



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE ECOLOGÍA

BIOLOGÍA EVOLUTIVA

**ANÁLISIS EVOLUTIVO DE LOS GENES DE LA FAMILIA CDPK EN CONÍFERAS DE
AMBIENTES CONTRASTANTES**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

JOSÉ MIGUEL AMARO ESTRADA

TUTOR PRINCIPAL DE TESIS: DR. JUAN PABLO JARAMILLO CORREA
INSTITUTO DE ECOLOGÍA, UNAM

COMITÉ TUTOR:

DRA. ALICIA GAMBOA DE BUEN
INSTITUTO DE ECOLOGÍA, UNAM

DR. ARTURO CARLOS II BECERRA BRACHO
FACULTAD DE CIENCIAS, UNAM

CIUDAD UNIVERSITARIA, CD. MX., OCTUBRE DE 2022



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE ECOLOGÍA

BIOLOGÍA EVOLUTIVA

ANÁLISIS EVOLUTIVO DE LOS GENES DE LA FAMILIA CDPK EN CONÍFERAS DE

AMBIENTES CONTRASTANTES

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

JOSÉ MIGUEL AMARO ESTRADA

TUTOR PRINCIPAL DE TESIS: DR. JUAN PABLO JARAMILLO CORREA
INSTITUTO DE ECOLOGÍA, UNAM

COMITÉ TUTOR:

DRA. ALICIA GAMBOA DE BUEN
INSTITUTO DE ECOLOGÍA, UNAM

DR. ARTURO CARLOS II BECERRA CAMACHO
FACULTAD DE CIENCIAS, UNAM

MÉXICO, CD. MX. ,2022

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS

ENTIDAD INSTITUTO DE ECOLOGÍA

OFICIO CPCB/829/2022

ASUNTO: Oficio de Jurado

M. en C Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM

Presente

Me permito informar a usted que, que el Comité Académico, del Posgrado en Ciencias Biológicas, en su reunión ordinaria del día 20 de junio de 2022, aprobó el siguiente jurado para el examen de grado de **MAESTRO EN CIENCIAS BIOLÓGICAS** en el campo de conocimiento de **Biología Evolutiva** del alumno **AMARO ESTRADA JOSÉ MIGUEL**, con número de cuenta: **309000185** con la tesis titulada: "**ANÁLISIS EVOLUTIVO DE LOS GENES DE LA FAMILIA CDPK EN CONÍFERAS DE AMBIENTES CONTRASTANTES**", bajo la dirección del **DR. JUAN PABLO JARAMILLO CORREA**, Tutor Principal, quedando integrado de la siguiente manera:

Presidente: DR. DAVID SEBASTIAN GERANDT

Vocal: DRA. ALICIA MASTRETTA YANES

Vocal: DRA. ALEJANDRA VÁZQUEZ LOBO YURÉN

Vocal: DR. LEV ORLANDO JARDÓN BARBOLLA

Secretario: DRA. ALICIA GAMBOA DE BUEN

Sin otro particular, me es grato enviarle un cordial saludo.

ATENTAMENTE
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Ciudad Universitaria, Cd. Mx., a 09 de septiembre de 2022

COORDINADOR DEL PROGRAMA



DR. ADOLFO GERARDO NAVARRO SIGÜENZA



AGRADECIMIENTOS INSTITUCIONALES

Al Posgrado en Ciencias Biológicas de la UNAM.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo recibido mediante la beca de maestría (CVU: **1005075**).

Este trabajo fue apoyado financieramente por becas del Consejo Nacional de Ciencia y Tecnología (CONACyT; CB-2016-284457 y 278987), la Dirección General de Asuntos del Personal Académico de la UNAM (proyecto PAPIIT: IN208416), y la presupuesto interno asignado (“presupuesto operativo”) de IE-UNAM al Dr. Juan Pablo Jaramillo Correa.

Este trabajo fue posible gracias a la orientación de mi tutor, el Dr. Juan Pablo Jaramillo, y de los integrantes de mi comité tutor, la Dra. Alicia Gamboa de Buen y el Dr. Arturo Carlos II Becerra Bracho.

AGRADECIMIENTOS PERSONALES

En primer lugar, agradezco a mis padres, por haberme dado todo el apoyo siempre para estudiar y dedicarme a lo que me apasiona.

A los integrantes del equipo oyamel, por el apoyo, las charlas, la orientación, y el ejemplo, gracias Jorge y Sebas. A Gustavo, además de lo anterior, por la amistad y todos los buenos momentos a lo largo de tanto años. A la maestra Verito, por todo lo anterior y por todo el apoyo y trabajo en conjunto, en muchos sentidos fuiste la co-tutora de este trabajo. Y en general, a mis compañeros del laboratorio de Genética y Ecología, por los ratos de convivencia que siempre sirvieron para mantener el ánimo y la cordura.

A la Dra. Nancy Gálvez, con quien tuve la suerte de contar para todo el apoyo técnico y logístico de mi trabajo en los difíciles tiempos de la pandemia por COVID-19, y también por siempre darme ánimos y mantener el entusiasmo.

Al Dr. Juan Pablo Jaramillo, mi tutor desde la licenciatura, por todo el aprendizaje, tanto académico como personal. A los integrantes de mi comité tutor, Dra. Alicia Gamboa y Dr. Arturo Becerra, sus comentarios y orientación durante la realización de este trabajo fueron muy valiosos.

También a los integrantes de mi jurado, Dr. David Gernandt, Dr. Lev Jardón, Dra. Alejandra Vázquez, por sus colaboración para mejorar mi trabajo y a la Dra. Alicia Mastretta, por lo anterior y también por ser una excelente maestra en el salón de clases.

TABLA DE CONTENIDO

1. RESUMEN	1
2. ABSTRACT	2
3. INTRODUCCIÓN	4
3.1. Duplicaciones y familias génicas	4
3.2. El genoma de las gimnospermas	8
3.3. Estrés vegetal y señalización por calcio	11
3.3.1. Generalidades sobre el estrés en plantas	11
3.3.2. El calcio como mensajero secundario en rutas de señalización por estrés	12
3.4. La familia CDPK	14
3.4.1. Estructura y regulación de las CDPK	14
3.4.2. Las proteínas CDPK en la regulación de la respuesta a estrés y del desarrollo	17
3.4.3. Origen y evolución de la familia CDPK	18
4. MÉTODO	22
4.1. Obtención e identificación de las secuencias	22
4.2. Elaboración de matrices y alineamiento	24
4.3. Análisis filogenéticos	24
4.4. Pruebas de selección y tasas evolutivas	25
4.4.1. Modelos evolutivos por codones	25
4.4.2. Modelos por rama	25
4.4.3. Modelos por sitio	26
4.4.4. Sitios bajo selección episódica	27
4.4.5. Divergencia funcional entre genes duplicados y efectos deletéreos de las sustituciones	28
4.5. Localización relativa de los genes CDPK en mapas de ligamiento de coníferas	29
5. RESULTADOS	31
5.1. Número y características de las secuencias CDPK obtenidas	31
5.2. Filogenia de la familia CDPK en coníferas	33

5.3.	Tasas evolutivas y pruebas de selección	34
5.4.	Divergencia funcional entre genes duplicados y efectos deletéreos de las sustituciones.	38
5.5.	Localización de los genes CDPK en mapas de ligamiento de las coníferas	39
6.	DISCUSIÓN	41
6.1.	Características, filogenia y ubicación genómica relativa de los genes <i>CDPK</i> en coníferas	41
6.2.	Tasas evolutivas y selección natural en los genes <i>CDPK</i> de coníferas	45
7.	CONCLUSIÓN	49
8.	FUENTES	50
9.	ANEXOS	64

1. RESUMEN

Las proteínas cinasas dependientes de calcio (**CDPK**) tienen un papel fundamental en la regulación de la respuesta a estrés biótico y abiótico, y del crecimiento y desarrollo de las plantas. Estas proteínas han sido ampliamente estudiadas en angiospermas modelo y de interés comercial; sin embargo, poco se sabe sobre estas proteínas en coníferas. Este linaje de plantas terrestres está adaptado a una gran diversidad de ambientes; desde latitudes boreales hasta tropicales, y desde ambientes de alta montaña, hasta nivel del mar. Por lo tanto, resulta importante estudiar el papel que han tenido las proteínas **CDPK** en la adaptación de las coníferas a estas diferentes condiciones ambientales.

En este trabajo se utilizaron datos transcriptómicos para identificar las secuencias de esta familia génica en 14 especies de coníferas, incluyendo tres familias; así como pares de especies adaptadas a distintos tipos de ambientes. Se identificaron 10 parálogos en las familias Taxaceae y Cupressaceae, y 11 en la familia Pinaceae, con evidencia de al menos una duplicación en tándem en esta última. Los genes **CDPK** de coníferas presentaron altos niveles de homología con los de angiospermas, y se agruparon cercanamente a estos dentro de los cuatro clados previamente reportados para esta familia génica. También presentaron la misma estructura canónica encontrada en otros linajes (un dominio cinasa y dos pares *EF-hand*) y se encontraron organizados con altos niveles de sintenia entre especies.

Se observaron altos niveles de conservación en estos genes, lo cual sugiere que su función ha sido importante durante la evolución de las coníferas. Sin embargo, se encontraron mayores tasas evolutivas después de la duplicación ocurrida en la familia Pinaceae, y en el patrón de divergencia no sinónima en uno de dichos duplicados (**CDPK34**). Esto sugiere que este gen se encuentra en proceso de pseudogenización. También se encontraron sitios bajo selección positiva recurrente en cinco genes, además de evidencia de selección episódica en diez de ellos. De estos últimos, tres casos mostraron la señal de selección de manera casi exclusiva en especies individuales y, de manera interesante, estas habitan ambientes contrastantes. El gen **CDPK6** mostró la mayor evidencia de selección en la especie subtropical *A. religiosa*; el **CDPK17**, lo hizo en la especie boreal *P. banksiana*; mientras que en el gen **CDPK10**, dicha evidencia se observó en la especie subtropical *P. patula*. Las funciones de estos genes incluyen la germinación del tubo polínico (**CDPK6** y **CDPK17**) y la tolerancia a la sequía

(*CDPK10*).

En conjunto, los resultados de este trabajo sugieren que la familia ***CDPK*** ha jugado un papel importante en la evolución de las coníferas y en su adaptación a diferentes condiciones ambientales. En el futuro, estos resultados podrán complementarse con análisis de expresión, para entender mejor el papel funcional de estas proteínas en la tolerancia de las coníferas a distintos tipos de estrés.

2. ABSTRACT

Calcium-dependent protein kinases (***CDPK***) play a fundamental role in the regulation of the response to both biotic and abiotic stress, as well as in growth and development in plants. These proteins have been widely studied in model and commercially important angiosperms; however, little is known about them conifers. This plant lineage is adapted to a great diversity of environments: from boreal to subtropical latitudes and from high mountain to sea-level environments. Therefore, it is important to understand the role that ***CDPK*** genes have played in the adaptation of conifers to such environmental variability.

In this work, we used transcriptomic data to pinpoint paralogs from this gene family in 14 conifer species, from three families, as well as pairs of species adapted to different types of environments. Ten paralogs were identified in the Taxaceae and Cupressaceae, and 11 more in the Pinaceae family, which included a possible tandem duplication. Angiosperm and conifer ***CDPK*** genes showed high levels of homology, being all closely grouped within the four previously reported clades. All paralogs further presented the same canonical structure reported for other lineages (one kinase domain and two EF-hand pairs), which were all syntenically organized among species.

High levels of conservation were observed in among paralogs, suggesting an important and conserved role during the evolution of conifers. However, higher evolutionary rates were found after the duplication event exclusive to the Pinaceae. Indeed, there was an excess of non-synonymous substitutions in one of these duplicates (*CDPK34*), pointing to a possible pseudogenization. Sites under recurrent positive selection were also found in five more paralogs; in addition to episodic selection in ten other genes. Three of such genes showed

selection signals in individual species all of which inhabited contrasting environments. The *CDPK6* gene showed the greatest evidence of selection in the subtropical species *A. religiosa*; *CDPK17* did so in the boreal species *P. banksiana*; while for the *CDPK10* gene, such evidence was observed in the subtropical species *P. patula*. The functions of these genes include pollen tube germination (*CDPK6* and *CDPK17*) and drought tolerance (*CDPK10*).

Altogether, our results suggest that the ***CDPK*** family has played an important role in the evolution of conifers, and their adaptation to contrasting environments. They further call for comparative expression analyses, to better explore the functional role of these proteins and their role in stress response in conifers.

3. INTRODUCCIÓN

3.1. Duplicaciones y familias génicas

Uno de los principales mecanismos mediante los que los organismos son capaces de desarrollar innovaciones evolutivas es la generación de nuevos genes. Si bien recientemente se han estudiado diversos procesos que pueden dar origen a genes nuevos, la duplicación de genes preexistentes ha sido el proceso mejor documentado mediante el cual estos se generan (Ding et al., 2012). De manera general, una duplicación génica produce dos copias funcionalmente redundantes de un gen; de las cuales al menos una puede evolucionar con menor restricción, con lo cual eventualmente puede desarrollar una nueva función a través de selección natural (Panchy et al., 2016).

La duplicación génica puede ocurrir a través de distintos procesos y a diferentes escalas. A gran escala, los procesos incluyen la duplicación del genoma completo (WGD) y las poliploidías. Estos son particularmente importantes en el caso de las plantas, ya que un gran porcentaje de las especies actuales son poliploides; mientras que los genomas modernos de las plantas terrestres muestran múltiples huellas de poliploidizaciones ancestrales (Flagel & Wendel, 2009). Debido a la prevalencia de las duplicaciones genómicas en plantas, estas tienen la mayor contribución a los genes duplicados en sus genomas (Bowers et al., 2003; Panchy et al., 2016). Por otro lado, y a una escala menor, los genes pueden duplicarse en tándem y generar bloques consecutivos en el genoma, lo cual suele darse cuando ocurre entrecruzamiento desigual entre los cromosomas (Y. Wang, Wang, et al., 2012). Por otro lado, pueden ocurrir duplicaciones de bloques de genes, lo cual se da mediante la duplicación segmental, en la cual, fragmentos o arreglos completos son copiados y dispersados hacia otras regiones del genoma (Flagel & Wendel, 2009). Finalmente, existen mecanismos de duplicación génica que involucran la acción de elementos transponibles (TE). Esto ocurre cuando un gen queda atrapado dentro de los límites de un TE, con lo cual, al amplificarse este, se generan copias del gen capturado. En plantas, el ejemplo más notable de esto es la captura génica por *Mutator-like elements* (MULE), de los cuales existen desde 46 en *A. thaliana* hasta 3,000 en *Oryza sativa* (Flagel & Wendel, 2009; Panchy et al., 2016). Otro mecanismo asociado a TE es la retroduplicación, en la cual, un RNA mensajero se retro-transcribe y se inserta en alguna otra región del genoma (Panchy et al., 2016). De manera general, se pueden determinar los mecanismos de duplicación que han generado dos copias mediante la evaluación de la sintenia y colinearidad de los genes, es decir, su localización en cromosomas y/o regiones cromosómicas

correspondientes y con un orden conservado (Y. Wang, Tang, et al., 2012). De esta manera, se pudo identificar que los genes duplicados mediante WGD o poliploidías presentan altos niveles de sintenia y colinearidad, que las duplicaciones segmentales no suelen conservar la sintenia, pero sí la colinearidad, y que las duplicaciones de genes individuales mediante transposones generan copias dispersas que no conservan su orden ni su localización (Freeling, 2009). Además, es posible que se generen copias génicas mediante combinaciones de los diferentes procesos arriba mencionados y, en ocasiones, pueden generarse copias quiméricas, es decir, que se originan a partir de múltiples loci parentales. Tal es el caso de los genes *jingwei* y *sphinx* en especies de *Drosophila* (H. Chen et al., 2017; Long et al., 2003). En otros casos, los genes quiméricos pueden generarse mediante uno solo de los procesos previamente discutidos (p. ej. entrecruzamiento desigual o duplicación segmental), como se ha reportado para algunos genes duplicados en especies de *Oryza* (C. Zhang et al., 2013).

Una vez generada una copia génica, diferentes procesos pueden llevar a su retención y fijación en el genoma de las especies. En ausencia de efectos selectivos o cambios funcionales, la deriva génica es el principal proceso que determina el destino de una copia génica. En este escenario, el tiempo de fijación para una copia es una función del tamaño poblacional efectivo (N_e); los efectos sobre la adecuación que tendría la pérdida de ambas copias y de la tasa de mutación en el locus duplicado, con lo cual, el tiempo promedio esperado de fijación se encuentra entre 3 y 20 N_e . Es decir, algunas copias génicas pueden sobrevivir por varios millones de años debido a deriva génica, y no porque su presencia sea benéfica (Panchy et al., 2016).

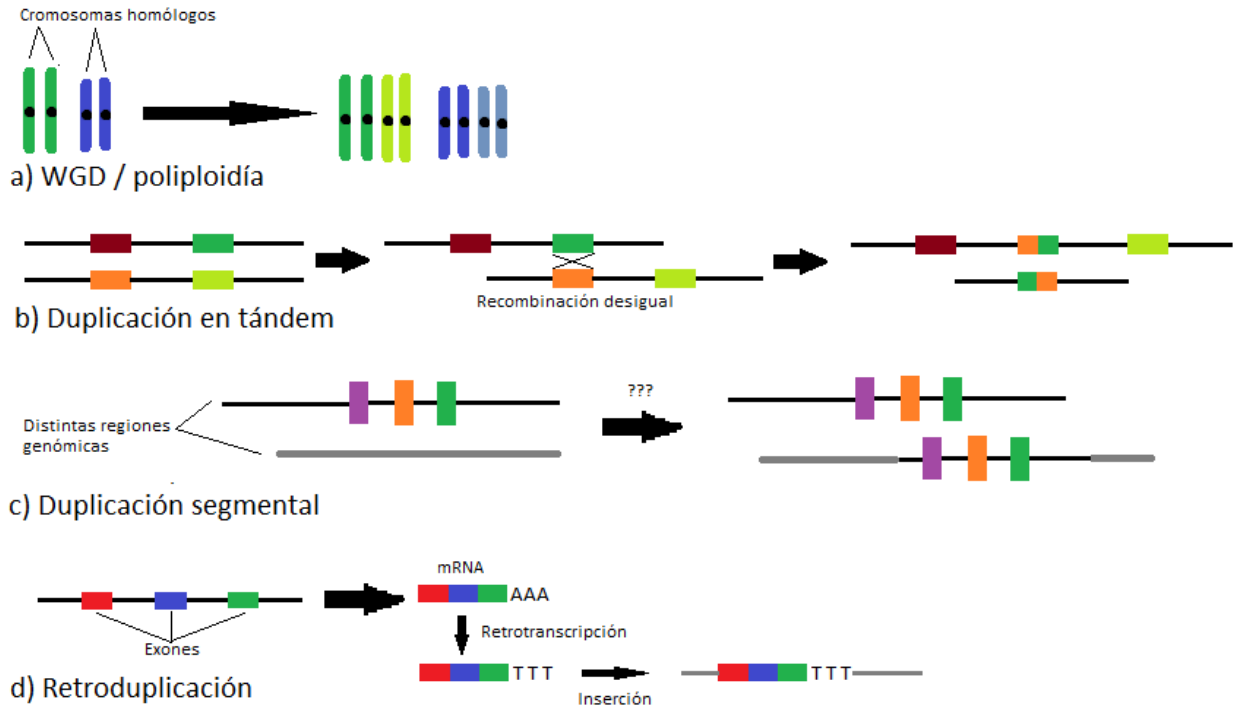


Figura 1. Mecanismos de duplicación génica. Modificada a partir de Panchy et al., 2016.

Alternativamente, la selección natural puede promover la retención de una copia génica, ya sea mediante selección sobre funciones preexistentes, o sobre funciones que se adquieren después de la duplicación. El primer caso está mejor representado por el modelo de subfuncionalización. Bajo este modelo, también llamado de duplicación-degeneración-complementación (DDC), la función original del gen ancestral se divide entre las nuevas copias, las cuales pierden, cada una, una parte distinta de dicha función original y, por lo tanto, ambos duplicados se conservan. Este modelo es atractivo para explicar el alto número de genes duplicados en eucariontes, ya que solo requiere de un gen progenitor y de mutaciones degenerativas ordinarias, sin necesidad de invocar un cambio funcional (Conant & Wolfe, 2008). Otros mecanismos que implican la retención de una copia génica por medio de la selección de funciones preexistentes incluyen la selección sobre el incremento de la dosis génica, el balance génico y la interferencia de los parálogos (Panchy et al., 2016).

En cuanto a la retención por medio de selección sobre funciones novedosas, el principal modelo es el de neofuncionalización. Bajo este escenario, ocurren mutaciones en una de las copias después de la duplicación, que le confieren una nueva función, mientras que la otra copia,

conserva la función ancestral. Si la nueva función contribuye a un aumento en la adecuación, entonces ambas copias se retienen (Conant & Wolfe, 2008; Panchy et al., 2016). Por otra parte, en el modelo EAC (*Escape from Adaptive Conflict*) se propone la adquisición de una función novedosa, pero en este caso, esta aparece en el gen ancestral, previo a la duplicación, con lo cual, la capacidad del gen para realizar la función original disminuye y ambas no pueden ser optimizadas simultáneamente (conflicto adaptativo). Una vez ocurrida la duplicación, cada una de las copias puede especializarse en una de las funciones del gen ancestral, para luego optimizarse de manera individual (Conant & Wolfe, 2008; des Marais & Rausher, 2008).

A pesar de los mecanismos mediante los cuales se pueden retener copias génicas, el principal destino de estas es su eliminación. En general, el proceso de pérdida de copias implica la delección total de una de ellas o la pseudogenización. Esta última consiste en la pérdida de función de una de las copias por medio de la acumulación de mutaciones deletéreas. Estos pseudogenes son mantenidos en el genoma como copias degeneradas o no funcionales, que pueden identificarse a partir de su similitud con genes anotados, pero caracterizados por la presencia de codones de *stop* prematuros o cambios en el marco de lectura (Panchy et al., 2016). Adicionalmente, los pseudogenes pueden ser regulados y silenciados por factores epigenéticos, como la metilación del ADN, la cual a su vez regula otros aspectos de la actividad genómica, como la proliferación de transposones (Davis et al., 2015). Por ejemplo, en *A. thaliana*, las secuencias identificadas como pseudogenes tienen mayores niveles de metilación que los genes funcionales, y la disrupción de la función de las metiltransferasas resulta en una mayor actividad transcripcional de los pseudogenes y los transposones, particularmente en las regiones centro y pericentroméricas (X. Zhang et al., 2006).

Después de duplicarse, las copias presentan diferentes características que las distinguen de los genes de copia única, y estas pueden estar influidas por el mecanismo de duplicación que genera las copias. Primero, los genes duplicados suelen tener tasas evolutivas más altas, al menos inicialmente, como resultado de la relajación de la presión selectiva, y en algunos casos, de la acumulación de mutaciones que producen funciones novedosas en alguna de las copias, mediante selección natural (Panchy et al., 2016). En cuanto al efecto de los mecanismos de duplicación en la tasa evolutiva, se ha encontrado que las copias individuales de genes tienen mayor número de sustituciones no sinónimas con respecto a los genes duplicados vía WGD, lo cual sugiere que los primeros están sujetos a menores restricciones

selectivas (Carretero-Paulet & Fares, 2012). En segundo lugar, los genes duplicados pueden presentar patrones divergentes de expresión, lo cual, en general, está relacionado positivamente con el nivel de divergencia no sinónima a nivel de secuencia y, en algunos casos, con la divergencia sinónima y la edad de las copias (Ganko et al., 2007; Y. Wang, Wang, et al., 2012). La divergencia en expresión también puede estar influida por el mecanismo de duplicación. En general, se ha reportado que los genes que se originan mediante procesos a gran escala, por ejemplo, vía WGD, tienen patrones de expresión más redundantes, sobrelapados, y correlacionados entre sí, así como en los mismos tejidos; mientras que los genes duplicados a menor escala o en segmentos no co-lineales tienden a divergir asimétricamente (Wang, Wang, et al., 2012). Estas diferencias en los patrones de divergencia en expresión están principalmente dadas por los efectos del proceso de duplicación sobre el contexto regulatorio de los genes. Por ejemplo, las duplicaciones de genes individuales producidas por entrecruzamiento desigual o por retrotransposiciones, tienden a generar disrupción en los promotores de las copias, mientras que los genes que se duplican vía WGD o poliploidía son menos propensos a esta disrupción (Casneuf et al., 2006). Finalmente, las categorías funcionales de los genes duplicados no son aleatorias y también suelen estar relacionadas con el mecanismo de duplicación. Por ejemplo, en *A. thaliana* los genes relacionados con la regulación transcripcional y la transducción de señales tienden a ser retenidos después de una duplicación genómica completa, mientras que genes involucrados en la respuesta a estrés ambiental se retienen preferentemente después de duplicaciones en tándem en *A. thaliana* y *O. sativa* (Panchy et al., 2016). De manera similar, los genes relacionados con actividades catalíticas suelen estar poco representados entre los genes duplicados mediante WGD, pero no para las duplicaciones sencillas, mientras que los genes con actividad cinasa suelen estar preferentemente retenidos, especialmente en duplicaciones a escala pequeña (Carretero-Paulet & Fares, 2012).

3.2. El genoma de las gimnospermas

Las gimnospermas tienen un genoma, en promedio, más grande que el de las angiospermas, estando entre los 10 y 35 Gbp (Neale & Wheeler, 2019; B. J. M. Zonneveld, 2012). Esto representa un genoma que es 200, 24 y 10 veces más grande que el de *A. thaliana*, *Populus trichocarpa* y *Zea mays*, respectivamente (Prunier et al., 2016). Sin embargo, el genoma de las gimnospermas presenta una variación de solamente 16 veces en su tamaño, comparado con el de las angiospermas, con una variación de hasta 2,400 veces, aunque con una distribución

sesgada hacia genomas pequeños (Leitch & Leitch, 2012). Además del tamaño del genoma, el número, tamaño y estructura de los cromosomas también se encuentran altamente conservado entre las gimnospermas, y particularmente entre las coníferas, cuyo número haploide va de los 9 a los 33 cromosomas (Prunier et al., 2016). La mayoría de las coníferas de la familia Pinaceae, por ejemplo, tiene 12 cromosomas ($2n = 24$), con excepciones como *Pseudotsuga menziesii*, que tiene un número diploide de 26 mientras que en *Pseudolarix amabilis*, este es de 44. Por su parte, en la familia Cupressaceae el número diploide es de 22 en la mayoría de las especies, sin embargo, dentro de esta familia está la especie *Sequoia sempervirens*, con 66 cromosomas (Neale & Wheeler, 2019).

A diferencia del genoma de las angiospermas, el de las coníferas presenta poca evidencia de duplicaciones genómicas completas o poliploidías recientes (salvo algunas excepciones como *Sequoia*) y, de manera similar, la especiación mediante poliploidización en gimnospermas es poco frecuente, con solo el 1.5% de las especies de coníferas reportadas como poliploides (Li et al., 2015; Pavy et al., 2012). En comparación, alrededor del 15% y el 31% de las especies de angiospermas y helechos, respectivamente, ha sido reportados como poliploides (Wood et al., 2009). Inicialmente, la única duplicación genómica reportada para las gimnospermas ocurrió antes de la divergencia entre estas y el linaje de las angiospermas (Nystedt et al., 2013). Sin embargo, posteriormente se han reportado tres más: una en el ancestro de la familia Pinaceae, otra en el ancestro de la familia Cupressaceae y una más en el linaje de *Welwitschia* (Li et al., 2015).

