



UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE CIENCIAS FORESTALES
INGENIERÍA EN BIOTECNOLOGÍA VEGETAL

**CARACTERIZACIÓN DE LA REGIÓN CODIFICANTE DE GENES
DEHIDRINA DESDE UNA COLECCIÓN DE SECUENCIAS RNASEQ
DE *PRUMNOPITYS ANDINA***

Tesis presentada a la Facultad de Ciencias Forestales de la Universidad de
Concepción para otorgar al título profesional de Ingeniero en Biotecnología
Vegetal

Por: Fernando Hugo Karim Muñoz Herrera
Profesora Guía: Marta Alejandra Fernández Reyes

Octubre, 2024
Concepción, Chile

© 2024, Fernando Hugo Karim Muñoz Herrera

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento

**CARACTERIZACIÓN DE LA REGIÓN CODIFICANTE DE GENES DEHIDRINA
DESDE UNA COLECCIÓN DE SECUENCIAS RNASEQ DE *PRUMNOPITYS*
*ANDINA***



Profesora Guía

Marta Alejandra Fernández Reyes

Profesora Asociada

Biólogo, PhD.



Profesora Guía

Martha Hernández de la Torre

Profesora Asociada

Ingeniero Químico, PhD.

AGRADECIMIENTOS

Quiero expresar mi más profundo agradecimiento a mi familia, cuya constante ayuda y apoyo han sido esenciales para la realización de esta tesis. A mi madre, Carmen Gloria Herrera Umaña, quien con su amor incondicional y sabios consejos siempre ha sido mi guía y mi inspiración. A mi abuela, María Petronila Umaña Jara, por su fortaleza y sabiduría, que han sido un faro en mi vida. Y a mis hermanas, Annais Muñoz y Valentina Muñoz, por su constante apoyo, ánimo y comprensión a lo largo de este arduo camino académico. Sin ustedes, nada de esto habría sido posible. Este logro es tanto mío como de ustedes.

Un agradecimiento especial a mi profesora guía, Marta Fernández Reyes, por su invaluable orientación, paciencia y dedicación durante todo el proceso de esta tesis. Su apoyo ha sido fundamental para el éxito de este trabajo.

También quiero agradecer de manera muy especial a mi prima, Lizbeth Coral Velásquez Ramos, Bióloga de la Universidad de Concepción, por estar siempre dispuesta a ayudarme y apoyarme en cada etapa de esta investigación. Su ayuda ha sido invaluable y siempre la recordaré con gratitud.

TABLA DE CONTENIDO

AGRADECIMIENTOS.....	iv
TABLA DE CONTENIDO	v
ÍNDICE DE TABLAS	vi
ÍNDICE DE FIGURAS.....	vii
RESUMEN.....	ix
ABSTRACT	x
I. INTRODUCCIÓN	1
HIPÓTESIS	10
OBJETIVO GENERAL.....	11
OBJETIVOS ESPECÍFICOS	11
II. METODOLOGÍA.....	12
III. RESULTADOS.....	17
IV. DISCUSIÓN	46
V. CONCLUSIONES.....	52
VI. GLOSARIO.....	54
VII. BIBLIOGRAFÍA.....	57

ÍNDICE DE TABLAS

Tabla 3.1. Secuencias putativas de cada proteína	17
Tabla 3.2. Resumen de las características bioquímicas de las secuencias.	41
Tabla 3.3. Porcentaje del índice de desordenamiento de la estructura secundaria de las secuencias deducidas.	42
Tabla 3.4. Resumen del tipo de configuración de cada proteína.	42

ÍNDICE DE FIGURAS

Figura 3.1. Alineamiento de las 9 secuencias deducidas hecho en Bioedit con la herramienta ClustalW de Bioedit.....	19
Figura 3.2. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN1. Enmarcadas rojas se detallan el segmento S y el K.	20
Figura 3.3. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN2. Enmarcadas rojas se detallan el segmento S y K.	21
Figura 3.4. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN3. Enmarcadas rojas se detallan el segmento F, S y K.	22
Figura 3.5. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN4. Enmarcadas rojas se detallan el segmento F, S y K.	23
Figura 3.6. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN5. Enmarcadas en color rojo se detalla el segmento F, S y K.	25
Figura 3.7. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN6. Enmarcadas en color rojo se detallan los segmentos K1 y K2.	26
Figura 3.8. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN7. Enmarcadas en color rojo se detalla el segmento F, S, K1 y K2.	27
Figura 3.9. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN8. Enmarcadas en rojo se detallan el segmento Y1, Y2, S y K.	28
Figura 3.10. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN9. Enmarcadas en rojo se detallan el segmento Y, S y K.	29
Figura 3.11. Logo de los motivos conservados encontrados en las proteínas deducidas junto a sus homólogos, además aparece el valor de significancia, numero de sitios donde se encontró el motivo y el ancho de este.	34
Figura 3.12. Representación de PrADHN1. En color rosado, los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados verdes, y los sitios de fosforilación de la caseína quinasa II están enmarcados negros, además de la estructura secundaria bajo los aminoácidos que la componen.	36
Figura 3.13. Representación de PrADHN2. Coloreado en rosado está enmarcado los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, mientras que los sitios de fosforilación de la caseína quinasa II están enmarcados en color negro, los sitios de fosforilación de la quinasa C están enmarcados con la línea negra discontinua y el sitio de glicosilación se representa enmarcado en color naranja, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.	37
Figura 3.14. Representación de PrADHN3, en rosado están los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están en verde, mientras que los sitios de fosforilación de la caseína quinasa II están en negro, y la estructura secundaria está bajo los aminoácidos que la componen.	37
Figura 3.15. Representación de PrADHN4. Coloreado en rosado está enmarcado los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, los sitios de fosforilación de la quinasa C están enmarcados con la línea negra discontinua, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.	38
Figura 3.16. Representación de PrADHN5. Coloreado en rosado está enmarcado los segmentos conservados con su respectiva letra en la parte superior, los sitios de	

amidación están enmarcados en color verde, mientras que los sitios de fosforilación de la caseína quinasa II están enmarcados en color negro, los sitios de fosforilación de la tirosina quinasa están enmarcados con la línea negra discontinua, los sitios de miristoilación están enmarcados en color rojo y finalmente el sitio de glicosilación enmarcado en color naranja, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.38

Figura 3.17. Representación de PrADHN6. Coloreado en rosado está enmarcado los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, mientras que el sitio de fosforilación de la quinasa C está enmarcado con la línea negra discontinua y los sitios de miristoilación están enmarcados en color rojo, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.39

Figura 3.18. Representación de PrADHN7. Coloreado en rosado está enmarcado los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, mientras que los sitios de fosforilación de la caseína quinasa II y los sitios de miristoilación están enmarcados en color rojo, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.39

Figura 3.19. Representación de PrADHN8. Coloreado en rosado está enmarcado los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, mientras que los sitios de fosforilación de la caseína quinasa II y los sitios de miristoilación están enmarcados en color rojo, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.40

Figura 3.20. Representación de PrADHN9. Coloreado en rosado está enmarcado los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, mientras que los sitios de fosforilación de la caseína quinasa II y los sitios de miristoilación están enmarcados en color rojo, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.40

Figura 3.21. Árbol filogenético de proteínas dehidrina homólogas a las proteínas deducidas en *Prumnopytis andina*. Al lado derecho de la imagen están los segmentos conservados, donde el color morado representa el segmento Y, el amarillo el segmento F, el celeste el segmento S y el verde el segmento K. Los segmentos se alinean con su respectivo código y especie de la proteína.45

