., Sena, J. and MacKay, J. J. 2015. Forest Tree Genomics: Review of Progress. *Advances in Botanical Research*, 74, pp. 39–92. <https://doi.org/10.1016/BS.ABR.2015.05.004>
- Plomion, C., Aury, J. M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., Faye, S., Francillon, N., Labadie, K., Le Provost, G., Lesur, I., Bartholomé, J., Faivre-Rampant, P., Kohler, A., Lepié, J. C., Chantret, N., Chen, J., Diévar, A., Alaeitabar, T. and Salse, J. 2018. Oak genome reveals facets of long lifespan. *Nature Plants*, 4(7), pp. 440–452. <https://doi.org/10.1038/s41477-018-0172-3>
- Proost, S., Bel, M. Van, Vanechoutte, D., Van De Peer, Y., Inzé, D., Mueller-Roeber, B. and Vandepoele, K. 2015. PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Research*, 43(Database issue), D974. <https://doi.org/10.1093/NAR/GKU986>
- Qin, L., Hu, Y., Wang, J., Wang, X., Zhao, R., Shan, H., Li, K., Xu, P., Wu, H., Yan, X., Liu, L., Yi, X., Wanke, S., Bowers, J. E., Leebens-Mack, J. H., dePamphilis, C. W., Soltis, P. S., Soltis, D. E., Kong, H. and Jiao, Y. 2021. Insights into angiosperm evolution, floral development and chemical biosynthesis from the *Aristolochia fimbriata* genome. *Nature Plants*, 7(9), pp. 1239–1253. <https://doi.org/10.1038/s41477-021-00990-2>
- Ragauskas, A. J., Beckham, G. T., Biddy, M. J., Chandra, R., Chen, F., Davis, M. F., Davison, B. H., Dixon, R. A., Gilna, P., Keller, M., Langan, P., Naskar, A. K., Saddler, J. N., Tschaplinski, T. J., Tuskan, G. A. and Wyman, C. E. 2014. Lignin valorization: Improving lignin processing in the biorefinery. *Science*, 344(6185), p. 1246843. <https://doi.org/10.1126/SCIENCE.1246843>
- Ramage, M. H., Burrige, H., Busse-Wicher, M., Fereday, G., Reynolds, T., Shah, D. U. and Scherman, O. 2017. The wood from the trees: The use of timber in construction. *Renewable and Sustainable Energy Reviews*, 68, pp. 333–359.
- Raven, P. and Wackernagel, M. 2020. Maintaining biodiversity will define our long-term success. *Plant Diversity*, 42(4), pp. 211–220. <https://doi.org/10.1016/J.PLD.2020.06.002>
- Regalado, J., Lundberg, D. S., Deusch, O., Kersten, S., Karasov, T., Poersch, K., Shirsekar, G. and Weigel, D. 2020. Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe–microbe interaction networks in plant leaves. *The ISME Journal*, 14(8), p. 8. <https://doi.org/10.1038/s41396-020-0665-8>
- Rhoads, A. and Au, K. F. 2015. PacBio sequencing and its applications. *Genomics, proteomics & bioinformatics*, 13(5), pp. 278–289. <https://doi.org/10.1016/j.gpb.2015.08.002>

- Romo-Lozano, J. L., Vargas-Hernández, J. J., López-Upton, J., Ávila Angulo, M. L., Romo-Lozano, J. L., Vargas-Hernández, J. J., López-Upton, J. and Ávila Angulo, M. L. 2017. Estimación del valor financiero de las existencias maderables de cedro rojo (*Cedrela odorata* L.) en México. *Madera y Bosques*, 23(1), pp. 111–120. <https://doi.org/10.21829/MYB.2017.231473>
- Rueda Sánchez, A., Gallegos Rodríguez, A., González Eguiarte, D., Ruiz Corral, J. A., Benavides Solorio, J. de D., López Alcocer, E. and Acosta Mireles, M. 2014. Estimación de biomasa aérea en plantaciones de *Cedrela odorata* L. y *Swietenia macrophylla* King. *Revista Mexicana de Ciencias Forestales*, 5(25), pp. 8–17.
- Sanger, F. and Coulson, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3), pp. 441–448. [https://doi.org/10.1016/0022-2836\(75\)90213-2](https://doi.org/10.1016/0022-2836(75)90213-2)
- Secretaría de Medio Ambiente y Recursos Naturales. 2018. *Anuario estadístico de la producción forestal*. <https://dsiappsdev.semarnat.gob.mx/datos/portal/publicaciones/2021/2018.pdf>
- Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A. and Waterston, R. H. 2017. DNA sequencing at 40: past, present and future. *Nature*, 50(7676), pp. 345–353. <https://doi.org/10.1038/nature24286>
- Simpson, M. 2010. Plant systematics: Second edition. *Plant Systematics: Second Edition*, pp.1–741.