El genoma de las gimnospermas también se caracteriza por una gran acumulación de transposones, particularmente del tipo LTR-RT (*long terminal repeat retrotransposons*), los cuales representan alrededor del 41% y 58% del genoma total en *Pinus taeda* y *Picea abies*, respectivamente (Nystedt et al., 2013; Wegrzyn et al., 2014). Estos transposones se han expandido con una tasa lenta y constante a lo largo de la evolución del linaje de las gimnospermas. En las angiospermas, esta expansión ha sido contrarrestada por eficientes mecanismos de remoción (por ejemplo, mediante recombinación), sin embargo, estos han sido poco eficientes en las gimnospermas, y la expansión de transposones parece ser el principal responsable del crecimiento de su genoma (Li et al., 2015).

En cuanto al contenido génico en las gimnospermas, este se encuentra entre los 37,936

(*S. sempervirens*) y los 80,495 genes (*P. tabuliformis*) (Niu et al., 2022; Scott et al., n.d.). Además, la evolución de familias génicas ha sido relativamente lenta en coníferas, en comparación con la rápida proliferación de estas en angiospermas. De hecho, la mayoría de los genes en coníferas son de copia única o pertenecen a familias pequeñas, con entre 2 y 10 miembros, mientras que solo unas pocas familias contienen más de 100 genes (De la Torre et al., 2015). Consistente con lo anterior, la mayoría de las duplicaciones génicas en coníferas preceden a la divergencia entre las angiospermas y las gimnospermas, además de que la mayoría de los genes se encuentran organizados con altos niveles de sintenia y co-linearidad entre genomas de distintas especies (Pavy et al., 2012). En cuanto a su estructura, los genes de las coníferas tienden a acumular intrones largos, entre los que se encuentran alrededor de 2,800 genes con intrones de más de 5 kbp de longitud encontrados en *Picea abies* (Nystedt et al., 2013). Además, los genes suelen estar separados por grandes regiones intergénicas, ricas en elementos transponibles, y ambas características están relacionadas con la actividad y acumulación de transposones previamente mencionada (Nystedt et al., 2013). Además, la fracción del genoma de las coníferas anotada como codificante suele ser mayor que lo esperado dado el número estimado de genes que contienen. Esto se ha explicado por una abundancia de pseudogenes, la cual también estaría relacionada con la inserción de transposones dentro de los límites de los genes, así como con la baja eficiencia de los mecanismos para remover estas secuencias duplicadas (Prunier et al., 2016).

En general, es posible que la relativa estasis del genoma de las coníferas, por ejemplo, la falta de duplicaciones genómicas y rearrreglos cromosómicos, estén relacionados con las bajas tasas de especiación y diversidad morfológica que estos organismos presentan (Nystedt et al., 2013). Sin embargo, estos procesos no necesariamente representan una desventaja adaptativa, ya que las coníferas son elementos dominantes en una gran diversidad de ecosistemas, los cuales van desde bosques subtropicales hasta ecosistemas boreales, y desde el nivel del mar, hasta locaciones de alta montaña, por ejemplo, 3000 msnm (*Taxus baccata*) en los Himalayas, más de 4000 y 4,500 msnm en México (*P. hartwegii* y *Juniperus monticola*) en México (Adams, 2014; Prunier et al., 2016; Shaheen et al., 2015). Esto revela que las coníferas poseen un gran potencial adaptativo, el cual posiblemente esté relacionado con los altos niveles de diversidad genética dentro de sus poblaciones, así como a una gran plasticidad fenotípica (Nystedt et al., 2013).

3.3. Estrés vegetal y señalización por calcio

3.3.1. Generalidades sobre el estrés en plantas

El estrés implica la presencia de un factor externo a la planta, proveniente del medio ambiente, que ejerce una influencia negativa sobre su desarrollo óptimo (Taiz et al., 2018). Las plantas responden a estos cambios ambientales mediante una alteración en su fisiología o morfología para mejorar su supervivencia y reproducción bajo estas nuevas condiciones. Cuando estos cambios no son permanentes, es decir, no requieren cambios genéticos y pueden ser reversibles si las condiciones ambientales previas se reestablecen, el proceso se denomina aclimatación (Taiz et al., 2018). En estos casos es importante mencionar la plasticidad fenotípica. Esta puede definirse como la habilidad de un genotipo individual para generar distintos fenotipos al ser expuestos a distintos estímulos ambientales. Esto incluye la capacidad de modificar la trayectoria del desarrollo de un individuo, o su capacidad de modificar su estado fenotípico (por ejemplo, su metabolismo) ante distintas condiciones ambientales (Fusco & Minelli, 2010). Por otro lado, si los cambios son permanentes o a largo plazo e implican la fijación de variantes alélicas en una población, el proceso se denomina adaptación (Taiz et al., 2018).

La manifestación de la respuesta ante el estrés pasa por una serie de etapas ordenadas. La primera etapa es la percepción del estímulo por parte de la planta. Esta percepción puede ser de tipo físico (p.ej. contracción de la membrana plasmática), biofísico (p. ej. cambios en la estructura proteica), metabólico (p. ej. acumulación de productos secundarios) y bioquímico (p. ej. canales iónicos que alteran la homeostasis de los iones) (Taiz et al., 2018). La segunda, consiste en el procesamiento de la señal percibida, lo cual implica su amplificación e integración a rutas de señalización. Estas rutas están formadas por diferentes moléculas mensajeras. Por ejemplo, los estímulos de estrés inducen cambios en la concentración intracelular de iones de calcio (Ca^{2+}) y de especies reactivas del oxígeno (EROS), las cuales indican a la célula que ha sido detectada una señal de estrés (Azcón Vieto & Talón, 2013). Estos cambios en la concentración de EROS y Ca^{2+} activan diferentes proteínas cinasas y fosfatasa, las cuales transmiten la señal río abajo mediante la fosforilación o defosforilación de diferentes proteínas; entre las que se incluyen factores de transcripción (Taiz et al., 2018). Esto lleva a la tercera etapa, que consiste en la modificación de la expresión génica. Esta implica tanto la activación como la represión transcripcional de genes cuyos productos proteicos son apropiados o no para la supervivencia de la planta bajo las nuevas condiciones ambientales (Azcón Vieto & Talón,

2013).

Las rutas de señalización no son aisladas, sino que existen diferentes mecanismos mediante los cuales estas establecen una comunicación cruzada. Por un lado, las cinasas activadas por mitógenos (MAPK) juegan un papel crucial en esto, ya que se ha encontrado que, por ejemplo, las respuestas a estrés por temperatura, salinidad, el estrés osmótico y la sequía, están regulados por módulos de la misma ruta MAPK básica (Taiz et al., 2018). Por el otro lado, las hormonas vegetales también regulan distintas respuestas a estrés. Por ejemplo, el ácido abscísico (ABA), está involucrado en la regulación de la apertura y cierre de estomas ante el estrés hídrico, así como en la adaptación al estrés por congelamiento y salinidad, mientras que las giberelinas regulan procesos que ligan eventos del desarrollo con las respuestas a distintos tipos de estrés (Taiz et al., 2018).

3.3.2. El calcio como mensajero secundario en rutas de señalización por estrés

Como se mencionó anteriormente, los iones de calcio (Ca^{2+}) son uno de los principales mensajeros secundarios en la respuesta ante distintos tipos de estrés, tanto bióticos como abióticos. Debido a que este ion forma fácilmente complejos insolubles con proteínas, membranas y ácidos orgánicos, resulta tóxico cuando sus concentraciones citosólicas son altas; por lo cual generalmente se encuentra almacenado en compartimentos como la vacuola, el retículo endoplásmico y la pared celular (Kudla et al., 2010; Taiz et al., 2018).

Estímulos como altas o bajas temperaturas, estrés salino y osmótico, desecación, y la acción de hormonas y elicitores derivados de patógenos, desencadenan cambios en la concentración citosólica de los iones de calcio (Hashimoto & Kudla, 2011). Estos cambios de concentración presentan patrones espaciotemporales específicos para cada tipo de estímulo, conocidos como “firmas de calcio”, los cuales se caracterizan por su duración, amplitud, frecuencia y distribución espacial (Kudla et al., 2010). Por ejemplo, en las células de la raíz de *A. thaliana*, el estrés por frío induce una respuesta monofásica, la cual consiste en un pico en la concentración de Ca^{2+} , tras lo cual la concentración vuelve a niveles de reposo (Kiegle et al., 2000). En contraste, el estrés osmótico genera una reacción bifásica específicamente en las células del periciclo y la endodermis, que consistente en un aumento inicial en la concentración de calcio, seguido de un segundo pico menor, antes de regresar a niveles ligeramente por arriba

de los de reposo (Kiegle et al., 2000). Estos patrones tan específicos se logran mediante un estricto control del eflujo y aflujo del ión a través de transportadores y proteínas canal localizadas tanto en la membrana plasmática como en la membrana de organelos celulares que actúan como almacenes (Hashimoto & Kudla, 2011). El aflujo es controlado por distintos canales iónicos, los cuales, según su mecanismo de activación, pueden clasificarse en canales dependientes de voltaje, canales dependientes de ligandos y canales activados por estiramiento (Kudla et al., 2010). De acuerdo con su mecanismo de activación, los canales de calcio pueden moldear los parámetros del aflujo, con lo cual la planta puede traducir un amplio rango de señales en distintas firmas de calcio. Por otro lado, el eflujo se logra mediante dos mecanismos. Primero, las Ca^{2+} -P-ATPasas median un eflujo de poca afinidad y alta capacidad, con el objetivo principal de mantener los niveles basales de calcio (Kudla et al., 2010). Segundo, los antiportadores de Ca^{2+} /protones median un eflujo de alta afinidad, pero baja capacidad, y están principalmente asociados al restablecimiento de los niveles de calcio, posterior a un aumento mediado por alguna señal (Kudla et al., 2010). En cuanto a la especificidad espacial, esta se logra mediante una combinación en los parámetros del eflujo y la liberación del calcio desde reservorios específicos, tanto desde el interior como del exterior de la célula, y la lenta difusión citoplasmática del ion, con lo cual se obtienen señales altamente localizadas (Hashimoto & Kudla, 2011).

Las plantas poseen un amplio repertorio de proteínas de unión a calcio, que detectan los cambios específicos en la concentración de este ión y traducen esta información en señales moleculares y bioquímicas específicas (Kudla et al., 2010). Estas, en última instancia, resultan en cambios en la expresión génica. Este repertorio está compuesto por familias complejas con dominios estructurales *EF-hand*, las cuales en conjunto suman hasta 250 miembros en el genoma de *A. thaliana*, en comparación con 132 proteínas con este dominio en *Drosophila melanogaster* (Hashimoto & Kudla, 2011). De las 250 observadas en *Arabidopsis*, al menos 100 han sido categorizadas como proteínas de respuesta a calcio, y se encuentran representadas en cuatro familias: las calmodulinas (CaM), las proteínas similares a calmodulina (CML), las calcineurinas tipo B (CBL) y las proteínas cinasas dependientes de calcio (*CDPK*). De estas, la familia CaM está altamente conservada en distintos linajes de eucariontes, mientras que las proteínas CML, CBL y *CDPK* solamente se han encontrado en plantas y algunos linajes de protistas (Batistič & Kudla, 2012).

Las diferentes proteínas de unión a calcio de las plantas difieren en múltiples aspectos, entre los que se incluyen su afinidad por el ion, su localización subcelular, su concentración y su especificidad por ciertos blancos. Con esto, aumenta la complejidad de las rutas de señalización y la planta puede ejercer un control estricto sobre la respuesta a los cambios ambientales que experimenta (Hashimoto & Kudla, 2011). Además, estas proteínas se encuentran clasificadas de acuerdo con su función dentro de las rutas de señalización. Por un lado, las denominadas proteínas censoras-retransmisoras, incluyen a las CaM y a las CML. Estas proteínas son capaces de unirse a calcio, pero carecen de dominios efectores; por lo cual, su función es retransmitir la señal a otras proteínas blanco para regular su actividad y continuar así con la señalización (Kudla et al., 2010). Por otro lado, las *CDPK* son denominadas proteínas censoras-efectoras, ya que combinan en una misma proteína las funciones de unión a calcio y de respuesta, mediante un dominio de actividad cinasa capaz de continuar la señalización por sí misma (Kudla et al., 2010). Las proteínas CBL también son proteínas censoras-retransmisoras; sin embargo, estas interactúan específicamente con un conjunto de proteínas llamadas cinasas interactuantes con CBL (CIPK); por lo cual también pueden considerarse como sensores-efectores bimoleculares (Hashimoto & Kudla, 2011).

3.4. La familia *CDPK*

3.4.1. Estructura y regulación de las *CDPK*

Dentro de las proteínas de unión a calcio presentes en las plantas, las *CDPK* son particularmente interesantes, ya que combinan la función de detectar cambios en la concentración de calcio, con una función cinasa (Y. Yang et al., 2017). Todos los miembros de esta familia tienen una estructura conservada, que consiste en tres dominios. Primero, un dominio *N*-terminal variable; segundo, un dominio cinasa de serina/treonina y; tercero, un dominio de activación (Liese & Romeis, 2013). Este último a su vez, está compuesto por un segmento de pseudosustrato para el dominio cinasa y por un dominio de unión a calcio altamente homólogo con la calmodulina y formado canónicamente por cuatro motivos *EF-hand* organizados en dos lóbulos, uno *N*-terminal y otro *C*-terminal (Liese & Romeis, 2013). Los diferentes miembros de esta familia de proteínas en *A. thaliana* presentan altos niveles de similitud en sus dominios cinasa (60%-98%), en su dominio de activación (42%-100%) y en el de unión a calcio (50%-

98%); sin embargo, para el dominio N-terminal, los niveles de similitud entre parálogos son bajos (Cheng et al., 2002a).

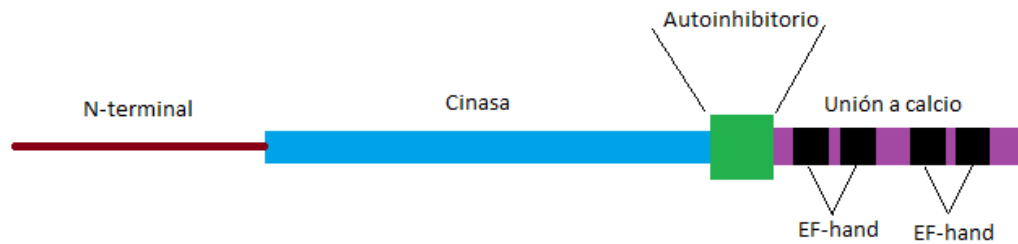


Figura 2. Estructura general de las proteínas CDPK

A partir de la estructura modular de las *CDPK*, se ha establecido que su funcionamiento está dado a partir de cambios conformacionales que se producen al unirse al ión de calcio, los cuales hacen que se libere el pseudosustrato del sitio activo del dominio cinasa (Liese & Romeis, 2013). Antes de la unión del ion, la región de pseudosustrato y el lóbulo *N*-terminal del dominio de unión a calcio, forman dos hélices alfa extendidas que bloquean el sitio activo de la cinasa. Sin embargo, esto no descarta que existan mecanismos regulatorios adicionales, tales como la fosforilación por otras *CDPK*, la unión a lípidos o la interacción con proteínas 14-3-3 (Liese & Romeis, 2013). Por ejemplo, en *A. thaliana*, las proteínas *CDPK11* y *CDPK24* funcionan en conjunto como reguladores negativos en el desarrollo del tubo polínico (Zhao et al., 2013). Sin embargo, solo la *CDPK11* presenta respuesta ante el calcio, mientras que la *CDPK24* solo se activa al ser fosforilada por la primera, y solo entonces se continúa con la fosforilación corriente abajo (Zhao et al., 2013). Adicionalmente, se ha encontrado que los dos lóbulos del dominio de unión a calcio tienen diferentes afinidades por este ion, sugiriendo que el lóbulo *C*-terminal tiene una mayor afinidad y que este se encuentra unido a iones de calcio en niveles basales de concentración (Liese & Romeis, 2013). Por otro lado, se han reportado proteínas *CDPK* con dominios de unión a calcio degenerados, particularmente en el primer motivo *EF-hand*. Estas son catalíticamente activas, pero no presentan una dependencia clara de calcio para su actividad (Liese & Romeis, 2013). En *A. thaliana*, estas *CDPK* con motivos *EF-hand* degenerados están especialmente representadas entre los miembros del grupo III de la familia génica (ver sección “Origen y evolución” para más información sobre los grupos filogenéticos de estas proteínas) (Liese & Romeis, 2013).

Otro aspecto de la regulación y funcionamiento de las *CDPK* está relacionado con el extremo *N*-terminal de estas proteínas. Esta región es altamente variable en secuencia y longitud entre los diferentes parálogos de una misma especie, y la mayoría de estas presentan sitios de *N*-miristoilación y *N*-palmitoilación, los cuales están relacionados con la asociación de estas proteínas a membranas (Schulz et al., 2013). La actividad de las *CDPK* se da tanto en el citosol como en membranas y se ha observado que la actividad de algunas de estas está adicionalmente regulada por la unión de fosfolípidos (Harmon et al., 2000). Sin embargo, en general, los fosfolípidos no funcionan como moléculas señalizadoras, por lo cual se ha sugerido que algunas *CDPK* tienen baja actividad citosólica y se activan al ser transportadas y asociadas a una membrana (Harmon et al., 2000). Por ejemplo, en *Solanum tuberosum*, la *NADPH*-oxidasa *RBOHB*, localizada en la membrana plasmática, es fosforilada por la *CDPK5*, por lo cual es necesario que esta se traslade correctamente hacia la membrana para realizar su función (Asai et al., 2013). De manera similar, en *Nicotiana benthamiana*, la mutación de los sitios de miristoilación del extremo *N*-terminal de la *CDPK2* no solo impide su correcta localización en la membrana, sino que evita su fosforilación por cinasas río arriba, lo cual es necesario para su correcta activación (Witte et al., 2010).

En relación con lo anterior, las *CDPK* presentan sitios de fosforilación a lo largo de toda su secuencia, y esta reacción puede ser catalizada tanto por la propia *CDPK* (autofosforilación) como por otras cinasas río arriba (Liese & Romeis, 2013). En el caso de la autofosforilación, se ha observado que puede ser tanto dependiente como independiente de Ca^{2+} (Liese & Romeis, 2013). La función precisa de las fosforilaciones en la regulación de la función de las *CDPK* no está del todo clara. Por ejemplo, en *Arachis hypogea* se ha visto que la fosforilación tiene un efecto de activación en algunas *CDPK*, mientras que en *Psophocarpus ragonolobus* se han observado efectos inhibitorios. En el primer caso, la fosforilación parece no tener efecto en la dependencia de calcio de la actividad de las *CDPK* (Cheng et al., 2002b). De manera similar, en el caso de la *CDPK2* de *N. benthamiana*, las variantes con sustituciones en los sitios de fosforilación tampoco presentan cambios en la actividad de fosforilación de su sustrato (Witte et al., 2010). En otras cinasas relacionadas con estrés, como las MAPK y las SnRK, la fosforilación dependiente de estímulo es necesaria para su activación (Vlad et al., 2009; K. Y. Yang et al., 2001). Sin embargo, este tipo de regulación no se ha encontrado para ninguna *CDPK* hasta el momento; por lo cual no está claro que la fosforilación juegue un papel directo en el control de

la actividad cinasa de las *CDPK* (Liese & Romeis, 2013). Sin embargo, es posible que las fosforilaciones jueguen un papel importante al promover cambios en la afinidad de las *CDPK* por el calcio, aumentando su sensibilidad, así como causando cambios estructurales que sean necesarios para acceder y secuestrar a sus proteínas blanco (Witte et al., 2010).

3.4.2. Las proteínas CDPK en la regulación de la respuesta a estrés y del desarrollo

Las proteínas *CDPK* tiene funciones que incluyen la señalización y respuesta, tanto a estrés biótico como abiótico, así como la señalización hormonal y de eventos del desarrollo vegetal (Schulz et al., 2013). Entre los principales tipos de estrés abiótico que experimentan las plantas, se encuentran la desecación y la salinidad, y se han reportado genes *CDPK* cuya expresión diferencial está asociada a estos estímulos (Singh et al., 2017). Entre los blancos de las proteínas *CDPK* relacionadas con la respuesta a estrés abiótico, se encuentran, principalmente, reguladores transcripcionales ABF (ABA Response Factor) y proteínas canal de iones y transportadores iónicos asociados a las células guarda que forman a los estomas (Schulz et al., 2013). Precisamente, la principal respuesta fisiológica de las plantas ante la sequía es la regulación de la apertura de los estomas, mientras que, ante la salinidad, se activa la expresión génica y acumulación de solutos compatibles, como la prolina (Singh et al., 2017; Verslues et al., 2006). Además, se ha reportado la expresión diferencial de genes *CDPK* bajo condiciones de estrés por altas y bajas temperaturas, sin embargo, en estos casos, el papel específico de estas proteínas no está del todo claro (Singh et al., 2017).

En cuando a la respuesta ante estrés biótico (en particular por la acción de patógenos), las reacciones de las proteínas *CDPK* incluyen cambios en la síntesis y señalización mediante fitohormonas, cambios en la expresión génica, así como la síntesis de EROS y muerte celular (Schulz et al., 2013). Por ejemplo, en *N. tabacum*, la expresión del gen *CDPK2* conduce a la acumulación de etileno y ácido jasmónico, lo cual a su vez promueve la expresión de genes relacionados con la defensa ante patógenos (Ludwig et al., 2005). Por otro lado, en *S. tuberosum*, la expresión de los genes *CDPK4* y *CDPK5* fosforila a la *NADPH* oxidasa *RBOHB*, con lo cual se promueve la producción de EROS y se desencadena la muerte celular como respuesta a la infección por el patógeno *Phytophthora infestans* (Kobayashi et al., 2007).

Finalmente, en cuanto los eventos del desarrollo en los que están involucradas las *CDPK*, uno de los más recientemente estudiados es el desarrollo del tubo polínico. Este proceso está caracterizado por un gradiente de oscilación en la concentración de iones de calcio (Schulz et al., 2013). Por ejemplo, en *A. thaliana* al menos 12 genes *CDPK* se expresan en el polen. De estos, se ha demostrado que los pares *CDPK17/CDPK34* y *CDPK11/CDPK24* tienen un rol específico en el desarrollo del tubo polínico y los dobles mutantes en estos genes presentan una menor adecuación relacionada con un bajo éxito en el crecimiento del tubo y en la fertilización del óvulo (Myers et al., 2009; Zhao et al., 2013). También en *A. thaliana*, se ha demostrado que las proteínas *CDPK* tienen un papel en el desarrollo floral, mediante la fosforilación del factor de transcripción *bZIP-FD*, el cual forma parte del complejo florígeno que se forma en el ápice del brote (Kawamoto et al., 2015). Otros procesos del desarrollo regulados por proteínas *CDPK* implican su papel en la regulación de la síntesis y señalización por fitohormonas, como es el caso de la proteína *CDPK1* en *N. tabacum*; la cual regula la biosíntesis de giberelinas mediante la fosforilación represiva del factor transcripcional *RSG* (REPRESSOR OF SHOOT GROWTH) (Schulz et al., 2013; Singh et al., 2017).

3.4.3. Origen y evolución de la familia CDPK

La teoría más aceptada sobre el origen de las *CDPK* sugiere que estas se formaron mediante la fusión de una proteína cinasa dependiente de calmodulina, con una calmodulina ancestral (Harmon et al., 2000). Debido a que las *CDPK* se encuentran presentes en todo el linaje de las plantas verdes (Viridiplantae, incluye a las algas verdes), así como en dos linajes de protistas, ciliados y apicomplejos, se ha propuesto que esta fusión se dio antes de la divergencia entre los linajes de plantas verdes y los protistas alveolados (X. S. Zhang & Choi, 2001).

El número de genes *CDPK* en plantas terrestres se encuentra entre 10 y 20 en algunos linajes menores y entre 30 y 40 en la mayoría de las angiospermas (Valmonte et al., 2014). Todos ellos se agrupan en cuatro grandes clados, dentro de los cuales están incluidos todos los genes (entre 23 y 25) de la licofita *Physcomitrella patens*; por lo cual se ha propuesto que cuatro es el número de genes *CDPK* en el ancestro de todas las plantas terrestres, y es a partir de estos que se dio la diversificación en diferentes linajes (Valmonte et al., 2014). Sin embargo, los genes *CDPK* de las plantas sin semilla tienden a formar sus propios grupos, ya sea dentro de alguno de los cuatro clados principales (por ejemplo, en los grupos I y IV), o en ramas independientes.

Estas últimas son, en estos casos, basales a las de algunos de los cuatro clados de las plantas con semilla, como ocurre en los clados II y III (Valmonte et al., 2014).

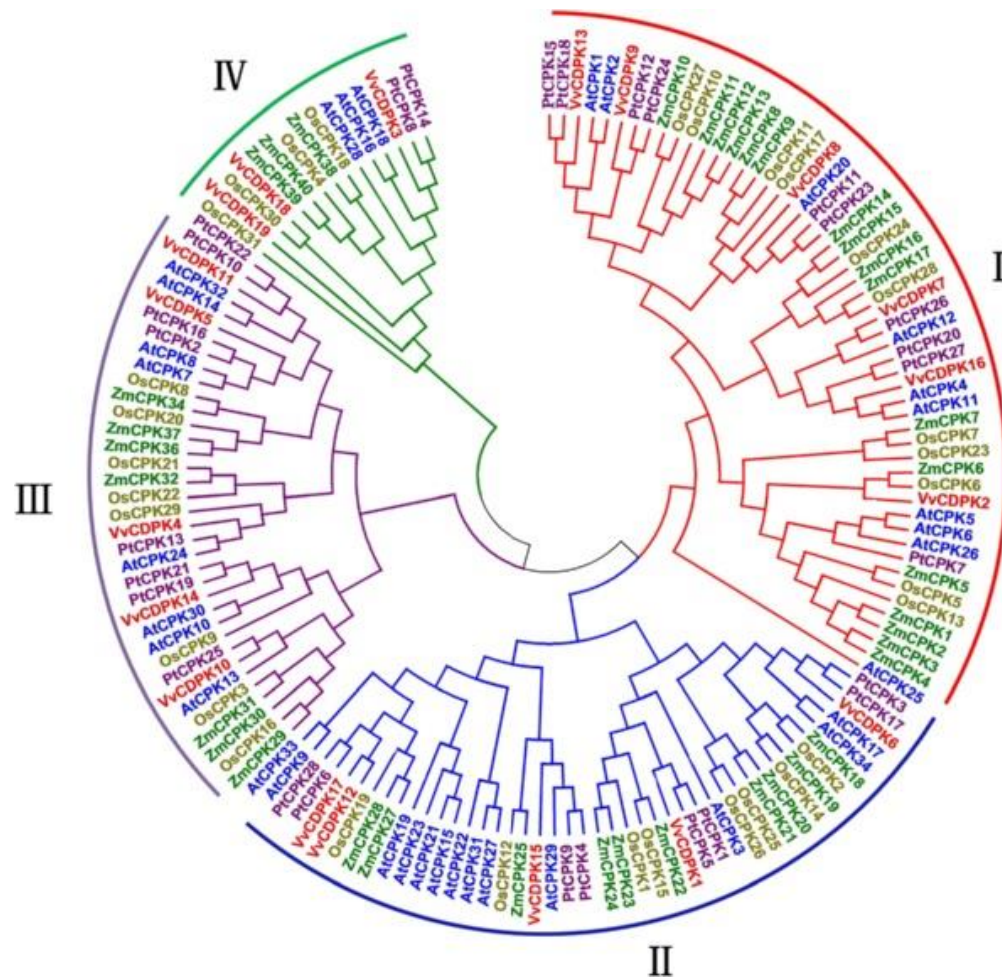


Figura 3. Filogenia de las CDPK de diversas especies de angiospermas. Los colores indican los genes de la misma especie. (*Vitis vinifera*, *Zea mays*, *Populus trichocarpa*, *Oryza sativa* y *Arabidopsis thaliana*), en la que se observan los cuatro clados principales de estas proteínas. Tomada de Zhang et al., 2015.