RESUMEN

Prumnopitys andina o Lleuque, conífera endémica del sur de Chile, ha sido escasamente investigada en sus mecanismos de tolerancia y/o resistencia frente a factores abióticos. Las deshidrinas son proteínas de respuesta a estrés ambiental, que han sido estudiadas en diversas especies vegetales. Esta investigación se centró en caracterizar proteínas dehidrinas en una colección de secuencias transcriptómicas disponible en la base de datos NCBI para *P. andina*. Se logró identificar nueve secuencias codificantes para proteínas homólogas a deshidrinas de especies previamente estudiadas, para luego caracterizar su secuencia genómica e identificar el tipo de deshidrinas presente en *P. andina*, caracterizar sus propiedades bioquímicas, compararlas con otras especies de coníferas, en cuanto a estructura primaria, secundarias y dominios funcionales, para entender la evolución de estas secuencias. Entre estas secuencias identificadas, la proteína COR47 mostró la mayor similitud con las deshidrinas homólogas en *P. andina*. Este estudio busca contribuir al conocimiento de los mecanismos de adaptación de *P. andina* frente a condiciones ambientales adversas, proporcionando una base para investigaciones futuras relacionadas con la función y regulación de las deshidrinas en esta especie.

Palabras claves: *Dehidrina, Lleuque, NCBI,*

ABSTRACT

Prumnopitys andina an endemic conifer from southern Chile, has been sparsely investigated regarding its tolerance or resistance mechanisms against various abiotic factors. This research focuses on gathering information about dehydrins in conifers and their presence in the reference transcriptome available in the NCBI database. Orthologs amino acid sequences to dehydrins from previously studied species were obtained to characterize these sequences and identify the type of dehydrin present in the *P. andina* transcriptome, as well as deduce the biochemical properties of these proteins. Additionally, phylogenetic analyses were conducted to understand the evolution of these sequences, along with the analysis of secondary structure. As a result, ten sequences that could encode for a dehydrin were identified, as they exhibit their distinctive features. Among these sequences, COR47 showed the highest similarity to Orthologs dehydrins and possessed characteristics similar to those already described. To further this research, laboratory assays are needed to verify the presence of these proteins in the species' tissues. This study aims to contribute to the understanding of *P. andina* adaptation mechanisms to adverse environmental conditions, such as cold and dehydration. Additionally, it provides a foundation for future research on the function and regulation of dehydrins in this understudied conifer species.

I. INTRODUCCIÓN

1.1. *Prumnopitys andina* una Podocarpácea chilena

Las coníferas datan del carbonífero, desarrollándose en el triásico tardío o jurásico temprano, hace unos 300 millones de años, siendo consideradas un componente ecológico crucial de los bosques mundiales (Farjon 2018). En Chile, las coníferas están representadas en tres familias: Araucariaceae, Cupressaceae y Podocarpaceae (MMA 2018), esta última es la segunda familia de coníferas más rica en especies, así como el clado más grande de coníferas del sur, comprendiendo 19 géneros y 187 especies (Christenhusz y Byng 2016), distribuidas por el Sur del pacífico, Nueva Zelanda, Nueva Caledonia, Este de Australia y América del Sur, habitando bosques húmedos (Álvarez *et al.* 2021), donde en Chile, la familia Podocarpácea está representada por nueve especies, presentando dos especies endémicas: *Prumnopitys saligna* y *Prumnopitys andina* (MMA 2018).

Prumnopitys andina (Poepp. Ex Endl.) de Laub , o de nombre vulgar Lleuque, es nativa de los bosques del centro Sur de Chile y de los Andes adyacentes de Argentina (Rodríguez 2004), en Chile específicamente su área de distribución principal es la precordillera andina entre las provincias de Linares y Cautín, con una población en la Cordillera de la Costa, ubicada en la vertiente oriental de la cordillera de Nahuelbuta, cercana a Angol (García 2008).

P. andina alcanza los 30 metros de altura y diámetros de hasta un metro, de copa ramosa, de aspecto piramidal o redondeada, su tronco es liso y la corteza es delgada, de coloración gris brillante a ligeramente azulada, es un árbol siempreverde, dioico y rara vez monoico (Donoso 2006), posee conos masculinos en espigas y conos femeninos con un fruto carnoso llamado “uva de cordillera”, de dos a tres cm de largo, amarillos cuando maduros, de sabor dulce, su floración es en noviembre, de semillas ovoides y recalcitrantes las cuales maduran entre enero y marzo (Hechenleitner 2005).

Esta especie produce conos de semillas oblongos, amarillos, donde cada óvulo está rodeado por una bráctea fértil, posee una gran exo-testa entre 16 a 22 capas de células, exo-testa carnosa, con cubierta seminal carnosa similar a la sarcotesta, una meso y endo-testa con células parenquimáticas y esclerificadas, posee una capa de células las cuales constituyen a una cubierta protectora del embrión (Khan *et al.* 2022).

El Lleuque está catalogado como una especie “Rara” en el Libro Rojo de la Flora Terrestre de Chile (Benoit 1989), siendo considerada como una especie con problemas de conservación (Benoit 1989), clasificándose como Vulnerable (A2cd) por la Unión Internacional para la Conservación de la Naturaleza (UICN) (Hernandez 2012, MMA 2012), por esto se prohíbe la corta, eliminación,

destrucción o descepado, al igual que la alteración de su hábitat (artículo 19, Ley N°20.283). Actualmente, las poblaciones costeras están severamente amenazadas por el pastoreo de ganado y la forestación con *Pinus radiata* y *Eucaliptus* sp., otras poblaciones han sido reducidas por acciones antropológicas (esquemas hidrológicos, tala ilegal y recolección del fruto) o por actividad volcánica, como el caso del Parque Nacional Conguillio (Cano *et al.* 2022).

1.2. La capacidad de las plantas para adaptarse a su entorno: el caso de la dehidrina.

La capacidad de sobrevivir y crecer de las plantas leñosas perennes en climas fríos es de vital importancia para la silvicultura (Wei *et al.* 2019), las especies de plantas leñosas de hoja perenne pueden adaptarse al duro invierno a través de un proceso llamado aclimatación al frío (AC), lo que se traduce como la tolerancia a las bajas temperaturas y heladas estacionales, aumentando su resistencia en otoño, llegando al punto máximo en invierno y alcanzando su nivel más bajo en verano (Arora y Taulavuori 2016).

Las plantas son organismos poiquilotermos, es decir, asumen la temperatura de su entorno. Se describen dos efectos causados por estrés a bajas temperaturas, asociados al frío y congelamiento (Hale 1987), la tolerancia a la congelación es la capacidad que tienen las plantas de resistir la formación de hielo extracelular y evitar la formación de hielo intracelular (Zamecnik 1992), por consecuencia de

la congelación extracelular se produce una deshidratación, dada la eliminación de agua del citoplasma hacia los cristales de hielo, produciendo un efecto de liofilización (Jarillo 1994). Para evitar la congelación, las plantas producen varios compuestos logrando proteger las células contra la formación de hielo intracelular e intercelular, acumulando azúcares, aminoácidos y compuestos anticongelantes (Thomashow 1998, Ewart 1999, Thomashow 2001, Yu y Griffith 2001).

Para lograr la tolerancia al congelamiento, las plantas se aclimatan a las temperaturas frías, acumulando proteínas cuya síntesis aumenta a bajas temperaturas (Jarillo 1994, Wisniewski *et al.* 2002), como las proteínas de la familia de *genes de embriogénesis tardía abundante* (LEA), descubiertos originalmente en *Gossypium hirsutum* durante la embriogénesis tardía, periodo que es marcado por la desecación de semillas (Galau 1986), del mismo modo se han encontrado proteínas LEA en tejidos vegetativos desecados (Tunnacliffe y Wise 2007).