- Slatko, B. E., Gardner, A. F. and Ausubel, F. M. 2018. Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, 122(1), p. e59. <https://doi.org/10.1002/CPMB.59>
- State of the World's Forests 2016 - World | ReliefWeb. (n.d.). Retrieved March 2, 2022, from <https://reliefweb.int/report/world/state-worlds-forests-2016>
- Soto Sedano, J. C. and López Carrascal, C. E. 2012. RNA-seq: herramienta transcriptómica útil para el estudio de interacciones planta-patógeno. *Fitosanidad*, 16(2), pp. 101–113. <https://www.redalyc.org/html/2091/209126216009/>
- Sundell, D., Street, N. R., Kumar, M., Mellerowicz, E. J., Kucukoglu, M., Johnsson, C., Kumar, V., Mannapperuma, C., Delhomme, N., Nilsson, O., Tuominen, H., Pesquet, E., Fischer, U., Niittylä, T., Sundberg, B. and Hvidsten, T. R. 2017. AspWood: High-Spatial-Resolution Transcriptome Profiles Reveal Uncharacterized Modularity of Wood Formation in *Populus tremula*. *The Plant Cell*, 29(7), pp. 1585–1604. <https://doi.org/10.1105/TPC.17.00153>
- Torales, S. L., Rivarola, M., Gonzalez, S., Inza, M. V., Pomponio, M. F., Fernández, P., Acuña, C. v, Zelener, N., Fornés, L., Hopp, H. E., Paniego, N. B. and Marcucci Poltri, S. N. 2018. De novo transcriptome sequencing and SSR markers development for *Cedrela balansae* C.DC., a native tree species of northwest Argentina. *PLoS One*, 13(12), p. e0203768. <https://doi.org/10.1371/journal.pone.0203768>
- Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, M., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A. and Rokhsar, D. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, 313(5793), pp. 1596–1604. https://doi.org/10.1126/SCIENCE.1128691/SUPPL_FILE/TUSKAN.SOM.PDF
- Uhlen M. and Quake S. R. 2023. Sequential sequencing by synthesis and the next-generation sequencing revolution. *Trends in Biotechnology*, in press, pp. 1-8. <https://doi.org/10.1016/j.tibtech.2023.06.007>
- Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van De Peer, Y., Coppens, F. and Vandepoele, K. 2018. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research*, 46(Database issue), D1190–D1196. <https://doi.org/10.1093/NAR/GKX1002>
- Vargas, P. and Zardoya, R., 2012. *El árbol de la vida: sistemática y evolución de los seres vivos*. Madrid:International unión of biological sciences. <https://bibdigital.rjb.csic.es/idurl/1/1742751>
- Voshall, A., Behera, S., Li, X., Yu, X. H., Kapil, K.,

- Deogun, J. S., Shanklin, J., Cahoon, E. B. and Moriyama, E. N. 2021. A consensus-based ensemble approach to improve transcriptome assembly. *BMC Bioinformatics*, 22(1), pp. 1–24. <https://doi.org/10.1186/S12859-021-04434-8>
- Wang, Y., Zhao, Y., Bollas, A., Wang, Y. and Au, K. F. 2021. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 2021 39:11, 39(11), pp. 1348–1365. <https://doi.org/10.1038/s41587-021-01108-x>
- Weber, J. L. and Myers, E. W. 1997. Human whole-genome shotgun sequencing. *Genome Research*, 7(5), pp. 401–409. <https://doi.org/10.1101/gr.7.5.401>
- Yasodha, R., Vasudeva, R., Balakrishnan, S., Sakthi, A. R., Abel, N., Binai, N. and Dev, S. A. 2018. Draft genome of a high value tropical timber tree, Teak (*Tectona grandis* L. f): insights into SSR diversity, phylogeny and conservation. *DNA Research*, 25(4), pp. 409–419. <http://doi.org/dnares/dsy013>
- Yaya Lancheros, M. L., Rai, K. M., Balasubramanian, V. K., Dampanaboina, L., Mendu, V. and Terán, W. 2021. De novo transcriptome analysis of white teak (*Gmelina arborea* Roxb) wood reveals critical genes involved in xylem development and secondary metabolism. *BMC genomics*, 22(1), 494. <http://doi.org/10.1186/s12864-021-07777-x>
- Zhang, Q. J. and Gao, L. Z. 2014. The complete chloroplast genome sequence of desert poplar (*Populus euphratica*). *Mitochondrial DNA A DNA Mapping, Sequencing and Analysis*, 27(1), pp. 721–7233. <https://doi.org/10.3109/19401736.2014.913159>