En angiospermas, los múltiples parálogos presentan altos niveles de homología y bajos niveles de diversificación, por lo cual es probable que se hayan originado a partir de eventos de duplicación recientes y particulares en ciertos linajes (Hamel et al., 2014). En varios casos, el número de parálogos es consistente con los eventos de duplicación genómica y poliploidización a lo largo de la historia evolutiva de las especies. Por ejemplo, entre las monocotiledóneas, *Zea mays* sp. *maiz* presenta el mayor número de CDPK, lo cual está probablemente asociado a dos duplicaciones genómicas: una en el linaje de los pastos y otra específica de esta especie (Paterson et al., 2010; Valmonte et al., 2014; Woodhouse et al., 2010). Similar a esto, entre las

eudicotiledóneas, *Glycine max* (soya) tiene el mayor número de genes *CDPK*, asociado también a varios ciclos de poliploidización reportados para esta especie (Schulz et al., 2013; Valmonte et al., 2014). En contraposición, *Vitis vinifera*, con 17 genes *CDPK*, es la eudicotiledónea estudiada hasta la fecha, con menor número de copias (de las que se han estudiado hasta el momento), lo cual probablemente se debe a que esta especie no ha experimentado duplicaciones genómicas desde la triplicación ancestral en el linaje de las eudicotiledóneas (Jaillon et al., 2007; Valmonte et al., 2014). Este mismo patrón puede observarse también en linajes de plantas sin semilla. Por ejemplo, *P. patens* tiene un número de genes *CDPK* cercano al de las angiospermas, a pesar de su estilo de vida relativamente más simple, lo cual puede explicarse dada la evidencia de un gran número de rearrreglos a gran escala que han ocurrido en su genoma (Hamel et al., 2014; Rensing et al., 2008). En contraste, la licofita *Selaginella moelendorffii* solamente presenta 10 genes *CDPK*, lo cual se relaciona con una probable ausencia de duplicaciones genómicas completas en esta especie (Banks et al., 2011; Hamel et al., 2014). Además de las duplicaciones genómicas completas, existe evidencia de genes *CDPK* generados mediante otros procesos de duplicación. Por ejemplo, existen duplicaciones segmentales en ocho pares en *A. thaliana* y nueve en *O. sativa*, mientras que cinco genes *CDPK* en *A. thaliana* y dos en *Populus trichocarpa* se encuentran arreglados en clústeres, lo cual sugiere que se originaron mediante duplicaciones en tándem (Singh et al., 2017).

Todos los genes *CDPK* de los linajes de algas verdes, los cuales cuentan con alrededor de 10 copias, se agrupan en cuatro grupos independientes de los de las plantas terrestres (Hamel et al., 2014). Esto sugiere que estos genes han evolucionado independientemente en ambos linajes, y en el caso particular de los de las plantas terrestres, es probable que la adaptación a las condiciones novedosas de la vida fuera del agua haya sido uno de los principales determinantes de su evolución y diversificación en este linaje (Hamel et al., 2014). Por lo tanto, el escenario más probable implica que los genes *CDPK* de las plantas terrestres partieron de genes ancestrales con funciones en el desarrollo de estructuras reproductivas y de mantenimiento de la homeostasis celular (funciones que comparten con las *CDPK* de las algas verdes). Posteriormente, mediante distintos eventos de duplicación, subfuncionalizaron en aspectos de desarrollo y de respuesta a estrés osmótico, o neofuncionalizaron hacia la respuesta a las condiciones de la vida terrestre, como la desecación, acceso reducido a agua y nutrientes, cambios abruptos de temperatura y mayor exposición a microorganismos (Valmonte et al., 2014). Si bien mucha de esta divergencia funcional entre los genes *CDPK* es

ancestral y compartida entre diversos linajes de plantas, muchas otras son recientes y específicas de especies individuales o cercanamente relacionadas, probablemente debido a las presiones ambientales específicas que estas experimentaron a lo largo de su evolución (Valmonte et al., 2014).

A pesar de los múltiples estudios que se han realizado de estos genes en diferentes especies de plantas, particularmente en angiospermas modelo y de interés comercial (uva, tomate, maíz, soya, trigo, etc), poco se ha investigado al respecto en otros grupos. En concreto, poco se sabe de esta familia génica en el linaje de las gimnospermas. El único dato disponible es el número aproximado de genes *CDPK* en el genoma de las coníferas, de alrededor de 12 en la especie *Picea sitchensis*, obtenido a partir de datos genómicos (Valmonte et al., 2014). Por lo tanto, aún está pendiente investigar a profundidad esta familia génica en este linaje. En particular, es de interés estudiar los procesos que han determinado la evolución de sus genes *CDPK*. Por un lado, considerando el tamaño y la dinámica del genoma de las coníferas, es importante determinar el número de genes *CDPK* que estas poseen y los procesos mediante los que se han generado, así como su relación evolutiva con los genes *CDPK* de otros linajes de plantas. Por otro lado, dada la función de los genes *CDPK*, es posible que estos hayan estado sujetos a distintas presiones selectivas y otros procesos evolutivos, por lo cual es importante estudiar el papel que ha tenido la selección natural en la evolución de estos genes en las coníferas, y en la adaptación de estas a los diferentes ambientes en los que habitan. Finalmente, si bien no se realizó en el presente trabajo, es importante analizar el papel funcional que los genes *CDPK* tienen en la respuesta de las coníferas a los diferentes tipos de estrés y estímulos ambientales que estas pueden experimentar en estos hábitats contrastantes y comparar esta respuesta entre distintos linajes de coníferas.

4. MÉTODO

4.1. Obtención e identificación de las secuencias

Se descargaron las secuencias anotadas como **CDPK** en el transcriptoma ensamblado y anotado de *Abies sachalinensis* (Todo Fir Gene; <http://plantomics.mind.meiji.ac.jp/todomatsu/index.html>), las cuales fueron identificadas de acuerdo con su anotación de origen en esta plataforma. Estas se utilizaron como *query* para realizar búsquedas con el algoritmo blastn, implementado en el programa BLAST+ (Camacho et al., 2009) sobre los transcriptomas de coníferas disponibles en la base de datos de NCBI. Se procuró tener representantes de ambientes boreales, templados y subtropicales para al menos un par de géneros e incluir especies de al menos tres familias de coníferas. La lista de las especies seleccionadas para esta búsqueda se encuentra en la Tabla 1, junto con el número de acceso de cada transcriptoma. El transcriptoma de *A. religiosa* se obtuvo a partir de datos generados para un análisis de expresión génica en poblaciones naturales sometidas a estrés por ozono (Reyes-Galindo, 2019). El RNAseq para estas muestras fue realizado en una línea de secuenciación Illumina Hi-Seq 4000 150PE. Las lecturas obtenidas (en promedio 33 millones por muestra) fueron analizadas y filtradas de acuerdo con su calidad (ver detalles en Reyes-Galindo, 2019). El ensamble se generó a partir de las lecturas limpias, para lo cual se utilizó Trinity v. 2.8.6 (Grabherr et al., 2011) con sus parámetros por defecto. El umbral (***e-value***) para las búsquedas BLAST fue de 10^{-50} y solo se retuvieron los cinco transcritos que mostraron la mayor afinidad para cada secuencia de referencia (*A. sachalinensis*) en cada especie.

Tabla 1. Especie y número de acceso de los transcriptomas utilizados para la búsqueda de las secuencias CDPK en coníferas. *El transcriptoma ensamblado de *A. religiosa* se generó a partir de datos de RNAseq previamente ensamblados por Reyes-Galindo, 2019. †Las secuencias de *A. sachalinensis* se obtuvieron de la base de datos TodoFirGene (<http://plantomics.mind.meiji.ac.jp/todomatsu/index.html>)

Especie	Familia	Número de acceso del transcriptoma (NCBI)	Tpo de ambiente
<i>Abies balsamea</i>	Pinacea	GGJG00000000.1	Boreal
<i>Abies sachalinensis</i>	Pinacea	NA†	Boreal
<i>Abies pinsapo</i>	Pinacea	GCZN00000000.1	Templado
<i>Abies religiosa</i>	Pinacea	NA*	Subtropical
<i>Larix laricina</i>	Pinacea	GGJA00000000.1	Boreal
<i>Picea glauca</i>	Pinacea	GGJI00000000.1	Boreal
<i>Pinus banksiana</i>	Pinacea	GGJH00000000.1	Boreal
<i>Pinus monticola</i>	Pinacea	GDQD00000000.1	Templado
<i>Pinus patula</i>	Pinacea	GECO00000000.1	Subtropical
<i>Pinus strobus</i>	Pinacea	GGJK00000000.1	Templado
<i>Pinus tecunumanii</i>	Pinacea	GFZL00000000.1	Subtropical
<i>Sequoia sempervirens</i>	Cupressaceae	GIBU00000000.1	Templado
<i>Taxus baccata</i>	Cupressaceae	GGKG00000000.1	Templado
<i>Thuja occidentalis</i>	Taxaceae	GGJL00000000.1	Boreal

Posteriormente se determinó el marco de lectura de cada transcrito retenido, utilizando la herramienta ORFfinder de NCBI; en cada caso, se buscó el marco de lectura más largo. Luego se eliminaron los transcritos para los que este fuera menor que 1000 bp.

A continuación, se evaluó la integridad de las secuencias retenidas. Para esto, se tradujeron las secuencias a aminoácidos, utilizando el paquete ape de R (Paradis & Schliep, 2019). Las secuencias traducidas fueron analizadas con el programa InterProScan (<https://www.ebi.ac.uk/interpro/>) para verificar que todas ellas presentaran los dominios cinaasa y de unión a calcio característicos de las proteínas **CDPK**. Para esto se analizaron cuatro perfiles distintos: Pfam, SMART, SUPERFAMILY y PROSITE.

Finalmente, se verificó si las secuencias retenidas sí correspondían a su respectivo gen **CDPK**. Para esto, se realizaron búsquedas BLAST de cada secuencia contra otras dos bases de datos de proteínas: la base Araport11, obtenida de The *Arabidopsis* Information Resource

(TAIR; <https://www.arabidopsis.org/index.jsp>), y la base revisada y anotada SwissProt, obtenida de <https://www.uniprot.org/>. Cada secuencia se anotó de acuerdo con su mejor *hit* en cada base de datos, así como con la identidad de la secuencia de búsqueda (*query*) original en el transcriptoma de *A. sachalinensis*. La anotación se tomó como válida cuando ambas coincidieron, o cuando al menos una coincidió y la otra obtuvo una anotación de un parálogo cercano. Finalmente, en algunos casos, la anotación se validó con base en su agrupamiento en las filogenias (ver más adelante).

4.2. Elaboración de matrices y alineamiento

Las secuencias filtradas y anotadas de cada especie fueron agrupadas en matrices individuales por especie; estas a su vez, se combinaron en una sola matriz multiespecie. Adicionalmente, se creó otra matriz multiespecie, que además de las secuencias **CDPK** de coníferas, contenía las secuencias ortólogas de *A. thaliana* y *P. trichocarpa*, de acuerdo con lo reportado para estas especies (Cheng et al., 2002a; Zuo et al., 2013). Las secuencias se obtuvieron respectivamente de TAIR y del “*Populus Genome Integrative Explorer*” (<https://popgenie.org/>). En esta segunda matriz, solo se incluyeron seis especies de coníferas: cuatro de la familia Pinaceae (*P. banksiana*, *P. patula*, *A. religiosa* y *A. balsamea*) y una de cada una de las familias Cupressaceae y Taxaceae (*T. occidentalis* y *Taxus baccata*, respectivamente). Esto debido a que el objetivo de esta matriz es simplemente generar una filogenia que represente las relaciones de los genes de las coníferas con respecto a los de las angiospermas.

Las dos matrices multiespecie fueron alineadas usando el paquete DECIPHER de Bioconductor en R (Wright, 2016), utilizando la opción AlignTranslation para conservar el marco de lectura de las secuencias. Posteriormente, estos alineamientos se tradujeron a aminoácidos utilizando MEGA X (Kumar et al., 2018).

4.3. Análisis filogenéticos

Primero se infirió una filogenia con cada una de las matrices, utilizando el método Neighbor-Joining (NJ) con el modelo TN93 (Tamura & Nei, 1993), en el paquete ape de R (Paradis & Schliep, 2019), usando los alineamientos de nucleótidos. A partir de esta filogenia inicial se identificaron secuencias que formaran grupos monofiléticos o parafiléticos de una misma especie y que tuvieran una misma anotación. Estas secuencias probablemente representaban transcritos alternativos y/o parciales de un mismo locus y no genes distintos, por lo cual, se

conservó solamente el transcrito con el menor *e-value* en los BLAST de identificación iniciales. Estas nuevas matrices multiespecie fueron re-alineadas y traducidas, utilizando el mismo método que para las matrices originales.

Estos nuevos alineamientos se usaron para construir filogenias bayesianas, con el programa MrBayes 3.2.7 (Ronquist et al., 2012). Para ambas filogenias, se utilizó una estrategia con modelos fijos mixtos, es decir, se permitió que las cadenas exploraran todos los modelos de aminoácidos disponibles en el programa. Al final de cada corrida, los modelos contribuyen al resultado en proporción a su probabilidad posterior. Se realizaron cadenas markovianas de 15 millones de generaciones (MCMC), muestreando cada 1000 árboles y descartando el 25% inicial de la muestra. En ambos casos, la convergencia de las cadenas se verificó utilizando Tracer 1.7.1 (Rambaut et al., 2018), y las filogenias fueron graficadas con el paquete ggtree de R (Yu et al., 2017).

4.4. Pruebas de selección y tasas evolutivas

4.4.1. Modelos evolutivos por codones

Para evaluar si existen diferencias en las tasas evolutivas entre los distintos genes *CDPK* de las coníferas y, en concreto, si hay evidencia de selección natural en alguno de estos genes entre las especies boreales, templadas y subtropicales, se estimaron los coeficientes dn/ds (sustituciones no sinónimas vs. sustituciones sinónimas). Para esto, se utilizaron varios de los modelos evolutivos por codones mediante máxima verosimilitud, implementados en el programa PAML 4 (Z. Yang, 2007).

4.4.2. Modelos por rama

Primero se probaron los modelos por rama (Z. Yang, 1998; Z. Yang & Nielsen, 1998). Estos permiten evaluar si existen ramas de la filogenia con coeficientes dn/ds significativamente diferentes de los del resto del árbol; estas representarían linajes cuyos genes han evolucionado bajo algún régimen de selección natural distinto. Para estos modelos se utilizaron el alineamiento y la filogenia de coníferas (sin los dos grupos de angiospermas). En primer lugar, se probó el modelo libre (modelo 1), el cual permite que cada rama de la filogenia tenga su propio valor dn/ds .

Luego, se probaron modelos por ramas únicamente en linajes con duplicaciones recientes exclusivas a alguna familia o género de coníferas; esto permitió evaluar si hubo algún

cambio en la tasa evolutiva después de dicha duplicación. Para esto se probaron tres escenarios sugeridos previamente (Bielawski & Yang, 2003). El primero utilizó dos coeficientes dn/ds : uno para todas las ramas anteriores a la duplicación, y otro para las ramas posteriores a esta. En el segundo se incluyeron tres coeficientes: uno para todas las ramas anteriores a la duplicación, y uno para cada grupo posterior a la duplicación (parálogos). En el tercero, se establecieron cuatro coeficientes: uno para todas las ramas anteriores a la duplicación, uno para ambas ramas inmediatamente posteriores a la duplicación, y dos más, uno para cada grupo formado dentro de estas ramas (parálogos).

Cada modelo fue probado estadísticamente contra un modelo nulo; es decir, un modelo con un solo coeficiente para todas las ramas de la filogenia. En cada caso, el *log-likelihood* del modelo alternativo se comparó contra el del modelo nulo mediante un *Likelihood Ratio Test* (LRT) con un número de grados de libertad igual a la diferencia de parámetros entre los modelos a comparar.

4.4.3. Modelos por sitio

Debido a que los modelos por ramas promedian los valores de los sitios en una rama determinada para calcular su coeficiente dn/ds , estos no son útiles para detectar los sitios específicos bajo selección. Por lo tanto, se utilizaron también modelos por sitio, los cuales permiten que los coeficientes dn/ds varíen entre los sitios a lo largo de las secuencias de los genes. Para esto, se ensamblaron y alinearon matrices multiespecie para cada gen individual, mediante el mismo método utilizado previamente. Estos se utilizaron para obtener árboles NJ con el modelo TN93 para cada alineamiento (Tamura & Nei, 1993).

Estas filogenias y sus respectivos alineamientos, se utilizaron para evaluar los modelos por sitio (Nielsen & Yang, 1998; Z. Yang et al., 2005) mediante dos pruebas. En la primera se comparó el modelo M1a (neutral) contra el modelo M2a (selección). El primero considera que una parte de los sitios es neutral, es decir, con coeficiente dn/ds igual a 1, y que la otra parte, está constreñida por selección negativa; es decir, con coeficientes dn/ds menores que 1. El segundo, presupone una tercera categoría de sitios: aquellos con coeficientes dn/ds mayores que 1, que se encontrarían bajo selección positiva (Nielsen & Yang, 1998; Yang et al., 2005).

La segunda prueba consistió en comparar los modelos M7 y M8 (Yang et al., 2005). El primero funciona de manera similar al modelo M1a, pero implementa una distribución beta para determinar los valores del coeficiente dn/ds de los sitios, dentro de un intervalo entre 0 y

1. Esto proporciona un modelo nulo más flexible para evaluar la acción de la selección natural en las secuencias (Yang & Swanson, 2002). El modelo M8 es una extensión del modelo M7, que permite una categoría adicional de sitios cuyo valor dn/ds es estimado a partir de los datos y es libre, con lo cual puede tener valores mayores que 1 (Z. Yang et al., 2005; Z. Yang & Swanson, 2002).

En ambos casos, se compararon los valores *log-likelihood* de los modelos nulos (M1a y M7) contra los de los modelos alternativos (M2a y M8), mediante un LRT con un número de grados de libertad igual a la diferencia de parámetros entre los modelos a comparar.

4.4.4. Sitios bajo selección episódica

Posteriormente se determinó si existen sitios que hayan evolucionado bajo selección en alguna rama o subconjunto de ramas, es decir, que estén sujetos a selección episódica. Para esto, se implementó el algoritmo MEME (Mixed Effects Models of Evolution) (Murrell et al., 2012) que forma parte del paquete HyPhy (Kosakovsky Pond et al., 2005). Este se basa en modelos filogenéticos de rama-sitio con efectos aleatorios, y permite que el valor del coeficiente dn/ds varíe entre sitios (efecto fijo) y entre ramas para un mismo sitio (efecto aleatorio). Esto le permite detectar efectos de selección únicamente en un conjunto de ramas, lo que contrasta con métodos tradicionales que asumen que la selección es constante a través del tiempo. Estos últimos suelen estimar coeficientes dn/ds menores que 1 cuando no hay suficientes sitios bajo selección, como es el caso en la mayoría de los genes codificantes (Murrell et al., 2012). Los modelos evolutivos de efectos mixtos se evaluaron para los datos individuales de cada gen, utilizando los mismos alineamientos y filogenias empleados para la evaluación de los modelos por sitio en CODEML. Se utilizó una significancia de $p < 0.05$ para presumir efectos selectivos en los sitios detectados.

A pesar del poder de MEME para detectar sitios bajo selección episódica, además de los sitios bajo selección recurrente, la capacidad para identificar sitios específicos bajo selección en una rama determinada (es decir, para detectar selección en un par rama-sitio) es estadísticamente limitada (Murrell et al., 2012). Esto se debe a que la información utilizada para realizar dicha inferencia proviene de la variación a lo largo de una rama en un solo sitio (Murrell et al., 2012). Por lo tanto, para identificar las ramas sujetas a selección, se evaluaron los modelos evolutivos aBSREL (por sus siglas en inglés *adaptive branch-site random effects likelihood*)

(Smith et al., 2015) que también están disponibles en HyPhy. Este método detecta las ramas bajo selección al inferir, para cada rama, el número óptimo de categorías de tasas evolutivas (i.e. sitios constreñidos, neutrales y bajo selección positiva) y la proporción de sitios dentro de cada categoría (Smith et al., 2015). De esta manera, aBSREL considera que la acción de la selección natural es heterogénea a lo largo de la filogenia y entre los sitios de un alineamiento. Mediante un LRT se determina, para cada rama, si el modelo alternativo que permite una determinada categoría de sitios bajo selección, se ajusta mejor a los datos que el modelo nulo, que no permite dicha categoría. Los modelos de aBSREL se probaron con el mismo set de datos que los modelos por sitio y que MEME, sin una selección de ramas *a priori*, por lo cual se aplicó una corrección de Holm-Bonferroni para pruebas múltiples a los valores de p del LRT de cada rama.

4.4.5. Divergencia funcional entre genes duplicados y efectos deletéros de las sustituciones

Se buscó evaluar si los genes **CDPK** de duplicación reciente (i.e. exclusivos de una familia o género) de las coníferas han estado bajo procesos de divergencia funcional. Para esto, se utilizó el programa DIVERGE 3, que permite evaluar la ocurrencia de dos tipos de divergencia funcional: tipo I y tipo II (Gu et al., 2013). El principio de estos análisis es que si un aminoácido es funcionalmente importante, se encontrará en un estado altamente conservado entre secuencias. Por lo tanto, un cambio en el estado de conservación de un residuo puede indicar un proceso de divergencia funcional (Gu, 2006). En el caso de la divergencia funcional tipo I, se postula que el cambio funcional entre dos duplicados resulta en cambios en las restricciones evolutivas en determinados sitios y se presupone que los sitios con mayor restricción funcional tienen menores tasas evolutivas. Por lo tanto, un relajamiento en dichas restricciones se reflejaría en una mayor tasa evolutiva, lo cual es una indicación de divergencia funcional. El patrón esperado en este caso es el de un sitio altamente conservado en uno de los clústeres duplicados, pero altamente variable en el otro (Gu, 1999; Gu et al., 2013). En contraste, en la divergencia funcional tipo II, una rápida acción de la selección natural lleva a la fijación de un aminoácido de características fisicoquímicas distintas en uno de los clústeres duplicados, llevando a un estado altamente conservado (pero distinto) en ambos clústeres (Gu, 2006). Debido a los patrones de divergencia encontrados entre los genes CDPK de las coníferas, solo fue posible analizar la ocurrencia de divergencia funcional tipo I. Para esto, se utilizó la filogenia completa de los genes **CDPK** de las coníferas y un alineamiento traducido a aminoácidos; dicha

traducción se realizó en MEGA X (Kumar et al., 2018). Para estimar los coeficientes de divergencia funcional tipo I, se utilizó el método Gu99 (Gu, 1999) y su significancia se determinó mediante un LRT con un grado de libertad a partir de los valores de ML, de acuerdo con lo sugerido en el manual del programa.

Posteriormente, con el objetivo de determinar los efectos de las sustituciones no sinónimas en los genes duplicados, se utilizó el programa PROVEAN en su versión de servidor en línea (Choi & Chan, 2015). Se evaluaron todas las sustituciones no sinónimas detectadas por MEME en el gen *CDPK34* en todos los linajes de coníferas, y se usó como secuencia de referencia la de *T. occidentalis*. Para determinar el carácter neutral o deletéreo de las sustituciones, se utilizó el umbral por defecto de -2.5. Es decir, las sustituciones con un umbral mayor que ese valor, se consideraron como neutrales. Posteriormente, se intentó identificar la ubicación aproximada de las sustituciones analizadas en la secuencia del gen; es decir, si se encuentran en el dominio cinasa, de unión a calcio, o alguna otra región.

4.5. Localización relativa de los genes *CDPK* en mapas de ligamiento de coníferas

Se trató de ubicar la posición relativa (en unidades de recombinación: cM) de los genes ***CDPK*** en los mapas de ligamiento disponibles para coníferas. En particular, era de interés identificar la posición de los genes duplicados, ya que esto proporcionaría información sobre el mecanismo mediante el cual se generaron las copias. Si las copias se generaron mediante poliploidía, estas se ubicarían en grupos de ligamiento equivalentes, y con un orden conservado. Si se generaron mediante duplicaciones segmentales, las copias presentarían un orden conservado, pero no se encontrarían en grupos de ligamiento correspondientes. Finalmente, si las copias se encuentran juntas sería indicación de duplicación en tándem, mientras que si se encuentran dispersas, podría tratarse de duplicaciones dispersas o mediadas por transposones (Freeling, 2009).

Se buscaron las secuencias ***CDPK*** retenidas en los tres mapas de ligamiento más densos reportados hasta la fecha: *Picea glauca* (Pavy et al., 2017), *Picea abies* (Bernhardsson et al., 2019) y un mapa compuesto para la familia Pinaceae (Pinaceae composite). Este último se generó a partir de la comparación de mapas de *Pinus* y *Picea* (de Miguel et al., 2015). Primero, se identificaron las secuencias de genes ***CDPK*** mapeadas mediante un BLAST, utilizando como referencia los genes de *A. sachalinensis*. Las secuencias retenidas se utilizaron para un segundo

BLAST, esta vez contra la base de proteínas Araport11, para verificar su identidad. Finalmente, se anotó el grupo de ligamiento y la posición en cM dentro de dicho grupo para cada secuencia retenida en cada mapa. La correspondencia entre los grupos de ligamiento de los diferentes mapas se anotó de acuerdo con lo reportado para cada uno (De Miguel et al., 2015). Esta información se utilizó para determinar si los genes se encuentran en grupos de ligamiento y posiciones equivalentes, o si su localización es distinta.

Nota: Los scripts, archivos de configuración y documentación adicional para realizar todos los análisis y figuras presentados en este trabajo están disponibles en el siguiente repositorio de GitHub: https://github.com/MickAmest/conifer_cdpk_repo.

5. RESULTADOS

5.1. Número y características de las secuencias CDPK obtenidas

Se obtuvieron secuencias **CDPK** para las 14 especies estudiadas. Estas especies incluyeron 11 de la familia Pinaceae, dos de la Cupressaceae y una de la familia Taxaceae. El número de secuencias obtenidas después de aplicar los filtros indicados en el método (p. ej. longitud mínima del marco de lectura igual a 1000 pb y presencia íntegra de los dominios cinasa y de unión a calcio) fue de un mínimo de siete (*P. monticola* y *Taxus baccata*) y un máximo de 16 (*A. balsamea* y *P. banksiana*; ver tabla 2; también ver Anexo 1 para el detalle de los transcritos totales recuperados para cada especie).