La expresión regulada al alza de una proteína LEA ocurre en respuesta a al menos una de las tres influencias abióticas: sequía, frío y estrés osmótico, procesos que conducen al agotamiento del agua intracelular (Close 1996). El estrés abiótico es el causante de daño celular, desencadenar el estrés osmótico, y posteriormente, estrés oxidativo, que a su vez conduce a la activación del mecanismo de defensa de las plantas, la respuesta que tienen las plantas al

estrés es la modulación del metabolismo celular y producción de proteínas, incluidas las dehidrinas (Karas *et al.* 2024). Los genomas de las plantas a menudo codifican varias familias de proteínas LEA, que no necesariamente están asociadas a estrés abiótico, siendo exclusivas en etapas específicas del crecimiento (Nylander 2001). El crecimiento y la germinación de las plantas mejoran en condiciones de estrés en las plantas que sobre expresan proteínas LEA (Cheng Z 2002).

La dehidrina es el nombre de las proteínas LEA (Proteínas abundantes en la embriogénesis tardía) pertenecientes al grupo dos (Amara *et al.* 2014), son proteínas altamente hidrófilas y ricas en glicina y lisina, además, pueden retener gran cantidad de agua, protegiendo biomoléculas y estructuras de la membrana celular frente a la pérdida de agua (Sun *et al.* 2021). Las dehidrinas son proteínas intrínsecamente desordenadas (IDP), es decir, que poseen muy poca estructura secundaria, sin estructura terciaria, lo que las hace muy flexibles en solución (Smith y Graether 2022). Al no tener estructura definida no pierden su integridad en condiciones de desnaturalización (Dyson 2002), pudiendo interactuar con los fosfatos de la membrana como también con las proteínas presentes en ella. Al sufrir una deshidratación celular, la membrana queda en forma hexagonal invertida (Steponkus 1998), comprometiendo permanentemente su integridad y permeabilidad, conduciendo a un daño oxidativo de lípidos, proteínas y ADN (Gordon-Kamm y Steponkus 1984). Smith y Graether (2022) demostraron como

las dehidrinas pueden proteger la membrana reduciendo la temperatura de transición (T_m) de diversas composiciones de la membrana, la disminución de la T_m permitiría que la membrana mantuviera su fluidez a una temperatura ambiental más baja, lo que le permitiría funcionar normalmente a pesar del frío y evitar la transición de fase.

Las dehidrinas están caracterizadas por los motivos conservados conocidos como segmentos Y, S, F y K (Dure 1989). Cada segmento lleva el nombre de los residuos conservados dentro del motivo. El segmento Y tiene un residuo aromático central que a menudo es Tirosina, el segmento S tiene de cinco a siete residuos de Serina seguidos, el segmento F contiene uno o dos residuos de Fenilalanina, y el segmento K que es rico en Lisina (Smith y Graether 2022).

Por definición una dehidrina debe contener al menos un segmento K. Este se describe mediante la secuencia canónica EKKGIMDKIKEKLPG, aunque un análisis detallado del segmento K sugiere que puede describirse con mayor precisión como XKXGXX(D/E)KIK(D/E)KXPG (Malik *et al.* 2017). En general, la función del segmento K es la protección de las membranas contra el daño por estrés por frío, así como la crioprotección enzimática (Smith y Graether 2022). De manera detallada, el segmento K se puede unir a las superficies de la membrana de fosfolípidos, creando una alfa hélice anfipática y así estabilizar la membrana mediante un efecto protector (Hara 2010). (Hernández Camacho 2017) examinó

la dehidrina Lti30, que tiene dipéptidos de histidina, flanqueando cinco de sus seis segmentos K, cuando estos dipéptidos fueron desprotonados la unión fue abolida, demostrando que los residuos que flanquean el segmento K son importantes para la unión a la membrana en ciertas dehidrinas. Gupta et al. (2019) demostraron que existe una correlación positiva entre la longitud de la hélice, el número de contacto con la membrana y la profundidad a la que se encuentra la hélice dentro de la membrana.

Por su parte, el segmento F recibe su nombre por la presencia de dos residuos de fenilalanina cerca de la mitad del motivo (Strimbeck 2017). Un análisis de 426 dehidrinas sugiere que este motivo podría mostrarse como EXXDRGXFDX(G/K), donde la séptima y décima posición son aminoácidos hidrofóbicos (Riley et al. 2019); sin embargo, aún no se ha descrito una función detallada para este motivo, aunque puede tener un papel en la crio protección enzimática (Smith y Graether 2022).

El segmento Y, al igual que el segmento K; generalmente se describe mediante una sola secuencia: DEYGNP, pero podría describirse mejor como: D(D/E)(Y/H/F)GNPX, donde X suele corresponder a un aminoácido hidrofóbico. El papel bioquímico del segmento Y aun no se ha determinado, teniendo algún propósito en las semillas recubiertas (Riley et al. 2019). Además, un análisis de

similitud de secuencia entre el segmento Y con los dominios de unión de nucleótidos de las chaperonas, arrojo un resultado positivo (Close 1996).

Finalmente, el segmento S se define por la presencia de cinco a siete residuos continuos de Serina, pero se ha demostrado que el motivo contiene otros residuos conservados, y describiendo su secuencia como LHR(S/T)GS(4-6)(S/D/E)(D/E)(3). No es sorprendente que el segmento S sea un objetivo para la fosforilación, probablemente por una cisteína quinasa 2, relacionada con SnF1 (SnRK2) (Maszkowska *et al.* 2019), este segmento suele sufrir una fosforilación, lo que conduce a la unión de iones de calcio (Alsheikh *et al.* 2003) y una translocación de desidrias al núcleo (Goday *et al.* 1994).

1.3. Deshidrias: Estrategias Bioinformáticas para su Identificación y Caracterización

Las desidrias son proteínas multifuncionales, con diversas funciones que cumplen con el objetivo de proteger la célula vegetal del estrés abiótico, siendo implicadas en la crioprotección de proteínas contra el estrés por congelación en varios estudios (Hughes y Graether 2011, Drira *et al.* 2013, Hughes *et al.* 2013, Shi *et al.* 2015, Hara *et al.* 2017, Palmer *et al.* 2019, Ohkubo *et al.* 2020). La enzima más utilizada es la Lactato Deshidrogenasa (LDH), una proteína lábil fría que se desnaturaliza y pierde casi toda su actividad después del estrés por congelación-descongelación (Melgar y Zelada 2021).

Existe interés científico en la comprensión de los segmentos conservados en la protección de LDH frente la congelación, se ha demostrado con un péptido de 15 residuos del segmento K que son esenciales los residuos hidrofóbicos, formando parches que se intercalan LDH, evitando que la LDH se agregue, generando una “hidratación hidrofóbica” en la superficie de la enzima (Farjon 2018). (Ohkubo *et al.* 2020) declararon que el segmento K junto con el F y sus aminoácidos hidrofóbicos, confieren actividad crioprotector a las dehidrinas de tipo FSKn, por su parte, Palmer *et al.* (2019) han demostrado que las características le atribuyen una actividad protectora son la naturaleza polar y desordenada de la dehidrina, más que el orden de los residuos.

En Chile, la conífera *P. andina* o Lleuque, se distribuye bajo un clima submediterráneo con condiciones déficit hídrico y bajas temperaturas (Abarzúa *et al.* 2020). En la base de datos NCBI existen 184 proteínas descritas para esta especie, donde ninguna de estas es una dehidrina. Con el avance de la biología molecular, junto con mejoras en los recursos computacionales, da lugar a una gran cantidad de datos biológicos con enfoques en el análisis para acelerar los avances en la biotecnología (Carvalho *et al.* 2019), por ello se desarrolló una estrategia bioinformática para determinar si *P. andina* presenta proteínas dehidrina, mediante el uso de la información genética presente en la biblioteca “SRA” (Sequence Read Archive), que contiene lecturas de secuenciación de transcritos, otorgando una referencia para la búsqueda de diferentes genes de

interés, utilizando la colección de genes que codifican a la proteína dehidrina en *P. andina*, permitiendo deducir la secuencia aminoacídica, realizar un análisis bioquímico de su composición, y describir las secuencias de los segmentos conservados, interpretando su posible utilidad bajo condiciones de estrés abiótico, particularmente bajo condiciones de baja temperatura.

Bajo este contexto, esta investigación busca poner a prueba la siguiente hipótesis:

HIPÓTESIS

La base de datos SRA posee dos bibliotecas de lecturas de secuenciación de secuencias expresadas de acícula de *P. andina*, de las cuales es posible ensamblar, identificar y analizar genes que se expresan en tejido foliar.-En base a los antecedentes mencionados, se plantea la siguiente hipótesis de trabajo:

En acículas de lleuque se expresan genes codificantes para proteínas dehidrinas del tipo FSK2, las cuales han sido relacionadas a la respuesta a estrés por congelación en otras especies de gimnospermas.

OBJETIVO GENERAL

Describir la secuencia transcriptómica de genes codificantes para proteínas dehidrina en *Prumnopitys andina* a partir de datos de secuenciación de segunda generación

OBJETIVOS ESPECÍFICOS

- 1) Comparar las secuencias codificantes para genes dehidrinados identificadas en la colección de lecturas secuenciación de *P. andina* con la secuencia aminoacídica de dehidrinados de otras especies vegetales.
- 2) Determinar grupos homólogos asociados a dominios dehidrinados específicos que han sido asociados a la respuesta a baja temperatura.

II. METODOLOGÍA

2.1. Búsqueda de proteínas dehidrinas en gimnospermas.

Al ser *P. andina* una especie gimnosperma, fue fundamental investigar la importancia y las funciones biológicas que cumplen las dehidrinas en este grupo de plantas. Para ello, se buscaron plantas con características similares a nuestro objetivo, filtrando la búsqueda por investigaciones que abordaran el rol de las dehidrinas en épocas de bajas temperaturas. Se utilizó la estrategia de búsqueda de minería de textos o “text mining”, campo interdisciplinario basado en la recuperación de información, datos, aprendizaje automático, estadística y la lingüística computacional (West y Bhattacharya 2016), encontrando los artículos: Chang *et al.* 2021, Perdiguero *et al.* 2014 y Deng Z 2006. Adicionalmente, se verificaron las dehidrinas presentes en coníferas a través de la base de datos NCBI, donde predominan las proteínas del género *Pinus*.

Para la recolección de datos, se realizó una búsqueda de proteínas dehidrina en la base de datos Uniprot (<https://www.uniprot.org>), utilizando los términos “dehydrin” y “Plant”, proporcionando todas las proteínas dehidrina descritas para plantas, incluyendo proteínas predichas o putativas. Los datos obtenidos se descargaron para realizar un BLASTx (Basic Local Alignment Search Tool), una herramienta que compara una base de datos de secuencias de aminoácidos con la biblioteca de secuenciación SRA de *P. andina*, descargada directamente

desde el NCBI (National Center for Biotechnology Information) con el código de acceso SRX9968514. Esta biblioteca contiene información de secuenciación de ARN ultra direccional NEBnext para Illumina, realizada por el Instituto de Botánica de la Academia de Ciencias de China. La base de datos contiene 24,208,655 lecturas con 7.3 Gb, publicada el 01-05-2021.

2.2 Proteínas DHN en la colección de lecturas de secuenciación de *P. andina*.^[OBJ]

Una vez que la herramienta BLASTx completó su tarea, se generó una tabla de Excel que contenía las posibles proteínas dehidrina presentes en la colección de secuencias de *P. andina*. Para el análisis de los datos, se eliminaron los datos duplicados, dejando solo las secuencias de proteínas que coincidían con las secuencias transcriptómicas de *P. andina*, permitiendo centrarse solo en las proteínas más representadas.

Con los códigos de las proteínas más representadas que mostraban similitud con las secuencias transcriptómicas de *P. andina*, se realizó una búsqueda de la secuencia de aminoácidos y la secuencia codificante de la proteína, permitiendo realizar un *BLASTn* local en la biblioteca SRA de *P. andina*. Los resultados se ensamblaron en el sitio web PRABI-Doua (<https://doua.prabi.fr/software/cap3>), cada ensamble resultante se verificó utilizando la herramienta del sitio web Bioinformatics (https://www.bioinformatics.org/sms2/orf_find.html).

Una vez confirmada la presencia de la huella dehidrina de manera visual, se realizó otro *BLASTn* local en la biblioteca SRA, esta vez alineando solo los

extremos del ensamblaje previo. Finalmente, se ensamblaron los resultados para llevarlos a la herramienta de lectura de marcos abiertos, obteniendo así la proteína putativa.

2.3 Alineamiento múltiple de secuencias dehidrina

Para el alineamiento múltiple de las secuencias deducidas, se realizó un BLASTp a cada proteína putativa, seleccionando secuencias descritas como dehidrina, que posean mejores puntajes de cobertura y porcentaje de identidad, seleccionando las cinco mejores, para luego descargarlas en un mismo archivo e incluir la dehidrina putativa, con lo cual se creó un archivo en formato *fasta*. Este archivo se utilizó para un alineamiento en *Bioedit*, en el cual se utilizó la herramienta *ClustalW*, donde se editaron las secuencias, marcando los sitios con identidad de un 70%.

2.4. Segmentos conservados de dehidrina y sitios de interacción

Para identificar las firmas o características distintivas de las proteínas predichas, se accedió a la plataforma Pfam de InterPro (<https://pfam-legacy.xfam.org/>). Esta herramienta proporciona los sitios conservados de la proteína y confirma si esta corresponde a una dehidrina, las cuales poseen diferentes sitios específicos con diversas funciones biológicas. Para conocer algunos sitios de amidación, fosforilación de la caseína quinasa II, fosforilación de la proteína quinasa C,

glicosilación, amidación y miristoilación, se consultó la base de datos Prosite (<https://prosite.expasy.org/>).

Con las proteínas putativas, se realizó un BLASTp de cada una de ellas, eligiendo las cinco con mejores puntajes de cobertura y mejores porcentajes de identidad, se descargaron, generando un archivo en formato fasta, donde se incluyó todas las proteínas que resultaron homologas, junto con las dehidrasas putativas de *P.andina*, eliminando las secuencias repetidas según sus códigos de búsqueda de NCBI, Luego, este archivo se analizó en la herramienta *MEMESuit* (<https://meme-suite.org/meme/tools/meme>), la cual permite realizar análisis de las secuencias basadas en motivos repetidos; modificando el número de motivos a encontrar, según los segmentos F, Y, S y K de las dehidrasa.