Tabla 2. Número de secuencias CDPK obtenidas para 14 especies de coníferas a partir de búsquedas BLAST realizados sobre transcriptomas ensamblados y después de aplicar los filtros de longitud e integridad de las secuencias (marco de lectura mayor que 1000 bp y presencia de los dominios cinasa y EF-hand).

Especie	Número de secuencias CDPK
<i>Abies balsamea</i>	16
<i>Abies sachalinensis</i>	14
<i>Abies pinsapo</i>	9
<i>Abies religiosa</i>	10
<i>Larix laricina</i>	10
<i>Picea glauca</i>	10
<i>Pinus banksiana</i>	16
<i>Pinus monticola</i>	7
<i>Pinus patula</i>	14
<i>Pinus strobus</i>	8
<i>Pinus tecunumanii</i>	13
<i>Sequoia sempervirens</i>	9
<i>Taxus baccata</i>	7
<i>Thuja occidentalis</i>	11

La variación en la longitud de los transcritos retenidos estuvo principalmente dada por la presencia, ausencia o longitud variable del extremo *N-terminal* de las secuencias. Por ejemplo, para un mismo gen utilizado como referencia (*query*), los diferentes transcritos en la búsqueda BLAST eran similares o iguales en la región de los dominios cinasa y de unión a calcio, pero diferían en la presencia o ausencia del extremo *N-terminal*. En estos casos, solo se retuvo la variante más larga y con el menor *e-value*.

De acuerdo con el análisis de InterProScan, se encontraron secuencias con dominios de unión a calcio alterados, caracterizados por la falta de uno o dos de los sitios de unión a calcio dentro de los dominios **EF-hands** (Figura 1). Esto se encontró sobre todo para los genes *CDPK10*

y *CDPK13* de las diferentes especies, lo cual sugiere que se trata de una característica estructural propia de estos dos genes en coníferas (ver Anexo 1). Estudios previos sugieren que estos genes *CDPK* atípicos, con dominios de unión a calcio degenerados, se encuentran en todos los linajes de plantas verdes, y representan alrededor del 21% de los genes *CDPK* en angiospermas (Hamel et al., 2014). Por lo tanto, estas secuencias fueron conservadas para los análisis posteriores.

Los BLAST realizados contra las bases de datos de proteínas Araport11 y SwissProt revelaron que, en todos los casos, las secuencias tuvieron como mejores *hits* a su gen correspondiente (de acuerdo con la referencia original en *A. sachalinensis*) o a algún parálogo cercano. Por ejemplo, las secuencias obtenidas a partir de la secuencia *CDPK10* de *A. sachalinensis*, tuvieron como mejor *hit* al gen *CDPK10* de *A. thaliana*, o en algunos casos, al *CDPK30*, su parálogo más cercano. Lo anterior indica que la identificación y anotación de las secuencias *CDPK* es confiable.

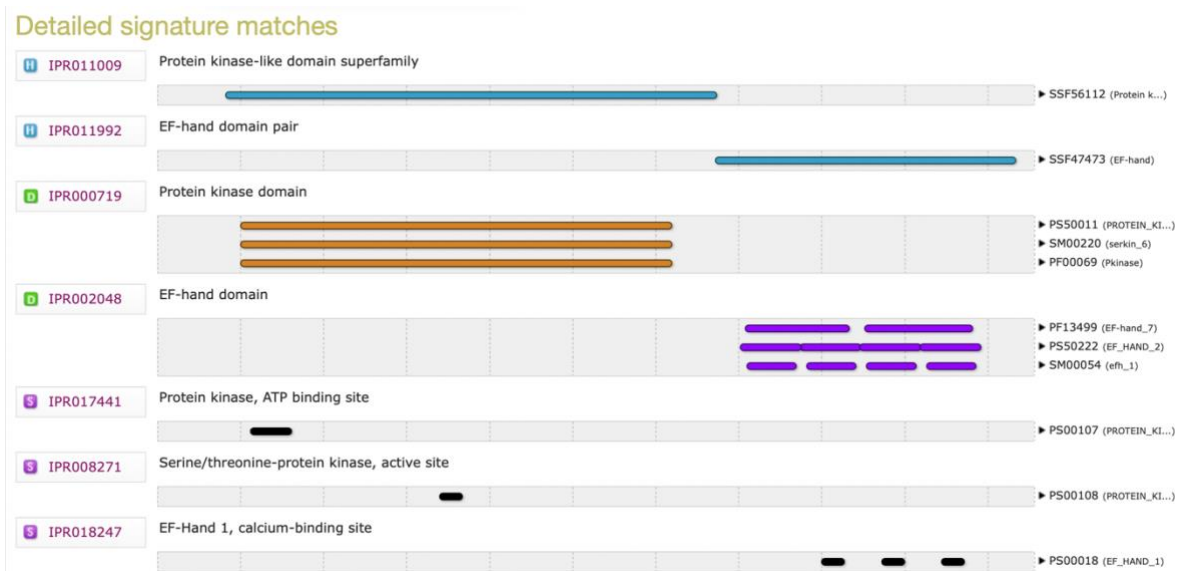


Figura 4. Ejemplo de un gen (*CDPK10* de *A. religiosa*) con dominio EF-hand degenerado, con tres sitios de unión a calcio, según el análisis de InterPro Scan. En azul se indican el dominio cinasa (arriba izq.) y el de unión a calcio (abajo der.) de acuerdo con el perfil de SUPERFAMILY. En anaranjado se indica el dominio cinasa de acuerdo con los perfiles de PROSITE, Pfam y SMART (PS, PF, SM). En morado se indica el dominio de unión a calcio (dos pares EF-hand) de acuerdo con los perfiles de Pfam, PROSITE y SMART. En negro se indican los patrones de PROSITE dentro de cada dominio: el sitio de unión a ATP y el sitio activo del dominio cinasa; y los sitios de unión a calcio de los dominios EF-hand. Estos últimos son solo tres, lo que sugiere que uno de estos dominios está degenerado (no tiene sitio de unión a calcio).

5.2. Filogenia de la familia CDPK en coníferas

La filogenia bayesiana mostró que los genes **CDPK** de las coníferas se distribuyen dentro de los cuatro clados previamente reportados para otras plantas terrestres. Dentro de cada clado, se identificaron entre 2 y 3 parálogos en coníferas, los cuales se agruparon entre sí y junto con sus respectivos ortólogos en angiospermas (Figura 2). Los genes *CDPK1*, *CDPK2* y *CDPK5* de las coníferas se agruparon en el clado I de la filogenia, mientras que los genes *CDPK17* y *CDPK34* lo hicieron en el clado II. En este último, se encontró una copia adicional en la familia Pinaceae que no se encontró en las familias Cupressaceae ni Taxaceae. En las Pináceas, el gen *CDPK6* aparece como una copia del gen *CDPK34*, lo cual sugiere un evento de duplicación específico para esta familia. La nomenclatura utilizada para los genes **CDPK** de coníferas se basó en las anotaciones iniciales del transcriptoma de *A. sachalinensis*, por lo que es importante señalar que el gen *CDPK6* de coníferas no tiene una relación evolutiva cercana con el gen *CDPK6* de *A. thaliana*, el cual, en esta especie, es parálogo del *CDPK5*, ambos ubicados dentro del grupo I. Finalmente, en el clado III se agruparon los genes *CDPK7*, *CDPK10* y *CDPK13*, mientras que en el IV, que es el grupo hermano de todos los demás de la familia **CDPK**, los genes *CDPK16* y *CDPK28* de las coníferas se agruparon junto con estos mismos genes de *A. thaliana*. En general, las relaciones dentro de los clados de cada gen concordaron con la taxonomía de las especies. Sin embargo, se observaron excepciones en la filogenia exclusiva de coníferas, en especial en las secuencias de *P. glauca* y *L. laricina*. Estas formaron un grupo monofilético en algunos casos (*CDPK17*) y, en otros, encontraron separadas (*CDPK10* y *CDPK13*). Además, hubo casos en los que las ramas terminales no fueron resueltas (p. ej. *CDPK1* y *CDPK5* en pinos).

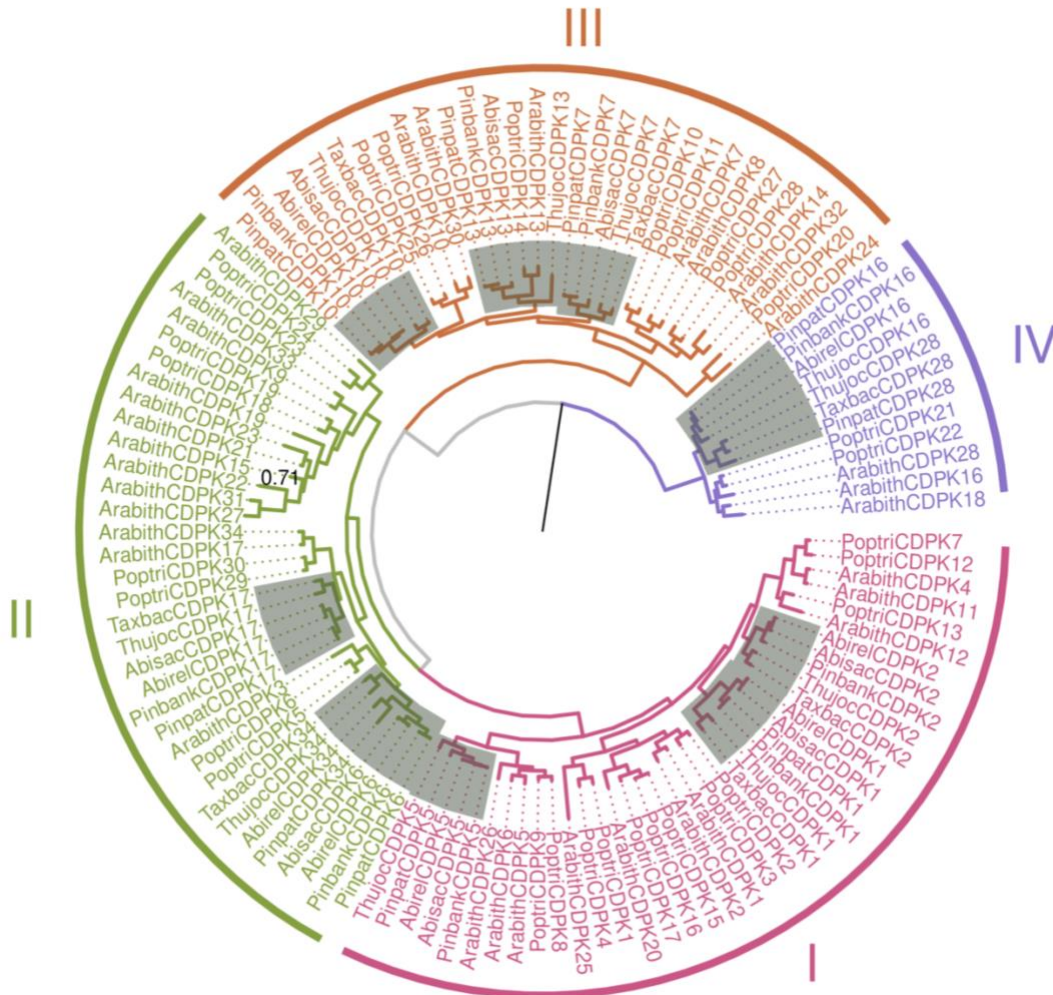


Figura 5. Filogenia bayesiana de la familia CDPK incluyendo secuencias de coníferas y angiospermas. En gris se resaltan los clados que contienen genes de coníferas exclusivamente. Los colores indican los cuatro grupos que componen a la familia CDPK en plantas terrestres. Se incluyeron dos especies de angiospermas: *Arabidopsis thaliana* (*Arabith*) y *Populus trichocarpa* (*Poptri*); cuatro especies de la familia Pinaceae: *Abies sachalinensis* (*Abisac*) y *A. religiosa* (*Abirel*), *Pinus banksiana* (*Pinbank*) y *P. patula* (*Pinpat*); y una de cada una de las familias Cupressaceae y Taxaceae: *Thuja occidentalis* (*Thujoc*) y *Taxus baccata* (*Taxbac*), respectivamente. Se indican los valores de soporte de rama solo en los casos en los que la probabilidad posterior fue menor que 0.9.

5.3. Tasas evolutivas y pruebas de selección

Los modelos por rama mostraron un aumento significativo en los coeficientes dn/ds para el clado que contiene la duplicación *CDK6-CDPK34*. De los tres modelos probados, el que mejor explicó los datos fue el de cuatro coeficientes. Según este modelo, los mayores valores de dn/ds corresponden al gen *CDPK34*, mientras que el *CDPK6* presenta coeficientes menores (Figura 3) lo que sugiere relajamiento en la selección purificadora en una de las copias luego de la duplicación.

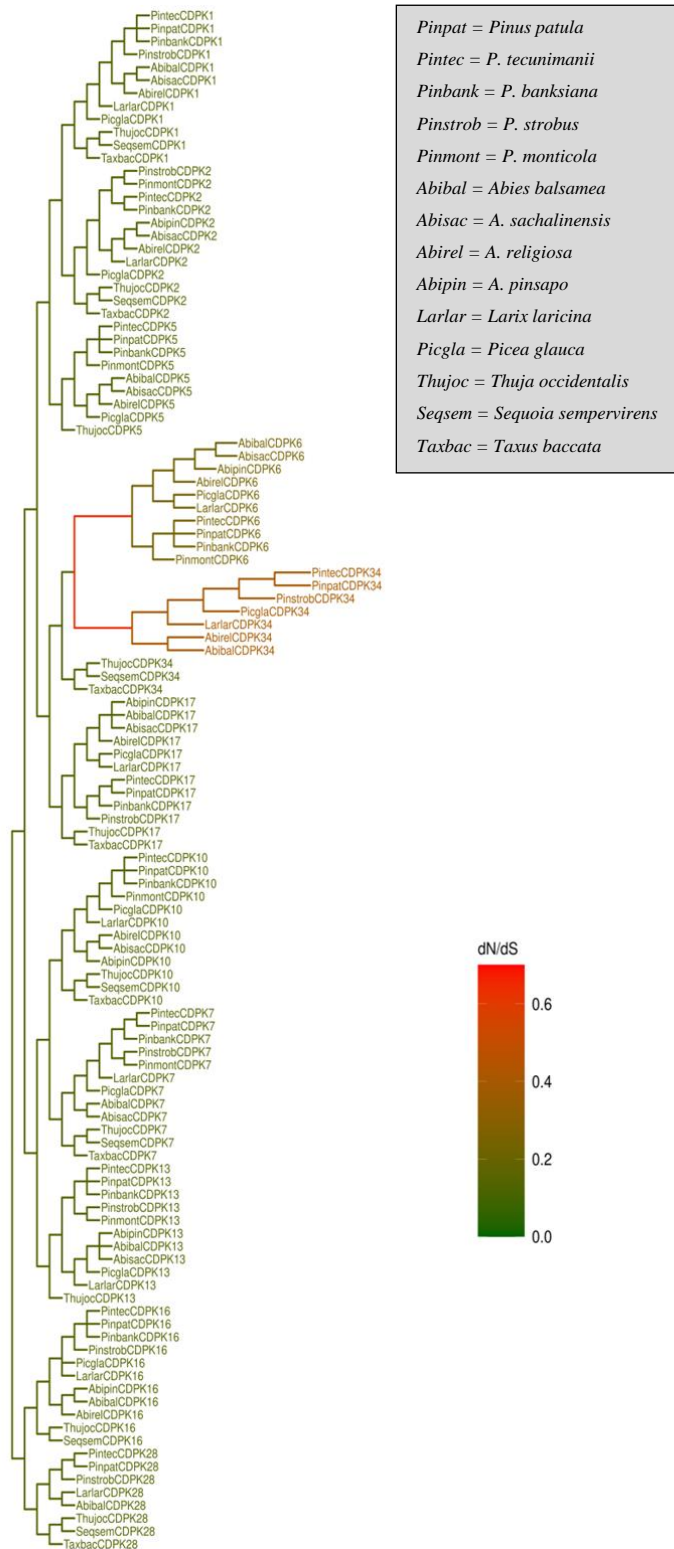


Figura 6. Filogenia de los genes CDPK de coníferas con longitud de rama reescalada y coloreada de acuerdo con sus coeficientes dN/dS , calculados con los modelos por rama de CODEML. El modelo de 4 coeficientes resultó el más adecuado sobre los modelos nulo, de dos, y de tres coeficientes ($p < 0.05$ en todas las comparaciones). Se observan mayores valores dN/dS después de la duplicación CDPK6-CDPK34 en la familia Pinaceae.

Los modelos por sitio detectaron huellas de selección positiva en siete genes: *CDPK2*, *CDPK7*, *CDPK13*, *CDPK16*, *CDPK28* y *CDPK34* (Tabla 3). En todos los casos, excepto *CDPK16*, el LRT resultó positivo para los dos pares de modelos comparados (M1a vs M2a y M7 vs M); para *CDPK16*, solo la segunda comparación resultó significativa. La prueba de Bayes Empirical Bayes (BEB), utilizada para detectar los sitios específicos que estarían bajo selección positiva, detectó además entre uno y seis sitios candidatos bajo selección para la mayoría de estas secuencias. Sin embargo, para los genes *CDPK16* y *CDPK34*, el análisis BEB no detectó sitios candidatos con ninguno de los modelos alternativos, a pesar de que ambos tuvieron un LRT significativo (Tabla 3).

Los Modelos Evolutivos de Efectos Mixtos (MEME) encontraron sitios bajo selección en todos los genes analizados, excepto *CDPK16* (Tabla 3). El número de sitios detectados por gen varió entre dos (*CDPK5*) y diez (*CDPK34*). Algunos de los sitios encontrados con los modelos de CODEML también se detectaron con MEME. En general, estos presentaron evidencia de selección en varias ramas a lo largo de la filogenia, indicando una selección constante a lo largo de la misma. Sin embargo, muchos sitios solo se detectaron con MEME (Tabla 3); estos presentaron evidencia de selección en una o pocas ramas específicas, indicando selección episódica en linajes particulares (ver Anexo 2).

Tabla 3. Sitios candidatos a estar bajo selección en once parálogos de la familia CDPK en coníferas. Se compararon dos modelos de selección (M2a y M8) contra dos modelos neutros (M1a y M7) en CODEML. *Indica los casos en los que el LRT de cada comparación resultó positivo ($p < 0.05$). En el análisis de CODEML, los sitios se detectaron mediante un análisis BEB. NA indica los casos en los que este análisis no detectó sitios bajo selección a pesar de que el LRT entre los modelos comparados fue significativo. En MEME, los sitios se detectaron con una significancia de $p < 0.05$. †Indica los sitios detectados con MEME que también se detectaron con al menos una de las comparaciones en CODEML. Los sitios se indican de acuerdo con la posición del codón en el alineamiento.

Gen	Modelos comparados (CODEML)	LRT de las comparaciones entre pares de modelos. (CODEML)	Sitios bajo selección (CODEML)	Sitios bajo selección (MEME)
CDPK1	M1a vs. M2a	0	-	56, 65, 121, 508, 530
	M7 vs. M8	4.4	-	
CDPK2	M1a vs. M2a	12.1*	48	4, 48†, 88, 89, 217
	M7 vs. M8	17.61*	29, 48, 49	
CDPK5	M1a vs. M2a	0	-	90, 262
	M7 vs. M8	0.22	-	
CDPK6	M1a vs. M2a	1.62	-	20, 30, 286
	M7 vs. M8	3.25	-	
CDPK7	M1a vs. M2a	19.91*	35, 52, 279	48, 52†, 214, 279†, 397, 468
	M7 vs. M8	34.13*	35, 38, 45, 46, 52, 279	
CDPK10	M1a vs. M2a	0	-	521, 522, 525, 526, 527
	M7 vs. M8	0.35	-	
CDPK13	M1a vs. M2a	9.23*	NA	9†, 24, 111, 308, 359, 372, 434
	M7 vs. M8	16.95*	9,12	
CDPK16	M1a vs. M2a	0	-	-
	M7 vs. M8	9.62*	-	
CDPK17	M1a vs. M2a	0	-	213, 232, 523, 524
	M7 vs. M8	5.37	-	
CDPK28	M1a vs. M2a	6.25*	NA	7, 90, 144
	M7 vs. M8	21.29*	33	
CDPK34	M1a vs. M2a	10.68*	NA	11, 12, 13, 87, 102, 277, 287, 289, 301, 499
	M7 vs. M8	22.02*	NA	

Entre los sitios detectados, destacan los de los genes *CDPK5*, *CDPK6*, *CDPK10*, *CDPK17* y *CDPK34*. En los primeros cuatro casos, porque los sitios presentaron una señal de selección en una o pocas ramas, sugiriendo selección en linajes particulares. Por ejemplo, el gen *CDPK6* presentó tres sitios bajo selección, dos de los cuales fueron exclusivos de la rama que lleva a *A. religiosa*, y para el gen *CDPK17*, dos de los cuatro sitios encontrados solo se observaron para la rama de *P. banksiana*. Igualmente, en el gen *CDPK10* los cinco sitios encontrados estuvieron en la rama de *P. patula*, y los dos sitios detectados en el *CDPK5* estuvieron en la rama que contiene las dos especies boreales de *Abies*. Por otro lado, el gen *CDPK34* presentó diez sitios bajo selección, aunque no se observó ningún patrón específico en la distribución de las señales dentro de la filogenia (ver Anexo 2). Todas estas señales putativas de selección fueron confirmadas de manera independiente por el análisis con aBSREL, excepto las del gen *CDPK5*, para el que no se observó nada significativo después de la corrección de Holm-Bonferroni. Sin embargo, la rama del gen *CDPK6* que contiene a *A. religiosa*, la del *CDPK17* de *P. banksiana* y la del *CDPK10* de *P. patula* mostraron nuevamente señal significativa de selección.

5.4. Divergencia funcional entre genes duplicados y efectos deletéreos de las sustituciones.

El análisis en DIVERGE indicó que hay divergencia funcional tipo I entre el par *CDPK6-CDPK34* ($\Theta=0.31$). Esta señal persistió al comparar ambos genes usando *CDPK17* como grupo externo. Sin embargo, en este caso se observó un patrón asimétrico, indicando mayor divergencia funcional para el gen *CDK6* que para el *CDPK34* ($\Theta_{CDPK6-CDPK17}=0.46$; $\Theta_{CDPK34-CDPK17}=0.43$). Al comparar los perfiles de los sitios individuales con un umbral de probabilidad posterior de 0.8, se encontró que solamente un sitio estaría involucrado en la divergencia funcional de estos genes; aunque se infirió un segundo sitio candidato al comparar *CDPK6* con *CDPK17* (Figura 4). Si se tienen en cuenta todos los sitios con probabilidad posterior mayor que 0.6, el mayor número de sitios se encontró para la divergencia funcional entre los genes *CDPK6* y *CDPK17* (17 sitios), seguido por la comparación entre *CDPK34* y *CDPK17* (13 sitios) y *CDPK6* y *CDPK34* (6 sitios). La posición de estos sitios a lo largo del alineamiento fue variable, sin evidencia de estar acumulados en alguna región específica (Figura 4).

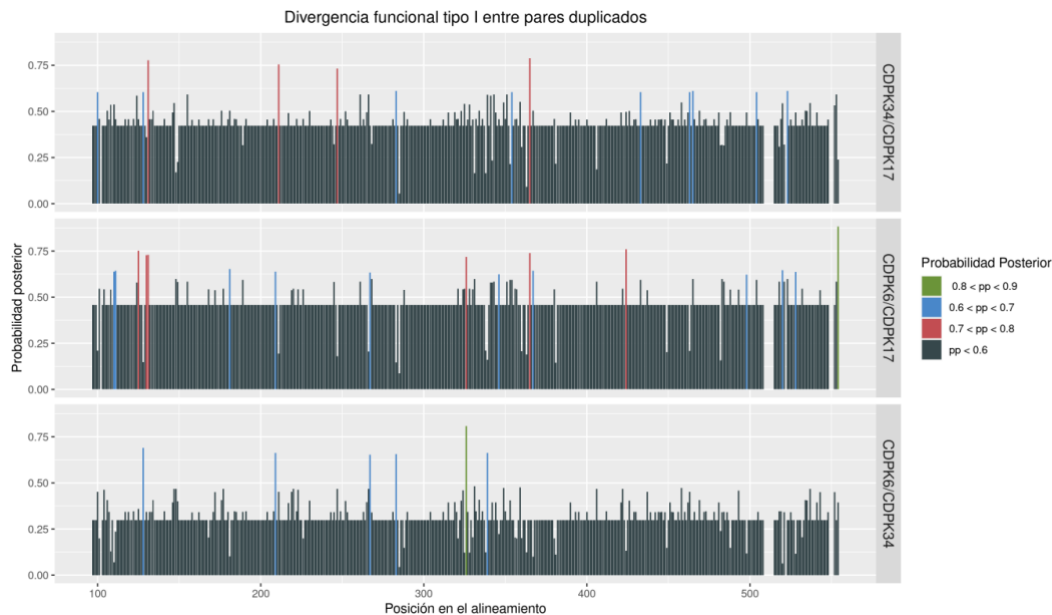


Figura 7. Perfiles específicos de cada sitio, de acuerdo con su probabilidad de estar involucrados en la divergencia funcional tipo I entre los genes CDPK duplicados en la familia Pinaceae. Se compararon los duplicados *CDPK6* y *CDPK34* entre sí, y cada uno con el *CDPK17* como grupo externo, utilizando el método Gu99 implementado en DIVERGE 3 para detectar divergencia funcional tipo I. Se analizaron las secuencias de todas las especies de coníferas estudiadas en este trabajo (11 de la familia Pinaceae, dos de la Cupressaceae y una de la Taxaceae). Los colores indican las diferentes categorías de probabilidad posterior para los sitios. El mayor número de sitios con probabilidad posterior mayor a 0.6, se encontró para la divergencia funcional entre los genes *CDPK6* y *CDPK17* (17 sitios), seguido por la comparación entre *CDPK34* y *CDPK17* (13 sitios) y *CDPK6* y *CDPK34* (6 sitios). Los sitios se indican según la posición del aminoácido en el alineamiento.

Por otro lado, al análisis en PROVEAN incluyó 23 sustituciones. De estas, 18 fueron clasificadas como neutrales, mientras que cinco tuvieron efectos deletéreos sobre los genes. Estas últimas no estuvieron distribuidas aleatoriamente a lo largo de las secuencias, sino que la mayoría se encontró acumulada hacia el final del dominio cinasa, mientras que una se encontró dentro del dominio autorregulatorio.

5.5. Localización de los genes CDPK en mapas de ligamiento de las coníferas

Se logró ubicar nueve genes *CDPK* en los mapas de ligamiento consultados: *CDPK1*, *CDPK2*, *CDPK6*, *CDPK7*, *CDPK10*, *CDPK13*, *CDPK16*, *CDPK17* y *CDPK34*. De manera individual, se encontraron cinco genes en el mapa compuesto para las pináceas, siete en el de *Picea abies* y tres en el de *P. glauca*. El análisis de correspondencia entre los grupos de ligamiento de cada mapa indicó que el grupo de ligamiento 5 del mapa de *P. abies*, corresponde al grupo de ligamiento 1 en los mapas de *P. glauca* y compuesto. Este grupo contiene a los genes duplicados *CDPK6* y *CDPK34*, que además parecen ser adyacentes (tabla 4). En ese mismo grupo de ligamiento, también se localizaron los genes *CDPK10* y *CDPK13*, aunque a distancias mucho mayores. El gen *CDPK17* solamente se identificó en el mapa compuesto para las pináceas, dentro del grupo de ligamiento 3, que es distinto al de los genes *CDPK34* y *CDPK6*. Los genes *CDPK10*, *CDPK16* y *CDPK34* también se encontraron en un solo mapa cada uno. Los genes *CDPK1* y *CDPK7* se encontraron en los mapas de *P. abies* y compuesto para las pináceas, en grupos de ligamiento equivalentes, pero en posiciones distintas en cada uno. El gen *CDPK13* fue el único que se encontró en los tres mapas, de igual manera en grupos equivalentes. Finalmente, el gen *CDPK2* se encontró en los mapas de *P. abies* y *P. glauca*, pero en grupos de ligamiento no equivalentes (tabla 4).