2.5. Caracterización de las propiedades bioquímicas a partir de la secuencia primaria

Para el análisis de las secuencias obtenidas como proteínas putativas, se emplearon diversas herramientas bioinformáticas para diferentes enfoques y objetivos. Para realizar un análisis bioquímico se utilizó la herramienta *ProtParam* (<https://web.expasy.org/protparam/>), la cual proporciona el punto isoeléctrico teórico, el peso molecular y el índice de hidrofobicidad (GRAVY), además de valores como el porcentaje de cada residuo y el total de aminoácidos presentes en la secuencia.

2.6 Análisis estructura secundaria

La estructura secundaria de las proteínas deducidas de la biblioteca SRA fue analizada utilizando el software Phyre2 ([PHYRE2 Protein Fold Recognition Server \(ic.ac.uk\)](http://www.ebi.ac.uk/Phyre2/)), esta herramienta proporciona información sobre la posible estructura tridimensional de las proteínas, además de ofrecer una representación visual del porcentaje de estructuras, como la hélice alfa o la lámina beta, incluyendo el porcentaje de desorden de los aminoácidos en la secuencia, una característica propia de las dehidrina.

2.7 Construcción del árbol filogenético

Para la construcción del árbol filogenético se utilizó el archivo en formato fasta creado en la sección 2.3, archivo con el cual se realizó un alineamiento múltiple de secuencias utilizando la herramienta online *Clustal Omega* de EMBI-EBI (<https://www.ebi.ac.uk/jdispatcher/msa/clustalo>), ajustando los parámetros para el formato de salida en *Pearson/FASTA*, obteniendo los resultados, los cuales fueron analizados en el servidor web *IQ-TREE* (<http://iqtree.cibiv.univie.ac.at/>), permitiendo realizar arboles filogenéticos rápidos y de máxima verosimilitud. El resultado es un árbol filogenético difícil de simplificar, por lo que se utilizó la herramienta *ETEtoolkit* (<http://etetoolkit.org/treeview/>), un visor de árboles filogenéticos, permitiendo obtener los resultados del árbol filogenético en formato PDF.

III. RESULTADOS

a. Obtención de las secuencias aminoacídicas

De la comparación entre las 2.816 proteínas dehidrina de la base de datos UNIPROT, junto con la colección de secuencias transcriptómicas de *P. andina* presente en NCBI, se obtuvo un total de 235.686 alineamientos, de los cuales 233 proteínas coincidieron con las secuencias de *P. andina*. De estas últimas las más representativas fueron: A0A9R0HTH9_SPIOL, Q3ZDL0_PICAB, Q3ZDK9_PICAB y Q2TLS1_GINBI, presentes en la base de datos UNIPROT, además del uso del código exclusivo de NCBI: XP_059075498.1. Con estos códigos se obtuvieron las secuencias aminoacídicas presentes en la tabla 3.1.

Tabla 3.1. Secuencias putativas de cada proteína

DHN	SECUENCIA AMINOACÍDICA
PRADHN1	MAEEVPPSSVPEQQDRALFGLFGKKKDETEHGSAPAHGGH EPQAAHVSAGGHGGHAAKLHRSDSSSSSSSDEEGEDGKKK KKGLKEKVKEKLPGHHGSSAEHHPPQPPEKKGLMEKIKEKLP GSGGH
PRADHN2	MADSNAPehtDRGFFGMFGKKDEEKHDQQSHAVDPHSGHA PAYTGEVHHAPPPAVAHQPQHYGQGTGGNAYAGEVHIPPP AVSHQPQHYGQGEGDGNKHPEGLKGKLHRSNSSSSSSSSSDE DESGGKKKEGRKKKGLKDKIKEKLPGHGGEKEPKEETEKKG LVDKIKEKLPGHKEGEKGH
PRADHN3	MADSGAPEHTDRGLFGLFGKKEEEKKGEHVQQSHAPAYT GSGEVHHAPPPAVSHQPQGE GANKHHEGLKEKLHRSSSSS SSSSSDEDESGGKKKKEGRKKKGLKDKLPGHGGEKESKEE GEKKGLMEKIKEKLPGHKEGEKPL

PRADHN4	MEQTDRLGFLGKKDEQKQHGGQGEYGNKQNEGMMEKIKDKL PGGHGGKATTVPGSQGATMTHEGPPKKGMMDKIKDKLPGG HGGKATTVPGSQGATMTHEGPPKKGMMEKIKEKLPGKK
PRADHN5	MADSNAPHEHKDRGLFGMFGKKDEEKKEGEYGQQQSHTADP YATHAPAYTGGVQYAPPSAASHQPQGYGQGEGGYKQDEPE RYSQGEKGKLNRRNNSSSSSSSDEDES GGKKKKEGRKKKGL KEKLHGRQDGQGIGKEEGEKKGLMDKIKEKLPGHKEGEREN REY
PRADHN6	MEQTDRLGFLGKKDEQKQHGGQGEYGNKQNEGMMDKIKDK LPGGHGGKATTVPGSQGATMTHEGPPKKGMMEKIKEKLPG KK
PRADHN7	MADEHHHKSHEYETTAVGDEGAVETKDRGLFDLFGKKEEEK PQDEVIVTEFEKVTVSEPEYKVVEEQKEEKKEEDEEEKKHGLL EQLRRSDSSSSSSSDEEEGEGEEKKKKKKKEKKGKKKVTDEK EEEKHEEDTSVPVYEEVAHPEPAQVVQPEEKKGFLKIKDK LPGQHKKTEEAVLSPAPPAPAEYTAAEASPQEVEAKEKKGIFE KIKEKLPGYHPKTEEEKEKEKEKDSASY
PRADHN8	MAELRDEYGNPIRQTDEYGNPVGTGAFGHGGHGAAAAGAG GYGAAGQQQAQPEKGLRRSGSSSSSSSEDDGMGGRKKK GLKDKIKDKLSGHK
PRADHN9	MAELRDEYGNPIRQTDEYGNPVGTGAFGHGGHGAAAAGAG GYGAAGQQQAQPEKGLRRSGSSSSSSSEDDGMGGRKKK GLKDKIKDKLSGHKPEEEQQQQQQHETAQHQPCKGLMAKI KEKLPGQSHQ

b. Alineamiento múltiple de proteínas

El alineamiento múltiple de las secuencias deducidas demostró las similitudes entre sus segmentos y la presencia de aminoácidos con propiedades hidrofílicas e hidrofóbicas en las mismas posiciones, observando ligeros cambios en los sectores más conservados de las secuencias, lo cual se presenta de mejor forma en los alineamientos de cada secuencia deducida junto a las cinco secuencias homólogas seleccionadas que resultaron del *BLASTp*.

PrADHN1 MAE EVPPSS - - - - - VPEQQDRALFGLFGKKKDET - - - - -
PrADHN2 MADSN - - - - - APEHTDRGFGMGFKKDEEK - - - - -
PrADHN3 MADSG - - - - - APEHTDRGLFGLFGKKEEK - - - - -
PrADHN4 - - - - - MEQTDRLGF - - - - - GKKDEQK - - - - -
PrADHN5 MADSN - - - - - APEHKDRGLFGMGFKKDEEKKEGEY - - - - -
PrADHN6 - - - - - MEQTDRLGF - - - - - GKKDEQK - - - - -
PrADHN7 MADEHHHKSHEYETTAVGDEGAVETKDRGLFDFLGKKKEEKPKQDEVIVTEFEKVTVSEPEYKVEE - - - - -
PrADHN8 MAEL - - - - - RDEYGNPIRQTDEY - - - - -
PrADHN9 MAEL - - - - - RDEYGNPIRQTDEY - - - - -

70 80 90 100 110 120 130

-HDQQSHAVDPHSGHAPAYTGEVHHAPPPAVAHQPQHYGQGTGGNAYAGEVHPIPPPAVSHQPPQHYYGQGDEG
-KGE--HVVQQS--HAPAYTG--SGEVHHAPPPAVSHQPP--GEGA
-QHG--GQGEYGNKQNEGMM--EKI
-GQQQSHTADPYATHAPAYTG--GVQYAPPSAASHQPPQGYGQGE
-QHG--GQGEYGNKQNE
QKEEKKEEDEEEKKHGLLEQLR--RSDSSSSSSSDEEEGEGEEKKKKKKEKKGKKKV
-GNPVGTTGAFGHGGHGAAG--AGGYGAAGQQQ--
-GNPVGTTGAFGHGGHGAAG--AGGYGAAGQQQ--

140 150 160 170 180 190 200

-----HGGHAALKHRSDSSSSSDEDEEDG-----KKKKKGLKEKKE-----KTPHHGSSAEHH
NK-----HPEGLKGKLHRSSSSSSSDEDESG-----GKKK-EGRKKKGKDKIKEKLPGHG-GEKEPK
NK-----HHEGLKEKLHRSSSSSSSDEDESG-----GKKKKEGRKKKGKDK-----KLPGHH-GEKESK
-----KDKLPGGHGGKATTVPGSSQGATMTHE-----GPPKKGMMDKIKD-----KLPGHGGKATTV
YKQDEPERYSQGEQGLNRRNSSSSSSSDEDESG-----GKKKKEGRKKKGKKE-----KPHGRDQGGIGK
-----GMMDDKIKD-----KLPGHGGKATTV
-----TDEKEEEKHEEDTSVPVVYE-EVAHPPEAQVVQPEEKKGFLEKIKD-----KLPQHKKTEEAV
-AQPEKG-LRRSGSSSSSSS-EDDGMG-----GRRKKKGLKDKIKD-----KLSGHK
-AQPEKG-LRRSGSSSSSSS-EDDGMG-----GRRKKKGLKDKIKD-----KLSGHKPEEEQQQ

210 220 230 240 250 260 270 280

P-----PQPEE-----KKGLMEKIKELPGSGGH-----
EET-----EKKGLVDKIKELKLPGHKEGEKGH-----
EEG-----EKKGLMEKIKELKLPGHKEGEKPL-----
P-----GSQGATMTHEGPPKKGMMEKIKELPGKK-----
EEG-----EKKGLMDKIKELKLPGHKEGERENREY-----
P-----GSQGATMTHEGPPKKGMMEKIKELPGKK-----
LSPPAPPAEYTAEEASPQVEAKEKKGI FEKIKELKLPGYHPKTEEEKEKEKDSASY
QQQH-----ETAQHQPKEKKGLMAKIKELPGQSHQ-----

19

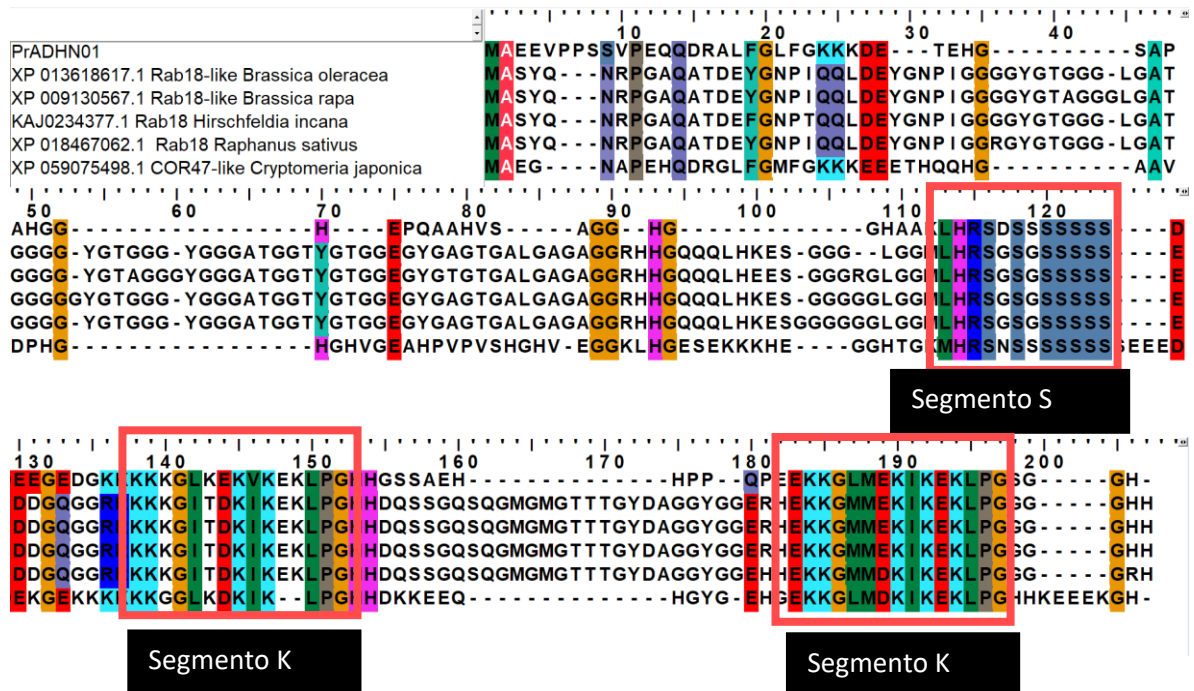


Figura 3.2. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN1. Enmarcadas rojas se detallan el segmento S y el K.

PrADNH1 se identificaron claramente tres segmentos conservados donde el segmento S posee pequeñas variaciones cambiando en la posición 117 una Glicina (G) por ácido aspártico (D) en esta proteína, mientras que los segmentos K se observa una leve diferencia cambiando en la posición 142 una Isoleucina (I) por Leucina (L), así mismo un cambio en la posición 144 de ácido aspártico (D) por ácido glutámico (E) y en la posición 146 una Isoleucina (I) por Valina (V) en el caso del segundo segmento K solo existe un cambio de aminoácidos en la posición 186 de Metionina (M) a Leucina (L), sin embargo la dehidrina de *C.japónica* también posee este cambio.



Figura 3.3. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN2. Enmarcadas rojas se detallan el segmento S y K.

PrADHN2 presentó alineamiento de cuatro dominios conservados el primero es el segmento F que en esta proteína tiene un cambio en la posición 16 de un aminoácidos de Glutamina (Q) a una Treonina (T), además de poseer tres aminoácidos Fenilalanina (F) dentro del segmento, también existe un cambio en la posición 23 de un aminoácido Leucina (L) a Metionina (M), mientras que el

segmento S presenta la misma composición salvo un cambio en la posición 117 de Treonina (T) a Serina (S), cambio que también está presente en la COR47-like de *C.japónica*, el primer segmento K no está tan bien definido pero si presenta la secuencia consenso KIKEKLPG en esta proteína putativa, en el segundo segmento K no existe mayor variación, cambiando en la posición 214 de Lisina (K) a ácido glutámico (E).