Tabla 4. Localización de los genes CDPK en tres mapas de ligamiento disponibles para conífera. Se indican el grupo de ligamiento (LG) y la posición relativa (en cM) dentro del grupo (POS). *Indica los casos en que los genes se ubican en grupos de ligamiento equivalentes entre mapas.

Gen	Pinaceae composite		<i>Picea abies</i>		<i>Picea glauca</i>	
	LG	POS	LG	POS	LG	POS
CDPK1	6*	64.5	10*	176.15	-	-
CDPK2	-	-	11	123.79	9	91.4
CDPK5	-	-	-	-	-	-
CDPK6	1*	15	5*	18.06	-	-
CDPK7	12*	115.1	9*	195.66	-	-
CDPK10	-	-	5	174.99	-	-
CDPK13	1*	72.2	5*	132.1	1*	98.16
CDPK16	-	-	-	-	1	93.13
CDPK17	3	152	-	-	-	-
CDPK28	-	-	-	-	-	-
CDPK34	-	-	5	17.91	-	-

6. DISCUSIÓN

En este trabajo se buscó caracterizar la familia **CDPK** en el linaje de las coníferas, con el objetivo de determinar el número parálogos y los procesos involucrados en su evolución. Se pudo determinar que las coníferas tienen entre 10 y 11 genes **CDPK**, y que esta diferencia es debida a una duplicación en tándem exclusiva de la familia Pinaceae. Además, se encontró evidencia de que esta familia génica se encuentra bajo fuerte restricción evolutiva, a pesar de lo cual, se observaron señales de selección positiva actuando en linajes específicos (selección episódica). Esto se observó de manera particular en especies que habitan en condiciones ambientales contrastantes (boreales vs. subtropicales). En conjunto, los resultados de este trabajo sugieren que la familia **CDPK** ha desempeñado una función importante en la evolución de las coníferas y en su adaptación a distintos ambientes.

Diversos trabajos han estudiado la evolución y función de familia **CDPK** en especies de plantas modelo y de interés comercial, tales como *A. thaliana* y *P. trichocarpa*, así como el trigo, la cebada y la uva (Cheng et al., 2002a; Y. Yang et al., 2017; K. Zhang et al., 2015; Zuo et al., 2013). En todos ellos, se ha reportado que los genes de esta familia tienen un papel importante en la regulación de la respuesta a distintos tipos de estrés, como el frío, calor, el congelamiento y algunos patógenos. De igual manera, se ha estudiado esta familia génica en especies de plantas como *Physcomitrella patens* y *Selaginella moellendorffii*, con lo cual se ha logrado entender el origen y la evolución de esta familia génica a lo largo de la historia evolutiva de las plantas terrestres (Hamel et al., 2014; Valmonte et al., 2014). Sin embargo, este es el primer trabajo en el que se estudia a profundidad esta familia génica en el linaje de las coníferas. Además, y a diferencia de la mayoría de los trabajos previos, que se enfocan en una sola especie, aquí se estudió a la familia **CDPK** en especies de distintas familias y géneros. Esto es importante, ya que las coníferas se han adaptado a una gran diversidad de ambientes, por lo que la familia **CDPK**, con su papel en la regulación de la respuesta a distintos tipos de estrés, es un candidato interesante para estudiar dicha adaptación.

6.1. Características, filogenia y ubicación genómica relativa de los genes *CDPK* en coníferas

Se logró determinar que existen al menos once parálogos **CDPK** en la familia Pinaceae y diez en las familias Cupressaceae y Taxaceae. Estos incluyen miembros de los cuatro clados previamente reportados en plantas terrestres, formando grupos cercanos a los de sus ortólogos

en angiospermas. Lo anterior indica que los genes **CDPK** de coníferas son derivados de los mismos genes ancestrales que dieron origen a los genes **CDPK** en angiospermas contemporáneas. De hecho, los genes **CDPK** de plantas como la briofita *P. patens* y la licofita *S. moellendorffii*, también se agrupan dentro de esos cuatro clados (Hamel et al., 2014; Valmonte et al., 2014). Esto sugiere que el ancestro común de las plantas terrestres poseía cuatro genes **CDPK** y, que a partir de estos, se multiplicaron y diversificaron en los diferentes linajes de este grupo (Valmonte et al., 2014).

Es importante señalar que el uso de transcriptomas para descubrir ortólogos y parálogos presenta dificultades para distinguir entre copias recientes, variantes alélicas, transcritos alternativos y artefactos del ensamble (Y. Yang & Smith, 2014). Para evitar la retención de secuencias que no fueran copias verdaderas, se eliminaron los agrupamientos monofiléticos de secuencias de una misma especie del árbol inicial. Sin embargo, es posible que esto haya implicado la pérdida de posibles parálogos más recientes (por debajo del nivel de familia). En efecto, previamente se han reportado 12 genes **CDPK** en coníferas (*P. sitchensis*), detectadas a partir de una colección de ESTs (Valmonte et al., 2014). También es posible que existan copias sin expresión que pueden no haber sido detectadas con el método utilizado en este trabajo. La falta de genomas de referencia hasta la fecha dificulta la determinación del número real de genes **CDPK** en coníferas, sin embargo, el número reportado previamente y el reportado en el presente trabajo, no representan una discrepancia importante.

Los genes **CDPK** de coníferas presentaron la misma estructura básica que sus homólogos en angiospermas, caracterizada por la presencia de dominios cinasa y de unión a calcio, así como de un extremo *N*-terminal y una región autoregulatoria. Si bien los dos últimos no pueden ser detectados por el análisis de InterPro Scan, el extremo *N*-terminal pudo visualizarse en los alineamientos, en los cuales se observó que este es similar en los ortólogos entre especies, y diferente entre los parálogos de una misma especie. Otra característica común entre coníferas, fue la presencia de genes con dominios *EF-hand* degenerados. En la mayoría de los casos, fue el primer dominio *EF-hand* el que presentó dicha degeneración (es decir, el más cercano al extremo *N*-terminal). Esto se observó de manera consistente en los genes *CDPK10* y *CDPK13* ya había sido reportado para sus ortólogos en *A. thaliana*, así como para otros parálogos de este mismo grupo (*CDPK7*, *CDPK8*, *CDPK30* y *CDPK32*; Liese & Romeis, 2013). Esto indicaría que esta característica estructural de los genes de este clado se originó antes de la divergencia de las gimnospermas y angiospermas.

En angiospermas, el número de genes **CDPK** es variable, y en general está asociado a eventos de duplicación genómica completa o de poliploidización. Por ejemplo, la soya (*Glycine max*) tiene 47 copias, lo cual se relaciona con diferentes eventos de poliploidización reportados para esta especie (Schmutz et al., 2010; Valmonte et al., 2014). En cambio, *Vitis vinifera*, para la cual no se han reportado duplicaciones genómicas desde la triplicación ancestral de las eudicotiledóneas, solamente presenta 17 copias (Jaillon et al., 2007; Valmonte et al., 2014). A diferencia de las angiospermas, no hay evidencia de duplicaciones genómicas recientes en gimnospermas (con excepciones como *Sequoia sempervirens*) y, por lo tanto, es poco probable que este proceso haya influido significativamente en la evolución de sus diferentes familias génicas (Nystedt et al., 2013). De hecho, la mayoría de familias génicas en coníferas son pequeñas, mostrando entre 2 y 10 miembros en promedio (De la Torre et al., 2015). De acuerdo con lo encontrado en este trabajo, la familia **CDPK** se encuentra dentro de este rango, para el caso de las cupresáceas y taxáceas; aunque, en el caso de las pináceas, el número está ligeramente por arriba del promedio. En comparación, las familias más expandidas en coníferas (con más de 100 copias), incluyen otras proteínas cinasas, las proteínas *PPR* (*pentatricopeptide repeat*) y las proteínas *LRR* (*leucin-reach repeat*) (De la Torre et al., 2015). Las primeras se encargan de regular la expresión génica a nivel de mRNA, en procesos relacionados con el desarrollo, en la fotosíntesis y la respuesta a estímulos ambientales, entre otros (Barkan & Small, 2014). Por otro lado, las proteínas *LRR* están involucradas en el reconocimiento y la inducción de respuesta ante patógenos, y también se encuentran ampliamente expandidas en otros linajes de plantas terrestres (Lee & Yeom, 2015).

Dada la falta de duplicaciones genómicas completas en coníferas, sus familias génicas probablemente se han expandido mediante duplicaciones de genes individuales o pequeños grupos; involucrando procesos como duplicaciones en tándem, segmentales y mediadas por transposones. De hecho, se ha reportado que la mayoría de los genes en coníferas que se han duplicado después de su divergencia con las angiospermas se encuentran próximos en el mismo grupo de ligamiento, frecuentemente arreglados en tándem (Pavy et al., 2012). Estos representan solo una pequeña parte del total de genes de las coníferas (2.9%-9.2%), y la mayoría de estos arreglos constan de solo dos genes (Pavy et al., 2012, 2017) lo cual es consistente con lo encontrado en la familia **CDPK**. La localización de los dos genes **CDPK** que mostraron duplicaciones exclusivas de la familia Pinaceae, están dentro del mismo grupo de ligamiento y a menos de un cM de distancia; lo que sugiere que, efectivamente, se duplicaron

en tándem. Estudios en otros grupos de plantas han revelado que los genes duplicados en tándem suelen estar enriquecidos para categorías funcionales relacionadas con la respuesta a estímulos ambientales y estrés biótico, incluida la actividad cinasa (Hanada et al., 2008; Rizzon et al., 2006). Sin embargo, en coníferas, las superfamilias cinasa y *EF-hand* no se encuentran dentro de las 15 familias más sobrerrepresentadas entre los genes en tándem (Pavy et al., 2017).

Además de los genes duplicados en tándem en la familia Pinaceae, se logró identificar otros siete genes **CDPK** en los mapas de ligamiento consultados. De los nueve genes en total, cuatro se identificaron en grupos de ligamiento equivalentes entre mapas, mientras que tres se encontraron solamente en un mapa, por lo que no se pudo determinar su homología de posición. En general, esto corresponde con los altos niveles de sintenia reportados previamente para los genomas de las coníferas (Pavy et al., 2012). Sin embargo, las posiciones de los genes dentro de los grupos de ligamiento equivalentes no fueron las mismas y, en algunos casos, la diferencia excedió los 50 y 100 cM (por ejemplo, las posiciones de los genes *CDPK1* y *CDPK13*). Esto puede deberse a las diferencias metodológicas en la construcción de los tres mapas consultados (como el número de marcadores utilizados), así como a la incertidumbre en la identificación de los marcadores como genes **CDPK**. Lo segundo puede ser más problemático en el caso de los parálogos más cercanos y con mayor similitud en sus secuencias. Por ejemplo, en el caso del gen *CDPK2*, que se localizó en grupos de ligamiento distintos en los mapas de *P. abies* y *P. glauca*, puede deberse a la dificultad para diferenciar entre los genes *CDPK1* y *CDPK2*. De hecho, el grupo de ligamiento en el que se ubicó el gen *CDPK2* en el mapa de *P. glauca*, es equivalente al grupo en el que se ubica el gen *CDPK1* en el mapa compuesto de las pináceas. Esto podría sugerir que el primero está mal identificado y en realidad se trata del gen *CDPK2*. En este caso, este gen (*CDPK2*) se encontraría en grupos de ligamiento equivalentes entre los tres mapas. Alternativamente, las diferencias en las localizaciones de los genes **CDPK** pueden deberse a re-arreglos a pequeña escala, debidos a otras duplicaciones en tándem o a la actividad de transposones, que son procesos recurrentes durante la evolución de los genomas, particularmente en coníferas (Liu et al., 2019; Nystedt et al., 2013). Finalmente, la distribución de los nueve genes **CDPK** no fue aleatoria; estos parecen acumularse en unos pocos grupos de ligamiento: cinco de ellos se encontraron en un mismo grupo (grupo 5 en el mapa de *P. abies* y 1 en los de *P. glauca* y compuesto de las pináceas), mientras que los cuatro restantes se ubicaron en otros tres grupos distintos. Los mapas de ligamiento de coníferas han mostrado previamente que los genes de una misma familia o superfamilia tienden a ubicarse dentro del mismo grupo

de ligamiento y, en algunos casos, en arreglos en tándem (Liu et al., 2019; Pavy et al., 2017). Esto corresponde con lo encontrado para la familia **CDPK** y podría sugerir que los parálogos que no se encontraron en los mapas, pueden localizarse en alguno de los grupos de ligamiento aquí reportados.

6.2. Tasas evolutivas y selección natural en los genes *CDPK* de coníferas

El coeficiente dn/ds de fondo (es decir, de las ramas constreñidas) para la familia **CDPK**, de acuerdo con el modelo por ramas de cuatro coeficientes, fue de 0.14. En comparación, estudios multigénicos han reportado que el coeficiente dn/ds promedio para el genoma de coníferas es de entre 0.26 y 0.31 (Buschiazzi et al., 2012; De la Torre et al., 2017). Esto coloca a la familia **CDPK** de las coníferas por debajo del promedio, y sugiere que está altamente conservada y se encuentra bajo restricción evolutiva. Por otro lado, los modelos por ramas sugieren que hubo un aumento significativo en los coeficientes dn/ds inmediatamente después de la duplicación de los genes *CDPK6* y *CDPK34* en la familia Pinaceae. Posteriormente, la selección purificadora se habría vuelto intensificar en al menos una de las dos copias (*CDPK6*), mientras que la otra permaneció relativamente menos constreñida. Este patrón sugiere que existe una presión selectiva para mantener la función en una de las copias, mientras que la otra ha seguido acumulando variación no sinónima. Este escenario corresponde en general con lo esperado después de una duplicación, ya que en general se observan mayores tasas evolutivas entre genes duplicados con respecto a genes de copia única. En coníferas específicamente, se ha encontrado una correlación positiva entre el número de copias en una familia génica, y sus coeficientes dn/ds (De la Torre et al., 2015). El coeficiente dn/ds mayor del gen *CDPK34* puede deberse a una mayor acumulación de sustituciones no sinónimas por medio de selección positiva, lo cual podría indicar neofuncionalización de este parólogo. Sin embargo, si el número de sustituciones no sinónimas está dado por un relajamiento de la selección purificadora, podrían fijarse sustituciones deletéreas y, por lo tanto, tratarse de un proceso de pseudogenización.

De manera individual, los diferentes parálogos de la familia mostraron coeficientes dn/ds variables, con un rango entre 0.1 y 0.34, de acuerdo con los modelos por sitio con mejor ajuste a los datos. Los valores más bajos los presentaron los genes *CDPK10* y *CDPK5*, mientras que los más altos corresponden a los genes *CDPK34* y *CDPK6*. Estudios previos han sugerido que la selección positiva es predominantemente episódica; estando caracterizada por períodos transitorios de evolución adaptativa en linajes específicos, enmascarados por los efectos

constantes de la selección purificadora en el resto de linajes (Murrell et al., 2012). Este patrón se ajusta a lo observado en la familia **CDPK** de las coníferas, ya que los genes en los que los modelos de selección se ajustaron mejor que los neutrales, tuvieron un número variable de sitios bajo selección (entre 1 y 7). Estos representarían sitios bajo selección recurrente; es decir, que ha actuado de manera relativamente constante a lo largo de la evolución de estos genes en coníferas. Varios de estos sitios se detectaron también con el análisis MEME, y presentaron señal de selección en varias ramas de la filogenia, confirmando la acción recurrente de la selección natural. Además, todos los genes, excepto *CDPK16*, presentaron sitios bajo selección episódica; es decir, que ha actuado solamente en ciertos linajes. En conjunto, lo anterior sugiere que, a pesar de estar conservada, la selección positiva ha actuado en la evolución de esta familia génica en coníferas, y que los diferentes parálogos se encuentran bajo diferentes regímenes selectivos entre especies. En la mayoría de los genes con sitios bajo selección episódica, los sitios presentaron señales en diferentes ramas, incluyendo terminales y profundas, indicando la acción de la selección en un mismo gen, pero en sitios distintos para cada linaje. Sin embargo, resultan particularmente interesantes los casos en los que todos o la mayoría de los sitios bajo selección episódica se encontraron en una misma rama. Por ejemplo, los genes *CDPK6* y *CDPK10*, mostraron sitios bajo selección casi exclusivamente en los linajes de *A. religiosa* y *P. patula*, respectivamente; es decir, en especies subtropicales. En contraste, en el gen *CDPK17*, la mayoría de sitios bajo selección se encontró en la especie boreal *P. banksiana*.

Se ha reportado previamente en angiospermas, que los genes *CDPK17* y *CDPK34* están involucrados principalmente en aspectos del desarrollo y, en particular, en el crecimiento y adecuación del tubo polínico (Schulz et al., 2013; Valmonte et al., 2014). De acuerdo con lo encontrado en este trabajo, el gen *CDPK6* es una copia reciente de *CDPK34*, y es exclusivo de la familia Pinaceae. Si bien se encontró evidencia de divergencia funcional entre ambos, este valor fue bajo, por lo cual es probable que el gen *CDPK6* de las coníferas tenga una función relacionada con la de *CDPK17* y *CDPK34*. Además, se ha reportado que estos dos genes, *CDPK17* y *CDPK34*, tienen una función parcialmente redundante, ya que solo los genotipos dobles mutantes presentan defectos en el crecimiento del tubo polínico en *A. thaliana* (Myers et al., 2009). Esto podría explicar en parte el patrón de divergencia encontrado en estos genes. En pináceas, el gen *CDPK6* tiene mayor restricción evolutiva que *CDPK34*, por lo tanto, el primero estaría reteniendo la función ancestral que, junto con la acción de *CDPK17*, mantendrían la adecuación del tubo polínico. Esta acción redundante provocaría un relajamiento en la selección purificadora sobre *CDPK34*, lo cual explica su alto número de sustituciones no sinónimas y la

distribución de estas en la filogenia. Además, el análisis en PROVEAN reveló cinco sustituciones deletéreas, las cuales están dentro del dominio cinasa y a la región autorregulatoria de estos genes, probablemente afectando su función. En conjunto, esto podría indicar que este parálogo se encuentra en un estado incipiente de pseudogenización, ya que no se encontró evidencia de codones de *stop* prematuros. Análisis de las secuencias genómicas de estos genes podrían ayudar a confirmar la pseudogenización de *CDPK34*, ya que permitirían evaluar aspectos como el nivel de homología con su gen parental, la presencia y características de su región regulatoria y la estructura de sus intrones; las cuales han sido utilizadas en estudios previos para identificar y caracterizar pseudogenes (Z. Zhang & Gerstein, 2004). Adicionalmente, los análisis de expresión diferencial podrían contribuir a esto. Si bien se ha encontrado evidencia de pseudogenes transcripcionalmente activos, también es posible que estos no se transcriban o que presenten actividad parcial (Pei et al., 2012). El último caso se ha interpretado como genes en proceso de “muerte” o pseudogenización (Pei et al., 2012). En efecto, el transcriptoma de *A. sachalinensis* muestran que el gen *CDPK34* tiene un nivel de expresión de hasta un orden de magnitud menor que sus parálogos *CDPK17* y *CDPK34*, lo cual reforzaría la idea de su pseudogenización (<http://plantomics.mind.meiji.ac.jp/todomatsu/index.html>).

Hasta la fecha, no se ha reportado si los genes *CDPK17* y *CDPK34* tienen un efecto en la adecuación del polen ante distintas condiciones ambientales y, debido a que no se ha encontrado expresión significativa en tejidos vegetativos, no se tienen reportes de su papel ante diferentes estímulos (Valmonte et al., 2014). Sin embargo, diversos estudios han indicado que las condiciones ambientales y el estrés abiótico pueden tener efecto sobre el desempeño del polen de las coníferas. Por ejemplo, se ha sugerido que tanto la falta como el exceso de humedad, pueden afectar la captura y entrada del polen en el micrópilo y, por tanto, su capacidad para fecundar el óvulo (Lora et al., 2016). Igualmente, en angiospermas se ha demostrado que temperaturas mayores tiene un efecto positivo en las tasas de germinación y crecimiento del tubo polínico, y que el polen proveniente de plantas bajo estrés por baja temperatura (incluso dentro del rango natural de temperaturas de la especie), presenta tubos polínicos más cortos y con menor éxito de fecundación (Delph et al., 1997; Hedhly et al., 2005). Por lo tanto, es posible que las señales de selección positiva en los genes *CDPK6* y *CDPK17* (en coníferas subtropicales y boreales, respectivamente), estén relacionadas con las presiones selectivas que ejercen las condiciones ambientales contrastantes en que viven ambas, sobre la germinación y desarrollo del tubo polínico. Sin embargo, será necesario confirmar esto mediante experimentos que

evalúen la expresión de estos genes en tejido reproductivo bajo distintos tipos de estrés. Por ejemplo, podrían realizarse experimentos de germinación de polen *in vitro* bajo diferentes temperaturas, y medir el éxito de la germinación y la expresión de estos genes en cada caso.

En cuanto al gen *CDPK10*, se ha reportado que está principalmente involucrado en la respuesta a estrés por sequía en especies como *A. thaliana* y *Oryza sativa*, mediante la regulación del cierre de estomas e interacción con proteínas de choque térmico (Bundó & Coca, 2017; Zou et al., 2010). En *Populus euphratica*, la única especie de este género distribuida en el límite de zonas desérticas, este gen también participa en la resistencia al congelamiento y a la salinidad (J. Chen et al., 2013). Además, se ha encontrado evidencia de selección positiva en este gen en linajes de plantas adaptadas a ambientes áridos y sujetos a estrés por sequía; por ejemplo, en la familia Cactaceae (N. Wang et al., 2019). Si bien *P. patula* habita principalmente en ambientes tropicales húmedos de montaña, existen poblaciones que habitan en condiciones de mayor aridez y temperatura. Además, plantaciones en otras zonas tropicales (Sudáfrica, Brasil, Colombia) han demostrado que esta especie puede establecerse en rangos de temperaturas más amplios que los presentes en su distribución natural (M. van Zonneveld et al., 2009). Experimentos de expresión comparada y diferencial en especies que estén adaptadas a ambientes contrastantes (p. ej. mayor o menor sequía, distintos rangos de temperatura) podrían contribuir a determinar el papel que este gen tiene en la respuesta de las coníferas a distintos estímulos ambientales y en su adaptación a condiciones ambientales contrastantes entre especies distantemente relacionadas.

7. CONCLUSIÓN

En este trabajo logró determinarse que en coníferas existen al menos 11 parálogos de la familia **CDPK** en pináceas y 10 en cupresáceas y taxáceas, y que esta diferencia se debe a una duplicación en tándem ocurrida en la familia Pinaceae. Se encontró que los genes **CDPK** de coníferas tienen altos niveles de homología con los de angiospermas, y que se agrupan junto a estos dentro de los cuatro grandes clados que componen a esta familia génica en plantas terrestres. Se encontraron altos niveles de conservación a nivel de secuencia, por lo cual puede concluirse que las funciones de los genes de esta familia han sido importantes a lo largo de la evolución de las coníferas. A pesar de lo anterior, también se encontró evidencia de selección positiva durante la evolución de esta familia en coníferas, particularmente actuando de manera episódica; es decir, en linajes específicos. Esto fue particularmente evidente en especies que habitan en ambientes contrastantes; en concreto, entre especies subtropicales y boreales. Los genes **CDPK** que presentaron evidencia de selección positiva están relacionados con el desarrollo (especies boreales y subtropicales) y con la resistencia a la sequía (especie subtropical). Por lo tanto, la evidencia sugiere que, al menos algunos de los parálogos de la familia **CDPK** podrían haber tenido un papel importante en la adaptación de las coníferas a distintos ambientes a lo largo de su historia evolutiva.

Debido a la gran capacidad adaptativa de las coníferas, evidenciada por la gran diversidad de ambientes en las que habitan, resulta importante comprender las bases genómicas responsables de dichos procesos adaptativos. Los resultados de este trabajo plantean la pregunta de qué otras familias génicas están involucradas en dicha capacidad adaptativa en coníferas, y cuáles han sido los procesos que han determinado la evolución de dichos genes (selección positiva, proliferación de copias, adquisición de nuevas funciones, etc.). Por lo tanto, es necesario realizar estudios que analicen a nivel genómico los genes de coníferas a escala filogenética, incluyendo especies cercana y lejanamente emparentadas, y que habiten en ambientes diversos. Además, estos estudios deberían complementarse con información funcional, como la expresión génica. En conjunto, esto permitirá tener una mejor comprensión de los mecanismos que les permitieron a las coníferas adaptarse a diferentes hábitats. Además, se podrá entender mejor la posible respuesta de las coníferas ante las nuevas condiciones ambientales que enfrentarán como consecuencia del cambio climático y, por lo tanto, puede contribuir al diseño e implementación de mejores medidas de conservación.