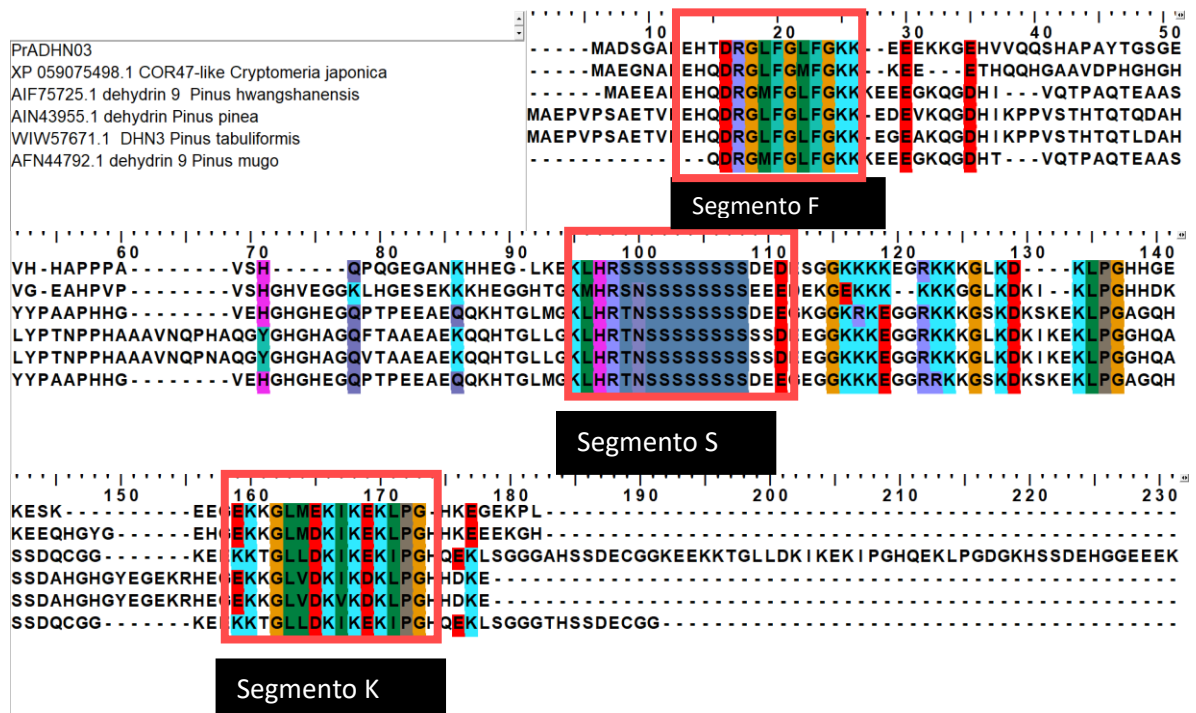


Figura 3.4. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN3. Enmarcadas rojas se detallan el segmento F, S y K.

PrADHN3 se alinea en tres sitios, el segmento F presenta un cambio en la posición 15 cambiando una Glutamina (G) por Treonina (T), mientras que en el

segmento S no hay mayores diferencias con sus proteínas homologas, salvo en la posición 100, donde todos los homólogos tienen Asparagina (N) y esta proteína putativa tiene Serina (S), el segmento K presenta dos cambios en la posición 164 y 165, respectivamente cambiando de Leucina (L) a Metionina (M) y de ácido aspártico (D) a ácido glutámico (E).

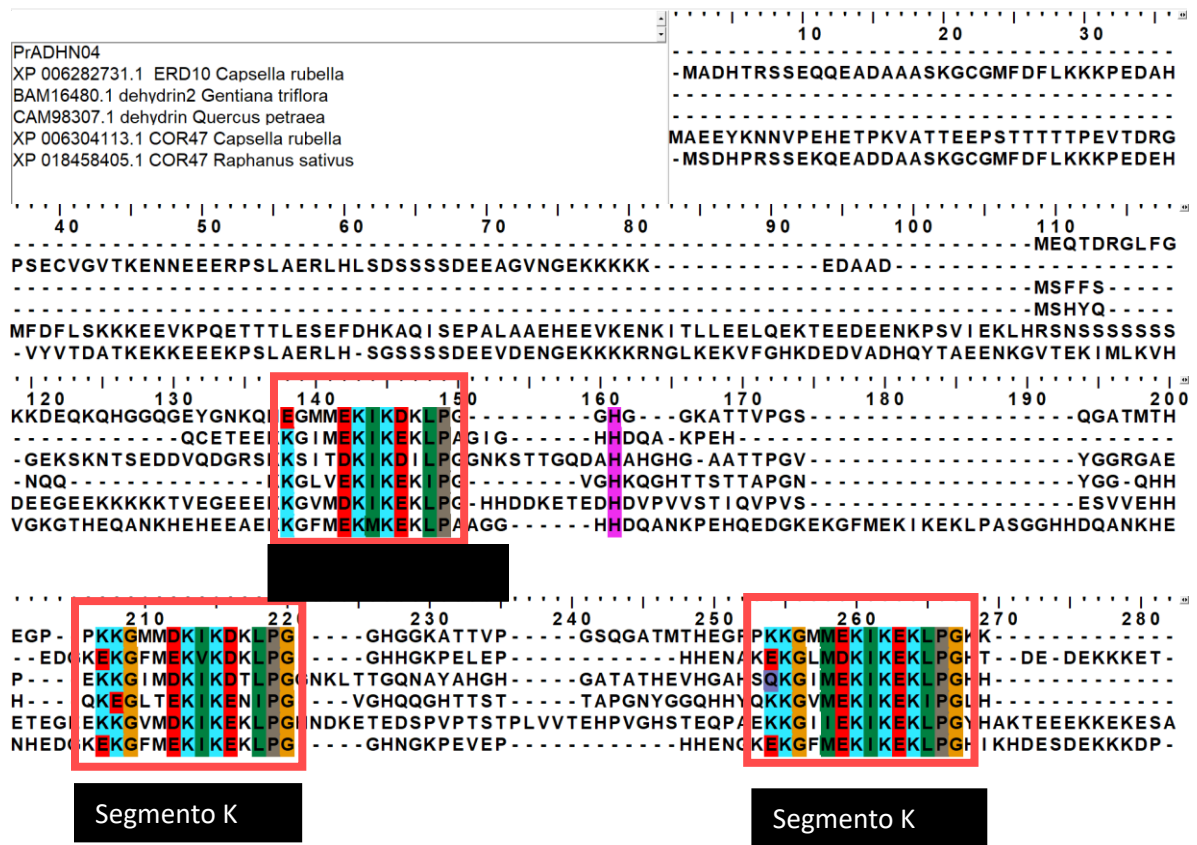


Figura 3.5. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN4. Enmarcadas rojas se detallan el segmento F, S y K.

PrADHN4 presentó tres segmentos K, el primero tiene muchos cambios en sus aminoácidos, en primer lugar cambia en la posición 138 una Lisina (K) por ácido

glutámico (E) luego una zona entre los aminoácidos 139 y 141 donde no hay identidad entre los homólogos, el segundo segmento K tiene como mayor diferencia el aminoácido en la posición 206, sin embargo ninguna de las otras proteínas homologas presenta un patrón similar, esta proteína presenta el aminoácido Prolina (P) mientras que las otras tiene Lisina (K), ácido glutámico (E) y Glutamina (Q), luego hay una zona donde no presenta identidad entre las posiciones 210 y 211, como también en la posición 217, con el tercer segmento K pasa algo similar teniendo un cambio en la posición 253 aminoácido Prolina (P) y luego otra zona en la posición 257 donde no hay identidad, donde esta proteína presenta el aminoácido Prolina (P) y en las otras los aminoácidos Lisina (K), Serina (S), Glutamina (Q) y ácido glutámico (E).