8. FUENTES

- Asai, S., Ichikawa, T., Nomura, H., Kobayashi, M., Kamiyoshihara, Y., Mori, H., Kadota, Y., Zipfel, C., Jones, J. D. G., & Yoshioka, H. (2013). The variable domain of a plant calcium-dependent protein kinase (CDPK) confers subcellular localization and substrate recognition for NADPH oxidase. *Journal of Biological Chemistry*, 288(20), 14332–14340. <https://doi.org/10.1074/jbc.M112.448910>
- Azcón Vieto, J., & Talón, M. (2013). *Fundamentos de fisiología vegetal (2a. ed.)*. McGraw-Hill España.
<http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=3229499>
- Banks, J. A., Nishiyama, T., Hasebe, M., Bowman, J. L., Gribskov, M., dePamphilis, C., Albert, V. A., Aono, N., Aoyama, T., Ambrose, B. A., Ashton, N. W., Axtell, M. J., Barker, E., Barker, M. S., Bennetzen, J. L., Bonawitz, N. D., Chapple, C., Cheng, C., Correa, L. G. G., ... Grigoriev, I. v. (2011). The Selaginella Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants. *Science*, 332(6032), 960 LP – 963. <https://doi.org/10.1126/science.1203810>
- Barkan, A., & Small, I. (2014). Pentatricopeptide repeat proteins in plants. *Annual Review of Plant Biology*, 65, 415–442. <https://doi.org/10.1146/annurev-arplant-050213-040159>
- Batistič, O., & Kudla, J. (2012). Analysis of calcium signaling pathways in plants. *Biochimica et Biophysica Acta - General Subjects*, 1820(8), 1283–1293.
<https://doi.org/10.1016/j.bbagen.2011.10.012>
- Bernhardsson, C., Vidalis, A., Wang, X., Scofield, D. G., Schiffthaler, B., Baison, J., Street, N. R., Rosario García-Gil, M., & Ingvarsson, P. K. (2019). An Ultra-Dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway spruce (*picea abies*). *G3: Genes, Genomes, Genetics*, 9(5), 1623–1632.
<https://doi.org/10.1534/g3.118.200840>
- Bielawski, J. P., & Yang, Z. (2003). Maximum likelihood methods for detecting adaptive evolution after gene duplication. *Journal of Structural and Functional Genomics*, 3(1), 201–212. <https://doi.org/10.1023/A:1022642807731>

- Bowers, J. E., Chapman, B. A., Rong, J., & Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, *422*(6930), 433–438. <https://doi.org/10.1038/nature01521>
- Bundó, M., & Coca, M. (2017). Calcium-dependent protein kinase OsCPK10 mediates both drought tolerance and blast disease resistance in rice plants. *Journal of Experimental Botany*, *68*(11), 2963–2975. <https://doi.org/10.1093/jxb/erx145>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 1–9. <https://doi.org/10.1186/1471-2105-10-421>
- Carretero-Paulet, L., & Fares, M. A. (2012). Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. *Molecular Biology and Evolution*, *29*(11), 3541–3551. <https://doi.org/10.1093/molbev/mss162>
- Casneuf, T., de Bodt, S., Raes, J., Maere, S., & van de Peer, Y. (2006). Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. *Genome Biology*, *7*(2). <https://doi.org/10.1186/gb-2006-7-2-r13>
- Chen, H., Tang, Y., Liu, J., Tan, L., Jiang, J., Wang, M., Zhu, Z., Sun, X., & Sun, C. (2017). Emergence of a novel chimeric gene underlying grain number in rice. *Genetics*, *205*(2), 993–1002. <https://doi.org/10.1534/genetics.116.188201>
- Chen, J., Xue, B., Xia, X., & Yin, W. (2013). A novel calcium-dependent protein kinase gene from *Populus euphratica*, confers both drought and cold stress tolerance. *Biochemical and Biophysical Research Communications*, *441*(3), 630–636. <https://doi.org/10.1016/j.bbrc.2013.10.103>
- Cheng, S. H., Willmann, M. R., Chen, H. C., & Sheen, J. (2002a). Calcium signaling through protein kinases. The *Arabidopsis* calcium-dependent protein kinase gene family. *Plant Physiology*, *129*(2), 469–485. <https://doi.org/10.1104/pp.005645>
- Cheng, S. H., Willmann, M. R., Chen, H. C., & Sheen, J. (2002b). Calcium signaling through protein kinases. The *Arabidopsis* calcium-dependent protein kinase gene

- family. *Plant Physiology*, 129(2), 469–485. <https://doi.org/10.1104/pp.005645>
- Choi, Y., & Chan, A. P. (2015). PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31(16), 2745–2747. <https://doi.org/10.1093/bioinformatics/btv195>
- Conant, G. C., & Wolfe, K. H. (2008). Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics*, 9(12), 938–950. <https://doi.org/10.1038/nrg2482>
- Davis, A. P., Benninghoff, A. D., Thomas, A. J., Sessions, B. R., & White, K. L. (2015). DNA methylation of the LIN28 pseudogene family. *BMC Genomics*, 16(1), 1–10. <https://doi.org/10.1186/s12864-015-1487-3>
- de La Torre, A. R., Lin, Y. C., van de Peer, Y., & Ingvarsson, P. K. (2015). Genome-wide analysis reveals diverged patterns of codon bias, gene expression, and rates of sequence evolution in *Picea* gene families. *Genome Biology and Evolution*, 7(4), 1002–1015. <https://doi.org/10.1093/gbe/evv044>
- de Miguel, M., Bartholomé, J., Ehrenmann, F., Murat, F., Moriguchi, Y., Uchiyama, K., Ueno, S., Tsumura, Y., Lagravelle, H., de Maria, N., Cabezas, J. A., Cervera, M. T., Gion, J. M., Salse, J., & Plomion, C. (2015). Evidence of intense chromosomal shuffling during conifer evolution. *Genome Biology and Evolution*, 7(10), 2799–2809. <https://doi.org/10.1093/gbe/evv185>
- Delph, L. F., Johannsson, M. H., & Stephenson, A. G. (1997). How environmental factors affect pollen performance: Ecological and evolutionary perspectives. *Ecology*, 78(6), 1632–1639. [https://doi.org/10.1890/0012-9658\(1997\)078\[1632:HEFAPP\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1997)078[1632:HEFAPP]2.0.CO;2)
- des Marais, D. L., & Rausher, M. D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature*, 454(7205), 762–765. <https://doi.org/10.1038/nature07092>
- Ding, Y., Zhou, Q., & Wang, W. (2012). Origins of new genes and evolution of their novel functions. *Annual Review of Ecology, Evolution, and Systematics*, 43, 345–363. <https://doi.org/10.1146/annurev-ecolsys-110411-160513>
- Flagel, L. E., & Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants.

- New Phytologist*, 183(3), 557–564. <https://doi.org/10.1111/j.1469-8137.2009.02923.x>
- Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology*, 60, 433–453. <https://doi.org/10.1146/annurev.arplant.043008.092122>
- Fusco, G., & Minelli, A. (2010). Phenotypic plasticity in development and evolution: facts and concepts. Introduction. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1540), 547–556.
<https://doi.org/10.1098/rstb.2009.0267>
- Ganko, E. W., Meyers, B. C., & Vision, T. J. (2007). Divergence in expression between duplicated genes in arabidopsis. *Molecular Biology and Evolution*, 24(10), 2298–2309. <https://doi.org/10.1093/molbev/msm158>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652.
<https://doi.org/10.1038/nbt.1883>
- Gu, X. (1999). Statistical methods for testing functional divergence after gene duplication. *Molecular Biology and Evolution*, 16(12), 1664–1674.
<https://doi.org/10.1093/oxfordjournals.molbev.a026080>
- Gu, X. (2006). A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. *Molecular Biology and Evolution*, 23(10), 1937–1945. <https://doi.org/10.1093/molbev/msl056>
- Gu, X., Zou, Y., Su, Z., Huang, W., Zhou, Z., Arendsee, Z., & Zeng, Y. (2013). An update of DIVERGE software for functional divergence analysis of protein family. *Molecular Biology and Evolution*, 30(7), 1713–1719. <https://doi.org/10.1093/molbev/mst069>
- Hamel, L. P., Sheen, J., & Séguin, A. (2014). Ancient signals: Comparative genomics of green plant CDPKs. *Trends in Plant Science*, 19(2), 79–89.
<https://doi.org/10.1016/j.tplants.2013.10.009>

- Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K., & Shiu, S. H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiology*, *148*(2), 993–1003.
<https://doi.org/10.1104/pp.108.122457>
- Harmon, A. C., Gribskov, M., & Harper, J. F. (2000). CDPKs - A kinase for every Ca²⁺ signal? *Trends in Plant Science*, *5*(4), 154–159. [https://doi.org/10.1016/S1360-1385\(00\)01577-6](https://doi.org/10.1016/S1360-1385(00)01577-6)
- Hashimoto, K., & Kudla, J. (2011). Calcium decoding mechanisms in plants. *Biochimie*, *93*(12), 2054–2059. <https://doi.org/10.1016/j.biochi.2011.05.019>
- Hedhly, A., Hormaza, J. I., & Herrero, M. (2005). The effect of temperature on pollen germination, pollen tube growth, and stigmatic receptivity in peach. *Plant Biology*, *7*(5), 476–483. <https://doi.org/10.1055/s-2005-865850>
- Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., ... Wincker, P. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, *449*(7161), 463–467. <https://doi.org/10.1038/nature06148>
- Kawamoto, N., Sasabe, M., Endo, M., Machida, Y., & Araki, T. (2015). Calcium-dependent protein kinases responsible for the phosphorylation of a bZIP transcription factor FD crucial for the florigen complex formation. *Scientific Reports*, *5*, 8341.
<https://doi.org/10.1038/srep08341>
- Kiegle, E., Moore, C. A., Haseloff, J., Tester, M. A., & Knight, M. R. (2000). Cell-type-specific calcium responses to drought, salt and cold in the Arabidopsis root. *Plant Journal*, *23*(2), 267–278. <https://doi.org/10.1046/j.1365-313X.2000.00786.x>
- Kobayashi, M., Ohura, I., Kawakita, K., Yokota, N., Fujiwara, M., Shimamoto, K., Doke, N., & Yoshioka, H. (2007). Calcium-dependent protein kinases regulate the production of reactive oxygen species by potato NADPH oxidase. *Plant Cell*, *19*(3), 1065–1080. <https://doi.org/10.1105/tpc.106.048884>
- Kosakovsky Pond, S. L., Frost, S. D. W., & Muse, S. v. (2005). HyPhy: Hypothesis testing

- using phylogenies. *Bioinformatics*, 21(5), 676–679.
<https://doi.org/10.1093/bioinformatics/bti079>
- Kudla, J., Batistič, O., & Hashimoto, K. (2010). Calcium signals: The Lead Currency of plant information processing. *Plant Cell*, 22(3), 541–563.
<https://doi.org/10.1105/tpc.109.072686>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Lee, H. A., & Yeom, S. I. (2015). Plant NB-LRR proteins: Tightly regulated sensors in a complex manner. *Briefings in Functional Genomics*, 14(4), 233–242.
<https://doi.org/10.1093/bfpg/elv012>
- Leitch, A. R., & Leitch, I. J. (2012). Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *New Phytologist*, 194(3), 629–646.
<https://doi.org/10.1111/j.1469-8137.2012.04105.x>
- Li, Z., Baniaga, A. E., Sessa, E. B., Scascitelli, M., Graham, S. W., Rieseberg, L. H., & Barker, M. S. (2015). Early genome duplications in conifers and other seed plants. *Science Advances*, 1(10), 1–8. <https://doi.org/10.1126/sciadv.1501084>
- Liese, A., & Romeis, T. (2013). Biochemical regulation of in vivo function of plant calcium-dependent protein kinases (CDPK). *Biochimica et Biophysica Acta - Molecular Cell Research*, 1833(7), 1582–1589.
<https://doi.org/10.1016/j.bbamcr.2012.10.024>
- Liu, J. J., Schoettle, A. W., Sniezko, R. A., Yao, F., Zamany, A., Williams, H., & Rancourt, B. (2019). Limber pine (*Pinus flexilis* James) genetic map constructed by exome-seq provides insight into the evolution of disease resistance and a genomic resource for genomics-based breeding. *Plant Journal*, 98(4), 745–758.
<https://doi.org/10.1111/tpj.14270>
- Long, M., Betrán, E., Thornton, K., & Wang, W. (2003). The origin of new genes: Glimpses from the young and old. In *Nature Reviews Genetics* (Vol. 4, Issue 11, pp. 865–875). <https://doi.org/10.1038/nrg1204>

- Lora, J., Hormaza, J. I., & Herrero, M. (2016). The diversity of the pollen tube pathway in plants: Toward an increasing control by the sporophyte. *Frontiers in Plant Science*, 7(FEB2016), 1–12. <https://doi.org/10.3389/fpls.2016.00107>
- Ludwig, A. A., Saitoh, H., Felix, G., Freymark, G., Miersch, O., Wasternack, C., Boller, T., Jones, J. D. G., & Romeis, T. (2005). Ethylene-mediated cross-talk between calcium-dependent protein kinase and MAPK signaling controls stress responses in plants. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), 10736–10741. <https://doi.org/10.1073/pnas.0502954102>
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, 8(7). <https://doi.org/10.1371/journal.pgen.1002764>
- Myers, C., Romanowsky, S. M., Barron, Y. D., Garg, S., Azuse, C. L., Curran, A., Davis, R. M., Hatton, J., Harmon, A. C., & Harper, J. F. (2009). Calcium-dependent protein kinases regulate polarized tip growth in pollen tubes. *Plant Journal*, 59(4), 528–539. <https://doi.org/10.1111/j.1365-313X.2009.03894.x>
- Neale, D. B., & Wheeler, N. C. (2019). The conifers: Genomes, variation and evolution. In *The Conifers: Genomes, Variation and Evolution*. <https://doi.org/10.1007/978-3-319-46807-5>
- Nielsen, R., & Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, 148(3), 929–936.
- Niu, S., Li, J., Bo, W., Yang, W., Zuccolo, A., Giacomello, S., Chen, X., Han, F., Yang, J., Song, Y., Nie, Y., Zhou, B., Wang, P., Zuo, Q., Zhang, H., Ma, J., Wang, J., Wang, L., Zhu, Q., ... Wu, H. X. (2022). The Chinese pine genome and methylome unveil key features of conifer evolution. *Cell*, 185(1), 204-217.e14. <https://doi.org/10.1016/j.cell.2021.12.006>
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., ... Jansson, S. (2013). The Norway spruce genome sequence and conifer

- genome evolution. *Nature*, 497(7451), 579–584. <https://doi.org/10.1038/nature12211>
- Panchy, N., Lehti-Shiu, M., & Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiology*, 171(4), 2294–2316. <https://doi.org/10.1104/pp.16.00523>
- Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Paterson, A. H., Freeling, M., Tang, H., & Wang, X. (2010). Insights from the comparison of plant genome sequences. *Annual Review of Plant Biology*, 61, 349–372. <https://doi.org/10.1146/annurev-arplant-042809-112235>
- Pavy, N., Lamothe, M., Pelgas, B., Gagnon, F., Birol, I., Bohlmann, J., Mackay, J., Isabel, N., & Bousquet, J. (2017). A high-resolution reference genetic map positioning 8.8 K genes for the conifer white spruce: structural genomics implications and correspondence with physical distance. *Plant Journal*, 90(1), 189–203. <https://doi.org/10.1111/tpj.13478>
- Pavy, N., Pelgas, B., Laroche, J., Rigault, P., Isabel, N., & Bousquet, J. (2012). A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. *BMC Biology*, 10(October). <https://doi.org/10.1186/1741-7007-10-84>
- Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X. J., Harte, R., Balasubramanian, S., Tanzer, A., Diekhans, M., Reymond, A., Hubbard, T. J., Harrow, J., & Gerstein, M. B. (2012). The GENCODE pseudogene resource. *Genome Biology*, 13(9). <https://doi.org/10.1186/gb-2012-13-9-r51>
- Prunier, J., Verta, J.-P., & MacKay, J. J. (2016). Conifer genomics and adaptation: at the crossroads of genetic diversity and genome function. *New Phytologist*, 209(1), 44–62. <https://doi.org/https://doi.org/10.1111/nph.13565>
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5), 901–904. <https://doi.org/10.1093/sysbio/syy032>
- Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., Nishiyama,

- T., Perroud, P.-F., Lindquist, E. A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin-I, T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., ... Boore, J. L. (2008). The Physcomitrella Genome Reveals Evolutionary Insights into the Conquest of Land by Plants. *Science*, 319(5859), 64 LP – 69. <https://doi.org/10.1126/science.1150646>
- Reyes Galindo, V. (2019). Análisis transcriptómico de la tolerancia a ozono troposférico en *Abies religiosa*. [tesis de maestría no publicada]. UNAM.
- Rizzon, C., Ponger, L., & Gaut, B. S. (2006). Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. *PLoS Computational Biology*, 2(9), 0989–1000. <https://doi.org/10.1371/journal.pcbi.0020115>
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., Xu, D., Hellsten, U., May, G. D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M. K., Sandhu, D., Valliyodan, B., ... Jackson, S. A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, 463(7278), 178–183. <https://doi.org/10.1038/nature08670>
- Schulz, P., Herde, M., & Romeis, T. (2013). Calcium-dependent protein kinases: Hubs in plant stress signaling and development. *Plant Physiology*, 163(2), 523–530. <https://doi.org/10.1104/pp.113.222539>
- Scott, A. D., Zimin, A. v, Puiu, D., Workman, R., Britton, M., Zaman, S., Caballero, M., Read, A. C., Bogdanove, A. J., Burns, E., Wegrzyn, J., Timp, W., Salzberg, S. L., & Neale, D. B. (n.d.). *The giant sequoia genome and proliferation of disease resistance genes*. <https://doi.org/10.1101/2020.03.17.995944>
- Shaheen, H., Sarwar, R., Firdous, S. S., Ejaz Ul Islam Dar, M., Ullah, Z., & Khan, S. M. (2015). Distribution and structure of conifers with special emphasis on *Taxus baccata* in moist temperate forests of Kashmir Himalayas. *Pakistan Journal of Botany*, 47(SI),

71–76.

- Singh, A., Sagar, S., & Biswas, D. K. (2017). Calcium Dependent Protein Kinase, a Versatile Player in Plant Stress Management and Development. *Critical Reviews in Plant Sciences*, 36(5–6), 336–352. <https://doi.org/10.1080/07352689.2018.1428438>
- Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, 32(5), 1342–1353. <https://doi.org/10.1093/molbev/msv022>
- Taiz, L., Zeiger, E., Møller, I. M., & Murphy, A. S. (2018). *Plant physiology and development*.
- Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, 10(3), 512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>
- Valmonte, G. R., Arthur, K., Higgins, C. M., & Macdiarmid, R. M. (2014). Calcium-dependent protein kinases in plants: Evolution, expression and function. *Plant and Cell Physiology*, 55(3), 551–569. <https://doi.org/10.1093/pcp/pct200>
- van Zonneveld, M., Jarvis, A., Dvorak, W., Lema, G., & Leibing, C. (2009). Climate change impact predictions on *Pinus patula* and *Pinus tecunumanii* populations in Mexico and Central America. *Forest Ecology and Management*, 257(7), 1566–1576. <https://doi.org/10.1016/j.foreco.2008.12.027>
- Verslues, P. E., Agarwal, M., Katiyar-Agarwal, S., Zhu, J., & Zhu, J. K. (2006). Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. *Plant Journal*, 45(4), 523–539. <https://doi.org/10.1111/j.1365-313X.2005.02593.x>
- Vlad, F., Rubio, S., Rodrigues, A., Sirichandra, C., Belin, C., Robert, N., Leung, J., Rodriguez, P. L., Laurière, C., & Merlot, S. (2009). Protein phosphatases 2C regulate the activation of the Snf1-related kinase OST1 by abscisic acid in *Arabidopsis*. *Plant Cell*, 21(10), 3170–3184. <https://doi.org/10.1105/tpc.109.069179>

- Wang, N., Yang, Y., Moore, M. J., Brockington, S. F., Walker, J. F., Brown, J. W., Liang, B., Feng, T., Edwards, C., Mikenas, J., Olivieri, J., Hutchison, V., Timoneda, A., Stoughton, T., Puente, R., Majure, L. C., Eggli, U., & Smith, S. A. (2019). Evolution of Portulacineae Marked by Gene Tree Conflict and Gene Family Expansion Associated with Adaptation to Harsh Environments. *Molecular Biology and Evolution*, *36*(1), 112–126. <https://doi.org/10.1093/molbev/msy200>
- Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T. H., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, *40*(7), 1–14. <https://doi.org/10.1093/nar/gkr1293>
- Wang, Y., Wang, X., & Paterson, A. H. (2012). Genome and gene duplications and gene expression divergence: A view from plants. *Annals of the New York Academy of Sciences*, *1256*(1), 1–14. <https://doi.org/10.1111/j.1749-6632.2011.06384.x>
- Wegrzyn, J. L., Liechty, J. D., Stevens, K. A., Wu, L. S., Loopstra, C. A., Vasquez-Gross, H. A., Dougherty, W. M., Lin, B. Y., Zieve, J. J., Martínez-García, P. J., Holt, C., Yandell, M., Zimin, A. v., Yorke, J. A., Crepeau, M. W., Puiu, D., Salzberg, S. L., de Jong, P. J., Mockaitis, K., ... Neale, D. B. (2014). Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics*, *196*(3), 891–909. <https://doi.org/10.1534/genetics.113.159996>
- Witte, C. P., Keinath, N., Dubiella, U., Demoulière, R., Seal, A., & Romeis, T. (2010). Tobacco calcium-dependent protein kinases are differentially phosphorylated in vivo as part of a kinase cascade that regulates stress response. *Journal of Biological Chemistry*, *285*(13), 9740–9748. <https://doi.org/10.1074/jbc.M109.052126>
- Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., & Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(33), 13875–13879. <https://doi.org/10.1073/pnas.0811575106>
- Woodhouse, M. R., Schnable, J. C., Pedersen, B. S., Lyons, E., Lisch, D., Subramaniam, S., & Freeling, M. (2010). Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. *PLoS Biology*, *8*(6).

<https://doi.org/10.1371/journal.pbio.1000409>

- Wright, E. S. (2016). Using DECIPHER v2.0 to analyze big biological sequence data in R. *R Journal*, 8(1), 352–359. <https://doi.org/10.32614/rj-2016-025>
- Yang, K. Y., Liu, Y., & Zhang, S. (2001). Activation of a mitogen-activated protein kinase pathway is involved in disease resistance in tobacco. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), 741–746. <https://doi.org/10.1073/pnas.98.2.741>
- Yang, Y., & Smith, S. A. (2014). Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution*, 31(11), 3081–3092. <https://doi.org/10.1093/molbev/msu245>
- Yang, Y., Wang, Q., Chen, Q., Yin, X., Qian, M., Sun, X., & Yang, Y. (2017). Genome-wide survey indicates diverse physiological roles of the barley (*Hordeum vulgare* L.) calcium-dependent protein kinase genes /631/337/1644 /631/443/810 /631/449/2661/2665 /38/39 /38/43 /38/77 /38/22 /45/29 /45/43 article. *Scientific Reports*, 7(1), 1–15. <https://doi.org/10.1038/s41598-017-05646-w>
- Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, 15(5), 568–573. <https://doi.org/10.1093/oxfordjournals.molbev.a025957>
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang, Z., & Nielsen, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution*, 46(4), 409–418. <https://doi.org/10.1007/PL00006320>
- Yang, Z., & Swanson, W. J. (2002). Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Molecular Biology and Evolution*, 19(1), 49–57. <https://doi.org/10.1093/oxfordjournals.molbev.a003981>
- Yang, Z., Wong, W. S. W., & Nielsen, R. (2005). Bayes empirical Bayes inference of

- amino acid sites under positive selection. *Molecular Biology and Evolution*, 22(4), 1107–1118. <https://doi.org/10.1093/molbev/msi097>
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). Ggtree: an R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data. *Methods in Ecology and Evolution*, 8(1), 28–36. <https://doi.org/10.1111/2041-210X.12628>
- Zhang, C., Wang, J., Marowsky, N. C., Long, M., Wing, R. A., & Fan, C. (2013). High occurrence of functional new chimeric genes in survey of rice chromosome 3 short arm genome sequences. *Genome Biology and Evolution*, 5(5), 1038–1048. <https://doi.org/10.1093/gbe/evt071>
- Zhang, K., Han, Y. T., Zhao, F. L., Hu, Y., Gao, Y. R., Ma, Y. F., Zheng, Y., Wang, Y. J., & Wen, Y. Q. (2015). Genome-wide Identification and Expression Analysis of the CDPK Gene Family in Grape, *Vitis* spp. *BMC Plant Biology*, 15(1), 1. <https://doi.org/10.1186/s12870-015-0552-z>
- Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W. L., Chen, H., Henderson, I. R., Shinn, P., Pellegrini, M., Jacobsen, S. E., & Ecker, J. R. R. (2006). Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in *Arabidopsis*. *Cell*, 126(6), 1189–1201. <https://doi.org/10.1016/j.cell.2006.08.003>
- Zhang, Z., & Gerstein, M. (2004). Large-scale analysis of pseudogenes in the human genome. *Current Opinion in Genetics and Development*, 14(4), 328–335. <https://doi.org/10.1016/j.gde.2004.06.003>
- Zhao, L. N., Shen, L. K., Zhang, W. Z., Zhang, W., Wang, Y., & Wu, W. H. (2013). Ca²⁺-dependent protein kinase11 and 24 modulate the activity of the inward rectifying K⁺ channels in *Arabidopsis* pollen tubes. *Plant Cell*, 25(2), 649–661. <https://doi.org/10.1105/tpc.112.103184>
- Zonneveld, B. J. M. (2012). Conifer genome sizes of 172 species, covering 64 of 67 genera, range from 8 to 72 picogram. *Nordic Journal of Botany*, 30(4), 490–502. <https://doi.org/10.1111/j.1756-1051.2012.01516.x>
- Zou, J. J., Wei, F. J., Wang, C., Wu, J. J., Ratnasekera, D., Liu, W. X., & Wu, W. H.

(2010). Arabidopsis calcium-dependent protein kinase cpk10 functions in abscisic acid- and Ca²⁺-mediated stomatal regulation in response to drought stress. *Plant Physiology*, 154(3), 1232–1243. <https://doi.org/10.1104/pp.110.157545>

Zuo, R., Hu, R., Chai, G., Xu, M., Qi, G., Kong, Y., & Zhou, G. (2013). Genome-wide identification, classification, and expression analysis of CDPK and its closely related gene families in poplar (*Populus trichocarpa*). *Molecular Biology Reports*, 40(3), 2645–2662. <https://doi.org/10.1007/s11033-012-2351-z>

9. ANEXOS

Anexo 1. Listas de los transcritos retenidos para cada especie, después de aplicar los filtros señalados en el Método. Se incluye el identificador del transcrito, su mejor *hit* en las bases de datos Araport11 y SwissProt, la anotación asignada al transcrito, la posición (bp) de los dominios cinasa y de unión a calcio (de acuerdo con InterproScan) y notas relacionadas con la estructura del gen.

TablaAnexa 1: Abies balsamea

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GGJG01008083.1	<i>CDPK1</i>	<i>CDPK1</i>	Abibal <i>CDPK1</i>	122-413	430-575	
GGJG01021475.1	<i>CDPK1</i>	<i>CDPK1</i>	Abibal <i>CDPK1-2</i>	113-408	423-570	
GGJG01023950.1	<i>CDPK2</i>	<i>CDPK2</i>	Abibal <i>CDPK2</i>	2-254	283-422	
GGJG01016093.1	<i>CDPK2</i>	<i>CDPK2</i>	Abibal <i>CDPK2-2</i>	11-308	325-465	
GGJG01004459.1	<i>CDPK5</i>	<i>CDPK4</i>	Abibal <i>CDPK5</i>	119-418	434-574	
GGJG01022112.1	<i>CDPK4</i>	<i>CDPK5</i>	Abibal <i>CDPK5-2</i>	122-369	NA	Sin dominio de unión a Ca ²⁺
GGJG01027264.1	<i>CDPK6</i>	<i>CDPK3</i>	Abibal <i>CDPK6</i>	23-284	323-464	
GGJG01005665.1	<i>CDPK6</i>	<i>CDPK3</i>	Abibal <i>CDPK6-2</i>	4-227	244-385	
GGJG01001842.1	<i>CDPK17/6</i>	<i>CDPK17</i>	Abibal <i>CDPK6-3</i>	63-325	364-504	
GGJG01000508.1	<i>CDPK7</i>	<i>CDPK7</i>	Abibal <i>CDPK7</i>	59-328	345-514	
GGJG01030713.1	<i>CDPK30</i>	<i>CDPK30</i>	Abibal <i>CDPK10</i>	42-336	339-454	Tres sitios de unión a Ca ²⁺
GGJG01013954.1	<i>CDPK30</i>	<i>CDPK3</i>	Abibal <i>CDPK10-2</i>	2-242	262-440	Tres sitios de unión a Ca ²⁺
GGJG01014548.1	<i>CDPK13</i>	<i>CDPK13</i>	Abibal <i>CDPK13</i>	45-340	339-515	
GGJG01000401.1	<i>CDPK16</i>	<i>CDPK16</i>	Abibal <i>CDPK16</i>	74-351	368-519	
GGJG01006300.1	<i>CDPK17</i>	<i>CDPK17</i>	Abibal <i>CDPK17</i>	81-371	389-530	
GGJG01018427.1	<i>CDPK28</i>	<i>CDPK28</i>	Abibal <i>CDPK28</i>	56-332	344-505	
GGJG01021395.1	<i>CDPK17</i>	<i>CDPK17</i>	Abibal <i>CDPK34</i>	63-335	364-505	

TablaAnexa 2: Abies pinsapo

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GCZN01010017.1	<i>CDPK2</i>	<i>CDPK2</i>	Abipin <i>CDPK2</i>	1-204	221-367	
GCZN01010018.1	<i>CDPK2</i>	<i>CDPK2</i>	Abipin <i>CDPK2-2</i>	93-390	407-546	
GCZN01039594.1	<i>CDPK5</i>	<i>CDPK4</i>	Abipin <i>CDPK5</i>	106-404	409-520	Tres dominios EF-hand
GCZN01013305.1	<i>CDPK6</i>	<i>CDPK3</i>	Abipin <i>CDPK6</i>	70-336	370-511	
GCZN01023682.1	<i>CDPK7</i>	<i>CDPK20</i>	Abipin <i>CDPK7</i>	1-196	205-374	
GCZN01023683.1	<i>CDPK30</i>	<i>CDPK3</i>	Abipin <i>CDPK10</i>	42-336	337-516	Tres sitios de unión a Ca ²⁺
GCZN01059103.1	<i>CDPK13</i>	<i>CDPK13</i>	Abipin <i>CDPK13</i>	45-340	339-515	
GCZN01031138.1	<i>CDPK16</i>	<i>CDPK16</i>	Abipin <i>CDPK16</i>	72-349	366-517	
GCZN01027992.1	<i>CDPK17</i>	<i>CDPK1</i>	Abipin <i>CDPK17</i>	82-372	390-531	

TablaAnexa 3: *Abies religiosa*

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
TRINITY_DN1368_c0_g1_i1	<i>CDPK2</i>	<i>CDPK2</i>	Abirel <i>CDPK2</i>	124-415	432-577	
TRINITY_DN1921_c0_g1_i1	<i>CDPK2</i>	<i>CDPK2</i>	Abirel <i>CDPK2-2</i>	93-390	407-546	
TRINITY_DN47229_c0_g1_i1	<i>CDPK5</i>	<i>CDPK4</i>	Abirel <i>CDPK5</i>	106-404	420-560	Tres sitios de unión a Ca ²⁺
TRINITY_DN1249_c0_g1_i3	<i>CDPK17</i>	<i>CDPK17</i>	Abirel <i>CDPK6-17</i>	70-331	370-511	
TRINITY_DN1249_c0_g1_i1	<i>CDPK6</i>	<i>CDPK3</i>	Abirel <i>CDPK6</i>	3-173	191-332	
TRINITY_DN1249_c0_g1_i2	<i>CDPK17</i>	<i>CDPK17</i>	Abirel <i>CDPK6-17-2</i>	63-325	364-505	
TRINITY_DN181_c0_g1_i2	<i>CDPK7</i>	<i>CDPK20</i>	Abirel <i>CDPK7</i>	1-196	205-374	
TRINITY_DN181_c1_g1_i1	<i>CDPK30</i>	<i>CDPK3</i>	Abirel <i>CDPK10</i>	42-336	337-516	Tres sitios de unión a Ca ²⁺
TRINITY_DN8242_c0_g1_i1	<i>CDPK13</i>	<i>CDPK13</i>	Abirel <i>CDPK13</i>	45-340	339-515	
TRINITY_DN8242_c0_g1_i3	<i>CDPK13</i>	<i>CDPK13</i>	Abirel <i>CDPK13-2</i>	51-308	NA	Sin dominio de unión a Ca ²⁺
TRINITY_DN926_c0_g2_i2	<i>CDPK16</i>	<i>CDPK16</i>	Abirel <i>CDPK16</i>	76-353	370-521	
TRINITY_DN486_c0_g1_i3	<i>CDPK17</i>	<i>CDPK17</i>	Abirel <i>CDPK17</i>	82-372	390-531	

TablaAnexa 4: *Larix laricina*

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GGJA01007453.1	<i>CDPK1</i>	<i>CDPK1</i>	Larlar <i>CDPK1</i>	124-390	422-570	
GGJA01014881.1	<i>CDPK2</i>	<i>CDPK2</i>	Larlar <i>CDPK2</i>	89-389	405-545	
GGJA01005252.1	<i>CDPK17</i>	<i>CDPK17</i>	Larlar <i>CDPK6-17</i>	71-332	371-512	
GGJA01001686.1	<i>CDPK34</i>	<i>CDPK34</i>	Larlar <i>CDPK6-34</i>	65-326	364-505	
GGJA01000498.1	<i>CDPK7</i>	<i>CDPK7</i>	Larlar <i>CDPK7</i>	61-323	346-515	
GGJA01012889.1	<i>CDPK1</i>	<i>CDPK20</i>	Larlar <i>CDPK10?</i>	43-336	337-516	Tres sitios de unión a Ca ²⁺
GGJA01013449.1	<i>CDPK13</i>	<i>CDPK13</i>	Larlar <i>CDPK13</i>	46-339	338-515	Tres sitios de unión a Ca ²⁺
GGJA01000392.1	<i>CDPK16</i>	<i>CDPK16</i>	Larlar <i>CDPK16</i>	127-404	421-572	
GGJA01005835.1	<i>CDPK17</i>	<i>CDPK17</i>	Larlar <i>CDPK17</i>	81-371	390-532	
GGJA01017203.1	<i>CDPK28</i>	<i>CDPK28</i>	Larlar <i>CDPK28</i>	55-333	343-501	

TablaAnexa 5: *Picea glauca*

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GGJI01007672.1	<i>CDPK2</i> isoform 2	<i>CDPK2</i>	Picgla <i>CDPK1</i>	130-396	428-575	Tres sitios de unión a Ca ²⁺
GGJI01015700.1	<i>CDPK2</i> isoform 2	<i>CDPK2</i>	Picgla <i>CDPK2</i>	92-392	408-548	
GGJI01004159.1	<i>CDPK5</i>	<i>CDPK4</i>	Picgla <i>CDPK5</i>	92-395	411-551	
GGJI01005328.1	<i>CDPK17</i>	<i>CDPK17</i>	Picgla <i>CDPK6</i>	72-342	372-513	
GGJI01000503.1	<i>CDPK7</i>	<i>CDPK20</i>	Picgla <i>CDPK7</i>	62-322	349-526	
GGJI01013551.1	<i>CDPK1</i>	<i>CDPK20</i>	Picgla <i>CDPK10</i>	43-336	337-515	Tres sitios de unión a Ca ²⁺
GGJI01014174.1	<i>CDPK13</i>	<i>CDPK13</i>	Picgla <i>CDPK13</i>	46-339	338-515	Tres sitios de unión a Ca ²⁺
GGJI01004090.1	<i>CDPK16</i>	<i>CDPK16</i>	Picgla <i>CDPK16</i>	75-351	369-520	
GGJI01005943.1	<i>CDPK17</i>	<i>CDPK17</i>	Picgla <i>CDPK17</i>	79-370	389-531	
GGJI01001628.1	<i>CDPK17</i>	<i>CDPK17</i>	Picgla <i>CDPK34</i>	60-329	361-501	

TablaAnexa 6: *Pinus banksiana*

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GGJH01001673.1	<i>CDPK1</i>	<i>CDPK1</i>	Pinbank <i>CDPK1</i>	113-408	423-570	
GGJH01038601.1	<i>CDPK2</i>	<i>CDPK2</i>	Pinbank <i>CDPK2</i>	92-390	391-506	Tres dominios EF-hand
GGJH01007737.1	<i>CDPK2</i>	<i>CDPK2</i>	Pinbank <i>CDPK2-2</i>	88-387	404-544	
GGJH01005842.1	<i>CDPK5</i>	<i>CDPK4</i>	Pinbank <i>CDPK5</i>	106-404	420-560	Tres sitios de unión a Ca ²⁺
GGJH01028602.1	<i>CDPK6</i>	<i>CDPK3</i>	Pinbank <i>CDPK6</i>	70-331	370-511	
GGJH01003374.1	<i>CDPK6</i>	<i>CDPK3</i>	Pinbank <i>CDPK6-2</i>	95-365	395-536	
GGJH01011650.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinbank <i>CDPK6-17</i>	1-197	222-363	
GGJH01002600.1	<i>CDPK7</i>	<i>CDPK7</i>	Pinbank <i>CDPK7</i>	58-321	344-513	
GGJH01004769.1	<i>CDPK30</i>	<i>CDPK3</i>	Pinbank <i>CDPK10</i>	43-331	337-515	Tres sitios de unión a Ca ²⁺
GGJH01025749.1	<i>CDPK13</i>	<i>CDPK3</i>	Pinbank <i>CDPK10-13</i>	2-146	146-325	Tres sitios de unión a Ca ²⁺
GGJH01035592.1	<i>CDPK13</i>	<i>CDPK13</i>	Pinbank <i>CDPK13</i>	45-340	339-515	
GGJH01001290.1	<i>CDPK13</i>	<i>CDPK13</i>	Pinbank <i>CDPK13-2</i>	59-349	348-516	Tres sitios de unión a Ca ²⁺
GGJH01005214.1	<i>CDPK16</i>	<i>CDPK16</i>	Pinbank <i>CDPK16</i>	78-355	372-523	
GGJH01005215.1	<i>CDPK16</i>	<i>CDPK16</i>	Pinbank <i>CDPK16-2</i>	8-247	250-408	
GGJH01027464.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinbank <i>CDPK17</i>	81-371	389-530	
GGJH01020041.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinbank <i>CDPK17-2</i>	102-369	402-543	

TablaAnexa 7: *Pinus monticola*

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GDQD01006074.1	<i>CDPK2</i>	<i>CDPK2</i>	Pinmont <i>CDPK1</i>	89-388	405-545	
GDQD01010129.1	<i>CDPK5</i>	<i>CDPK4</i>	Pinmont <i>CDPK5</i>	93-395	411-551	Tres sitios de unión a Ca ²⁺
GDQD01000430.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinmont <i>CDPK6-17</i>	57-360	378-519	
GDQD01002270.1	<i>CDPK7</i>	<i>CDPK20</i>	Pinmont <i>CDPK7</i>	65-325	348-517	
GDQD01008625.1	<i>CDPK1</i>	<i>CDPK20</i>	Pinmont <i>CDPK10</i>	43-332	337-515	Tres sitios de unión a Ca ²⁺
GDQD01003550.1	<i>CDPK13</i>	<i>CDPK13</i>	Pinmont <i>CDPK13</i>	60-350	353-529	Tres sitios de unión a Ca ²⁺
GDQD01004026.1	<i>CDPK16</i>	<i>CDPK16</i>	Pinmont <i>CDPK16</i>	1-154	166-317	

TablaAnexa 8: *Pinus patula*

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GECO01027391.1	<i>CDPK1</i>	<i>CDPK1</i>	Pinpat <i>CDPK1</i>	113-408	423-570	
GECO01040240.1	<i>CDPK5</i>	<i>CDPK4</i>	Pinpat <i>CDPK5</i>	96-395	411-551	
GECO01038217.1	<i>CDPK6</i>	<i>CDPK3</i>	Pinpat <i>CDPK6</i>	78-348	378-519	
GECO01002400.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinpat <i>CDPK6-17</i>	6-267	306-430	Tres dominios EF-hand
GECO01002399.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinpat <i>CDPK6-17-2</i>	6-271	290-400	Dos dominios EF-hand
GECO01027402.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinpat <i>CDPK6-17-3</i>	64-325	363-504	Tres sitios de unión a Ca ²⁺
GECO01035146.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinpat <i>CDPK6-17-4</i>	40-347	364-504	Tres sitios de unión a Ca ²⁺
GECO01017745.1	<i>CDPK7</i>	<i>CDPK7</i>	Pinpat <i>CDPK7</i>	58-327	344-513	
GECO01037008.1	<i>CDPK1</i>	<i>CDPK20</i>	Pinpat <i>CDPK10</i>	43-331	337-514	Tres sitios de unión a Ca ²⁺
GECO01025051.1	<i>CDPK13</i>	<i>CDPK13</i>	Pinpat <i>CDPK13</i>	43-336	335-503	Tres sitios de unión a Ca ²⁺
GECO01024576.1	<i>CDPK16</i>	<i>CDPK16</i>	Pinpat <i>CDPK16</i>	78-355	372-523	Tres sitios de unión a Ca ²⁺
GECO01045818.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinpat <i>CDPK17</i>	88-379	398-539	
GECO01019618.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinpat <i>CDPK17-2</i>	98-360	398-513	Un dominio EF-hand
GECO01040404.1	<i>CDPK16</i>	<i>CDPK16</i>	Pinpat <i>CDPK28</i>	54-334	343-501	

TablaAnexa 9: *Pinus strobus*

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GGJK01001545.1	<i>CDPK1</i>	<i>CDPK1</i>	Pinstrob <i>CDPK1</i>	11-407	423-568	
GGJK01006936.1	<i>CDPK2</i>	<i>CDPK2</i>	Pinstrob <i>CDPK2</i>	89-388	405-545	
GGJK01010425.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinstrob <i>CDPK6-17</i>	62-321	359-500	
GGJK01002355.1	<i>CDPK7</i>	<i>CDPK7</i>	Pinstrob <i>CDPK7</i>	62-325	348-517	
GGJK01001188.1	<i>CDPK13</i>	<i>CDPK13</i>	Pinstrob <i>CDPK13</i>	60-350	353-529	Tres sitios de unión a Ca ²⁺
GGJK01004699.1	<i>CDPK16</i>	<i>CDPK16</i>	Pinstrob <i>CDPK16</i>	73-350	367-518	
GGJK01022959.1	<i>CDPK17</i>	<i>CDPK17</i>	Pinstrob <i>CDPK17</i>	9-276	310-450	
GGJK01027806.1	<i>CDPK16</i>	<i>CDPK16</i>	Pinstrob <i>CDPK28</i>	57-334	343-501	

TablaAnexa 10: *Pinus tecunumanii*

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GFZL01005316.1	<i>CDPK1</i>	<i>CDPK1</i>	Pintec <i>CDPK1</i>	113-408	423-570	
GFZL01014342.1	<i>CDPK2</i>	<i>CDPK2</i>	Pintec <i>CDPK2</i>	88-387	404-544	
GFZL01008041.1	<i>CDPK5</i>	<i>CDPK4</i>	Pintec <i>CDPK5</i>	147-446	462-602	
GFZL01001961.1	<i>CDPK6</i>	<i>CDPK3</i>	Pintec <i>CDPK6</i>	95-365	395-536	
GFZL01027163.1	<i>CDPK17</i>	<i>CDPK17</i>	Pintec <i>CDPK6-17</i>	64-325	364-504	Tres sitios de unión a Ca ²⁺
GFZL01023164.1	<i>CDPK17</i>	<i>CDPK17</i>	Pintec <i>CDPK6-17-2</i>	66-325	363-504	
GFZL01021794.1	<i>CDPK17</i>	<i>CDPK15</i>	Pintec <i>CDPK6-17-3</i>	1-215	235-377	
GFZL01022062.1	<i>CDPK7</i>	<i>CDPK7</i>	Pintec <i>CDPK7</i>	58-327	344-513	
GFZL01010682.1	<i>CDPK30</i>	<i>CDPK3</i>	Pintec <i>CDPK10</i>	43-331	337-515	Tres sitios de unión a Ca ²⁺
GFZL01009033.1	<i>CDPK13</i>	<i>CDPK13</i>	Pintec <i>CDPK13</i>	60-350	349-517	Tres sitios de unión a Ca ²⁺
GFZL01002487.1	<i>CDPK16</i>	<i>CDPK16</i>	Pintec <i>CDPK16</i>	78-355	372-523	
GFZL01007699.1	<i>CDPK17</i>	<i>CDPK17</i>	Pintec <i>CDPK17</i>	92-383	402-546	
GFZL01003864.1	<i>CDPK16</i>	<i>CDPK16</i>	Pintec <i>CDPK28</i>	54-334	343-501	

TablaAnexa 11: *Sequoia sempervirens*

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GIBU01318870.1	<i>CDPK1</i>	<i>CDPK17</i>	Seqsem <i>CDPK1</i>	141-436	452-597	
GIBU01318872.1	<i>CDPK2</i>	<i>CDPK13</i>	Seqsem <i>CDPK2</i>	82-377	393-532	
GIBU01146325.1	<i>CDPK17</i>	<i>CDPK2</i>	Seqsem <i>CDPK6-17</i>	46-344	361-502	
GIBU01146326.1	<i>CDPK6</i>	<i>CDPK1</i>	Seqsem <i>CDPK6-low</i>	30-154	171-312	
GIBU01187568.1	<i>CDPK7</i>	<i>CDPK7</i>	Seqsem <i>CDPK7</i>	60-330	349-527	
GIBU01327310.1	<i>CDPK7/30</i>	<i>CDPK20</i>	Seqsem <i>CDPK10</i>	48-317	333-502	
GIBU01249822.1	<i>CDPK16</i>	<i>CDPK16</i>	Seqsem <i>CDPK16</i>	70-349	365-517	
GIBU01249824.1	<i>CDPK16</i>	<i>CDPK16</i>	Seqsem <i>CDPK16-2</i>	18-205	216-368	
GIBU01249828.1	<i>CDPK16</i>	<i>CDPK16</i>	Seqsem <i>CDPK28</i>	55-329	342-499	

TablaAnexa 12: *Taxus baccata*

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GGKG01002503.1	<i>CDPK1</i>	<i>CDPK1</i>	Taxbac <i>CDPK1</i>	146-443	459-605	
GGKG01004005.1	<i>CDPK2</i>	<i>CDPK13</i>	Taxbac <i>CDPK2</i>	102-371	400-539	
GGKG01002667.1	<i>CDPK17</i>	<i>CDPK17</i>	Taxbac <i>CDPK6-17</i>	47-346	359-501	
GGKG01001780.1	<i>CDPK7</i>	<i>CDPK20</i>	Taxbac <i>CDPK7</i>	58-330	349-527	
GGKG01001176.1	<i>CDPK32</i>	<i>CDPK7</i>	Taxbac <i>CDPK10</i>	47-316	332-501	
GGKG01004664.1	<i>CDPK17</i>	<i>CDPK17</i>	Taxbac <i>CDPK17</i>	69-367	384-526	
GGKG01001090.1	<i>CDPK28</i>	<i>CDPK28</i>	Taxbac <i>CDPK28</i>	55-328	342-499	

TablaAnexa 13: *Thuja occidentalis*

Locus	Araport11	SwissProt	Gen	Cinasa	EF-hand	Notas
GGJL01003948.1	<i>CDPK1</i>	<i>CDPK17</i>	Thujoc <i>CDPK1</i>	144-439	455-597	
GGJL01008098.1	<i>CDPK2</i>	<i>CDPK13</i>	Thujoc <i>CDPK2</i>	95-362	393-532	
GGJL01001987.1	<i>CDPK5</i>	<i>CDPK4</i>	Thujoc <i>CDPK5</i>	112-410	425-565	
GGJL01000707.1	<i>CDPK17</i>	<i>CDPK2</i>	Thujoc <i>CDPK6-17</i>	64-328	362-503	
GGJL01017095.1	<i>CDPK17</i>	<i>CDPK17</i>	Thujoc <i>CDPK6-17-2</i>	82-349	381-522	
GGJL01000257.1	<i>CDPK7</i>	<i>CDPK20/7</i>	Thujoc <i>CDPK7</i>	58-329	349-527	
GGJL01006977.1	<i>CDPK7/30</i>	<i>CDPK7</i>	Thujoc <i>CDPK10</i>	48-316	333-502	
GGJL01007280.1	<i>CDPK13</i>	<i>CDPK13</i>	Thujoc <i>CDPK13</i>	45-339	335-504	Tres sitios de unión a Ca ²⁺
GGJL01001937.1	<i>CDPK16</i>	<i>CDPK16</i>	Thujoc <i>CDPK16</i>	69-347	363-515	
GGJL01002979.1	<i>CDPK17</i>	<i>CDPK17</i>	Thujoc <i>CDPK17</i>	83-373	391-532	
GGJL01008972.1	<i>CDPK28</i>	<i>CDPK28</i>	Thujoc <i>CDPK28</i>	50-336	342-499	

Anexo 2. Resultados del análisis MEME para detectar sitios bajo selección en los genes CDPK de coníferas. Se indican para cada gen, los sitios que mostraron señal de selección positiva, y se resaltan en rojo las ramas de la filogenia en las que se detectaron dichas señales. El gen *CDPK16* no presentó sitios bajo selección en este análisis.

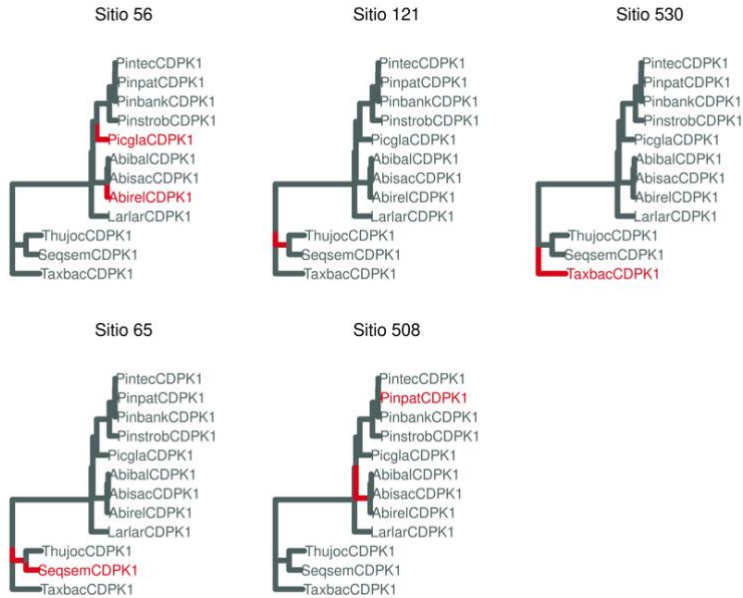


Figura Anexa 1. CDPK1

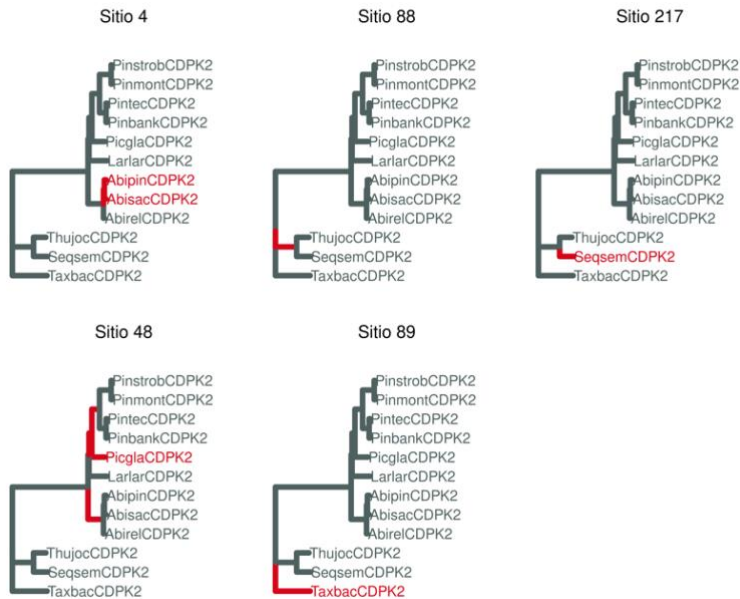


Figura Anexa 2. CDPK2

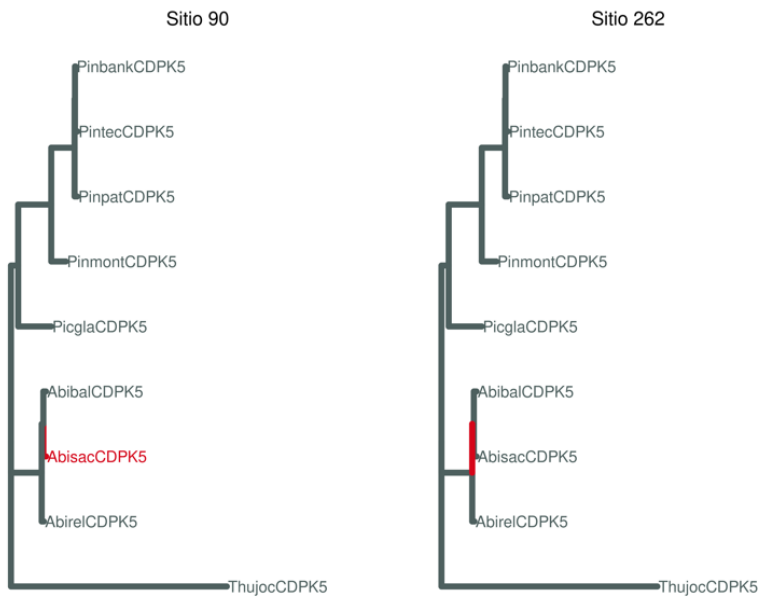


Figura Anexa 3. CDPK5

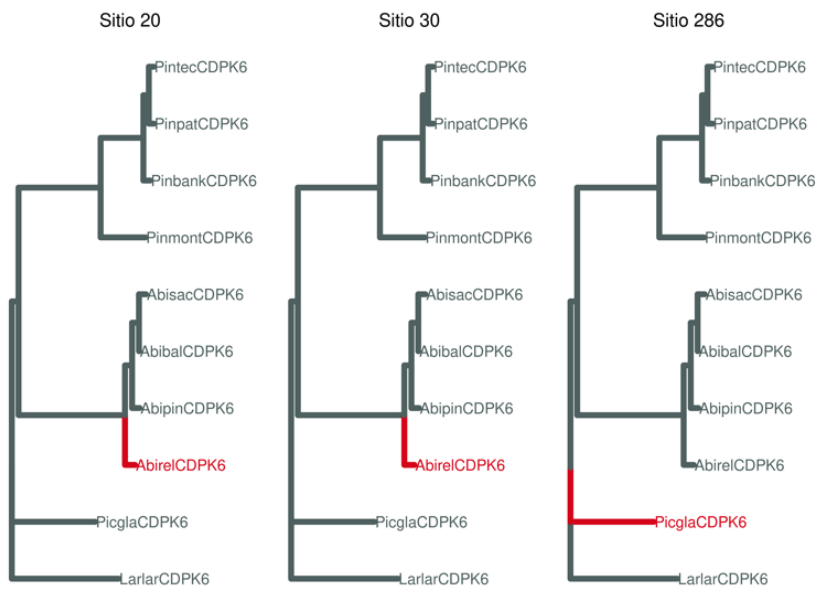


Figura Anexa 4. CDPK6

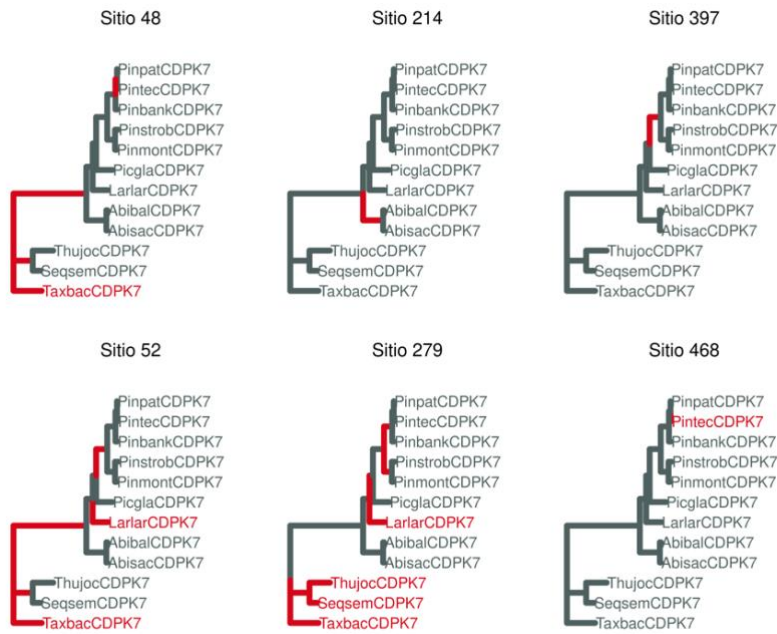


Figura Anexa 5. CDPK7

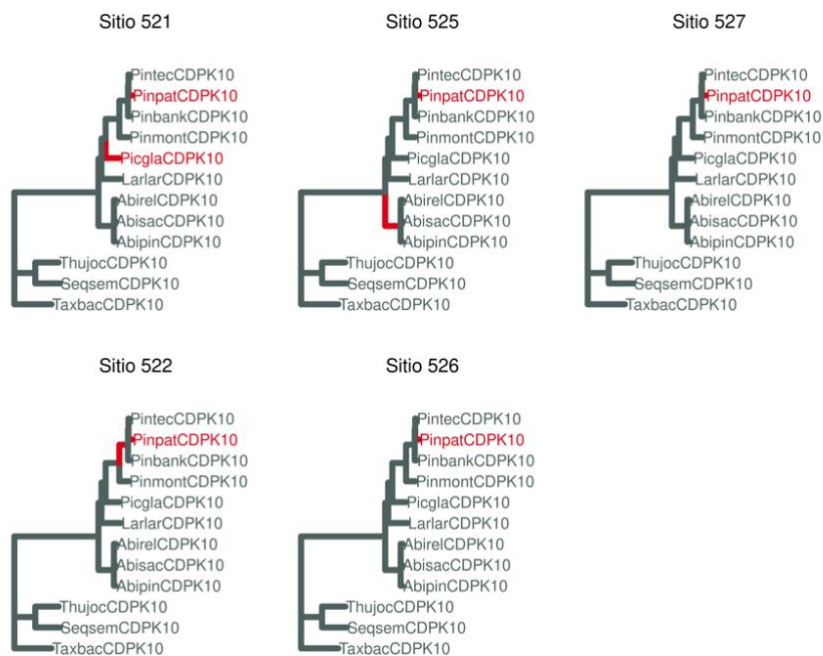


Figura Anexa 6. CDPK10

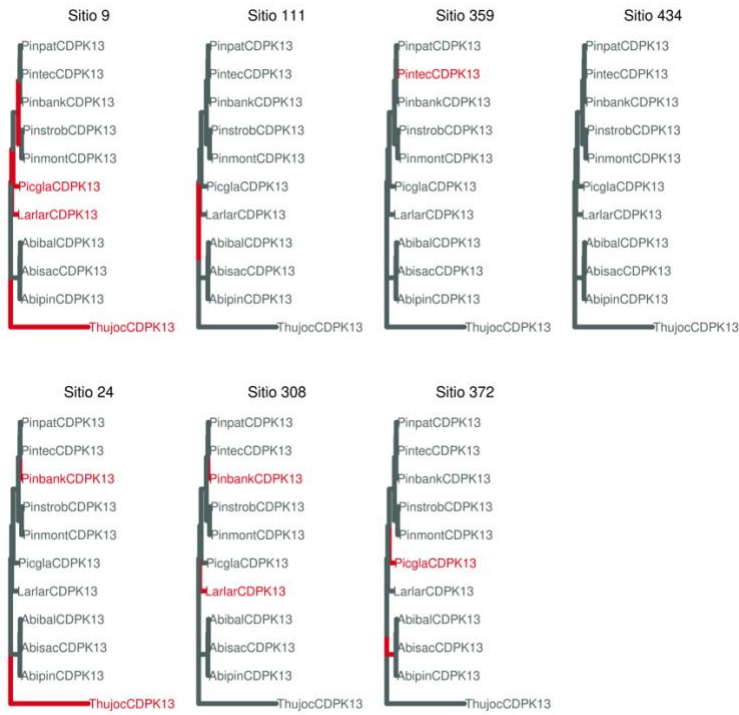


Figura Anexa 7. CDPK13

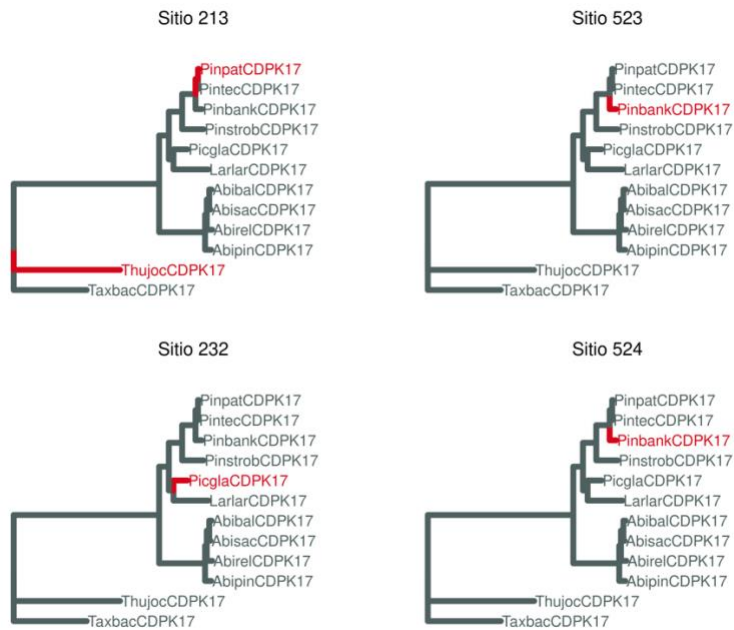


Figura Anexa 8. CDPK17

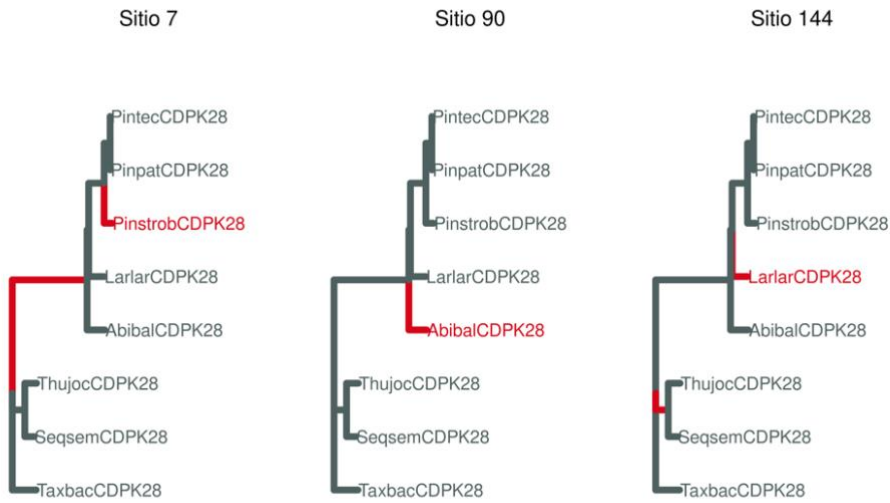


Figura Anexa 9. CDPK28

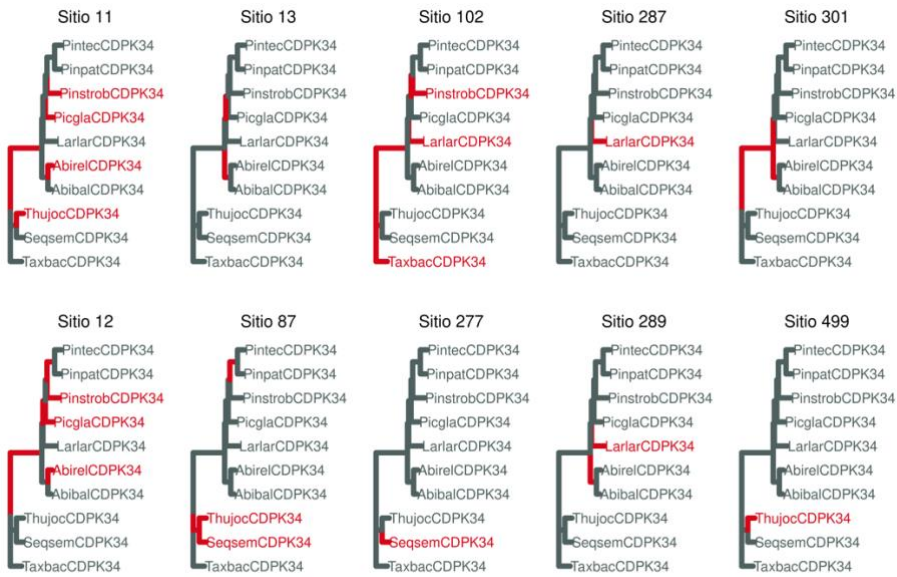


Figura Anexa 10. CDPK34

Anexo 3. Resultados del análisis aBSREL para detectar los linajes de coníferas en los que los genes CDPK estarían bajo selección positiva. Las ramas de las filogenias están coloreadas y re-escaladas de acuerdo con sus coeficientes ω (dn/ds). Solo las ramas más gruesas resultaron significativas después de la corrección de Holm-Bonferroni.

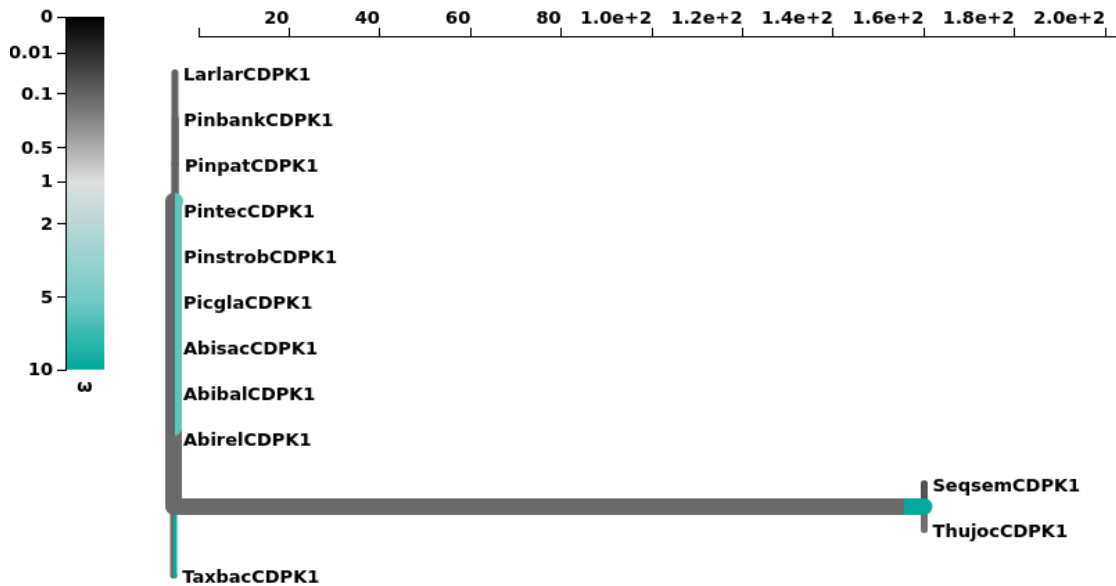


Figura Anexa 11. CDPK1

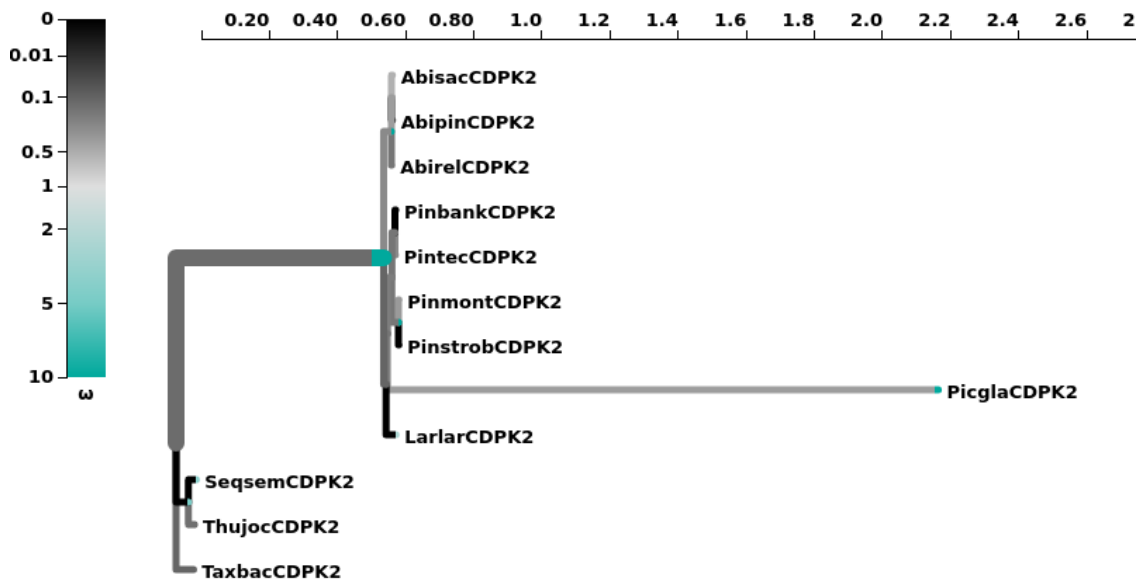


Figura Anexa 12. CDPK2

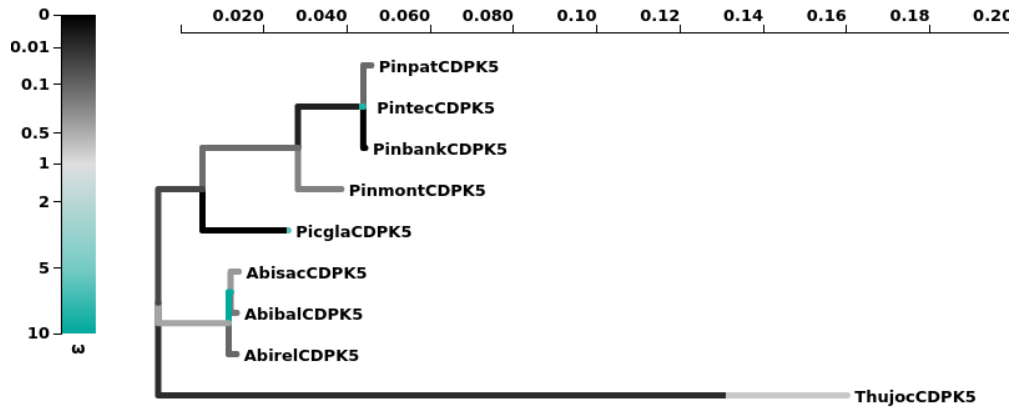


Figura Anexa 13. CDPK5

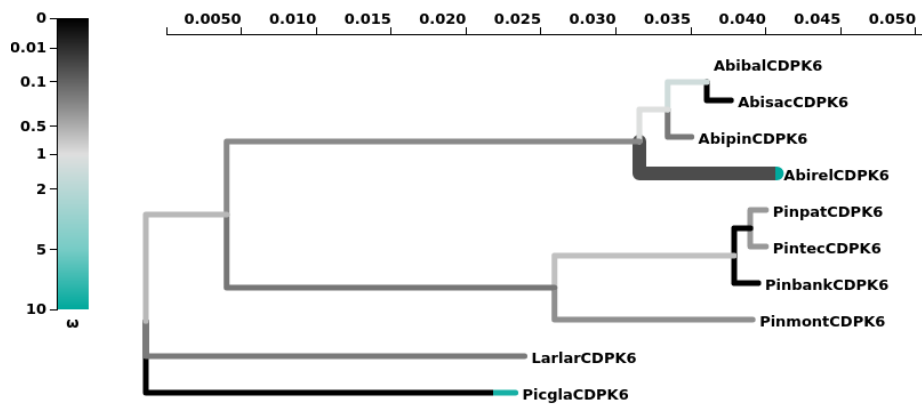


Figura Anexa 14. CDPK6

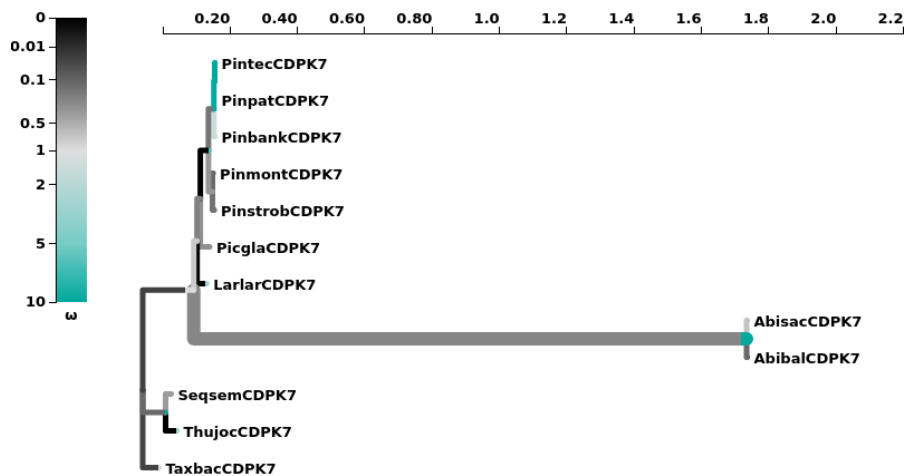


Figura Anexa 15. CDPK7

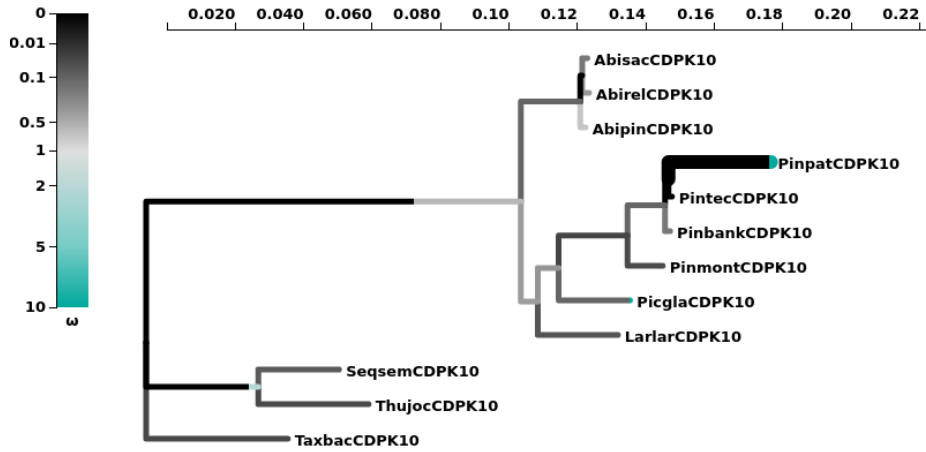


Figura Anexa 16. CDPK10

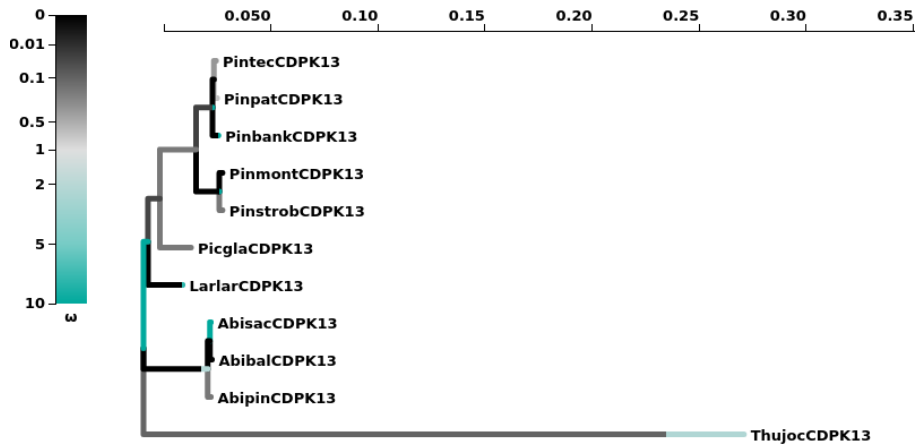


Figura Anexa 17. CDPK13

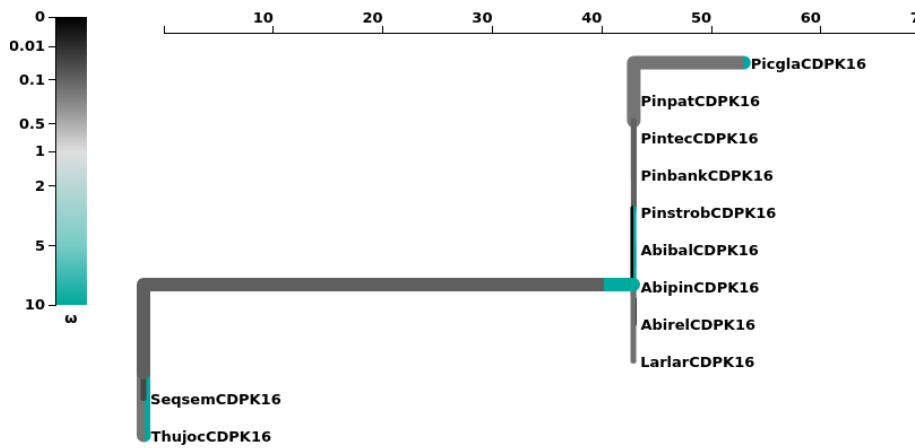


Figura Anexa 18. CDPK16

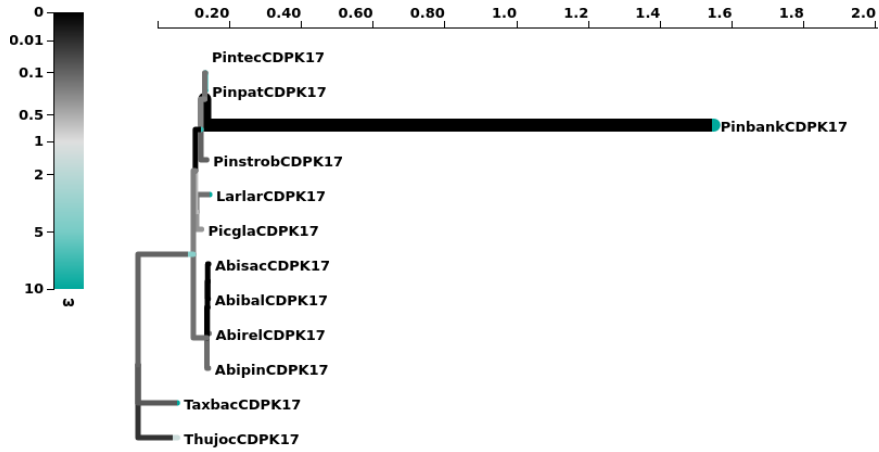


Figura Anexa 19. CDPK17

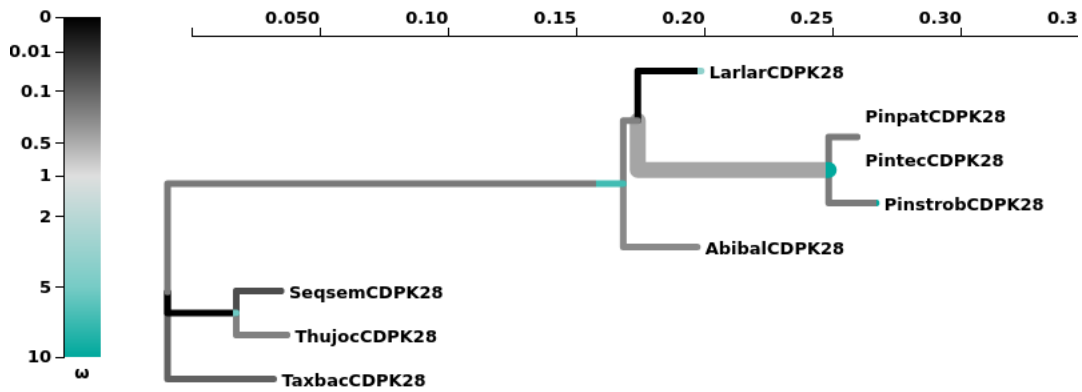


Figura Anexa 20. CDPK28

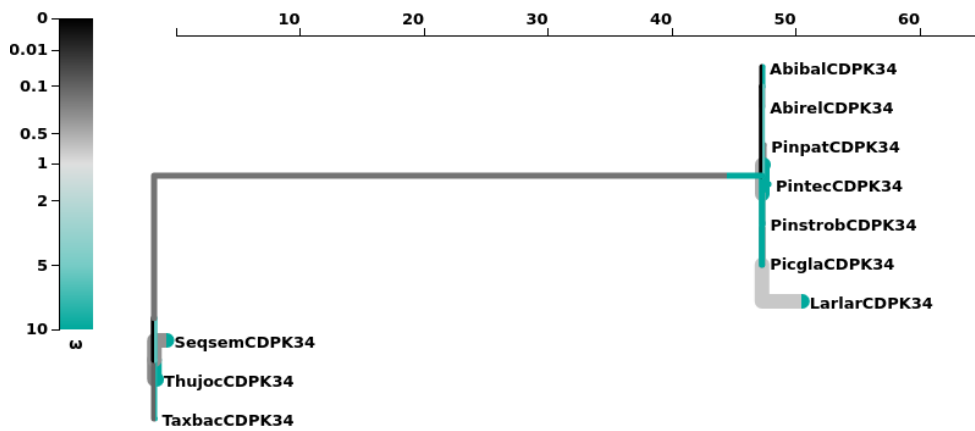


Figura Anexa 21. CDPK34

Anexo 4. Resultados del análisis PROVEAN para detectar sustituciones deletéreas en el gen CDPK34 de Pináceas y su posición aproximada en los diferentes dominios de este gen. Las sustituciones se indican en formato HGVS.

Sustitución	Score		Dominio
P11S	-0.512	Neutral	variable
P11R	-0.841	Neutral	variable
P11G	-1.623	Neutral	variable
P11Q	-0.192	Neutral	variable
E12S	0.114	Neutral	variable
E12D	0.06	Neutral	variable
X13N	1.042	Neutral	variable
X13Y	-0.178	Neutral	variable
X13S	1.416	Neutral	variable
H87L	-1.723	Neutral	kinase
H87K	0.56	Neutral	kinase
H87G	6.197	Neutral	kinase
T102R	-1.33	Neutral	kinase
T102N	0.378	Neutral	kinase
T102S	0.399	Neutral	kinase
T102G	-1.855	Neutral	kinase
L277Q	-5.547	Deleterious	kinase
P287V	-8.8	Deleterious	kinase
P289K	-7.529	Deleterious	kinase
P289T	-7.534	Deleterious	kinase
N301S	-0.217	Neutral	kinase/autoregulatory
N301W	-3.052	Deleterious	kinase/autoregulatory
C499A	3.46	Neutral	EF-hand