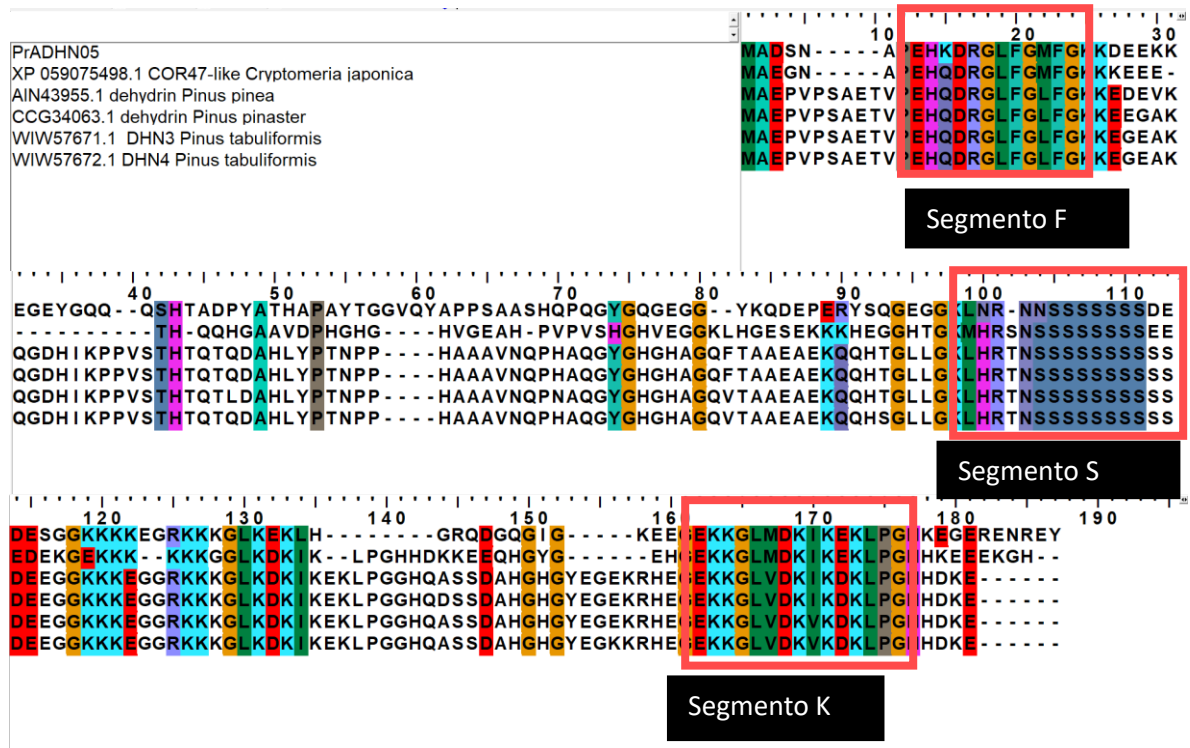


Figura 3.6. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN5, Enmarcadas en color rojo se detalla el segmento F, S y K.

PrADHN5 presentó tres segmentos conservados, el primer segmento es el F el cual mostró una modificación de dos aminoácidos cambiando en las posiciones 15 y 22 respectivamente de Glutamina (Q) a Lisina (K) y el segundo de un Leucina (L) a Metionina (M), cambio que también presentaba la COR47-like de *C.japonica*, en el segmento S existe un cambio en la posición 100 de Histidina (H) a Asparagina (N) y otro el cual no alineó con las proteínas en la posición 102, el segmento K existe un cambio en la posición 172 de ácido aspártico (D) a ácido glutámico (E).

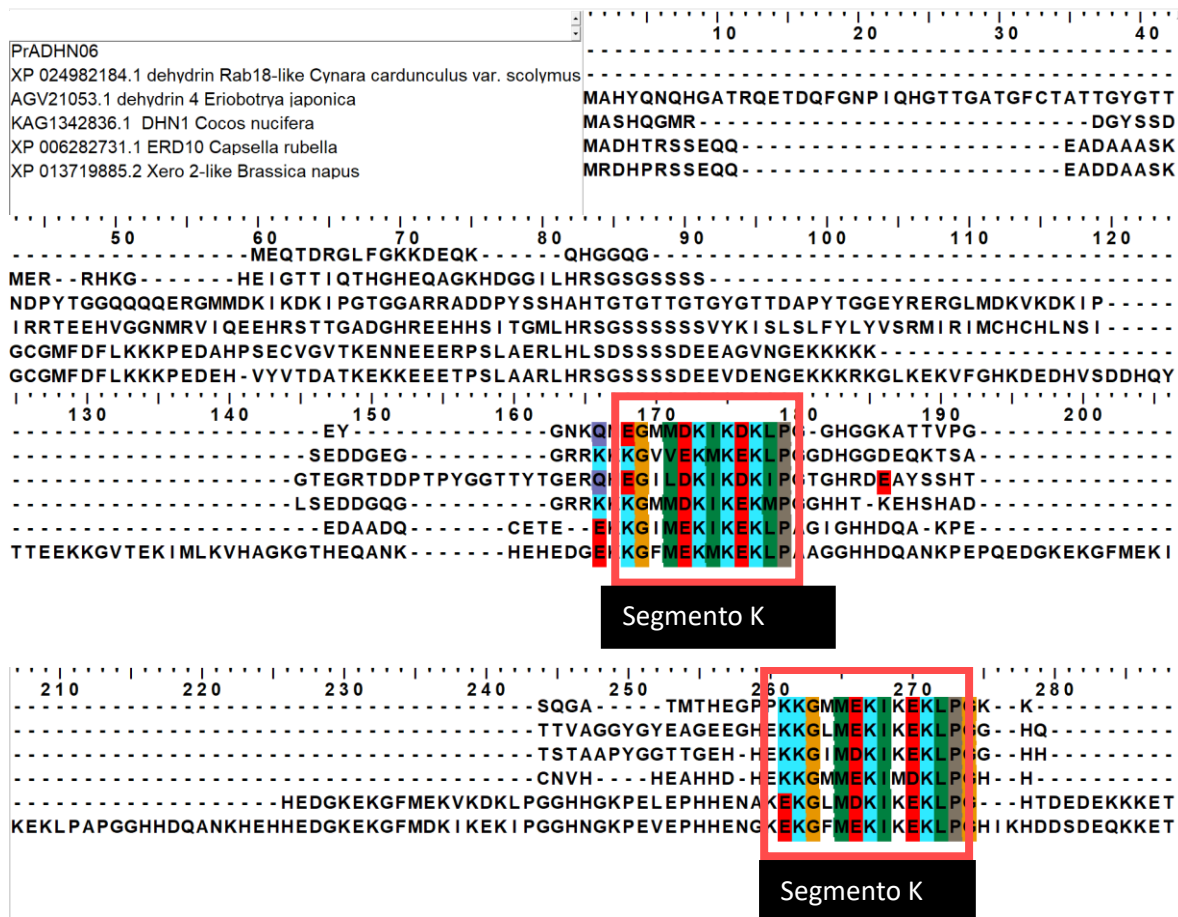


Figura 3.7. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN6. Enmarcadas en color rojo se detallan los segmentos K1 y K2.

PrADHN6 presentó dos segmentos K, el primer segmento K mostró cambios en la posición 168 de Lisina (K) a ácido glutámico (E), pero este cambio también lo presenta la dehidrina 4 de *E.japonica*, luego el otro cambio es la presencia de una Metionina (M) ya que las otras proteínas tienen Isoleucina (I), Valina (V) y Fenilalanina (F), el segundo segmento K no presenta identidad en dos posiciones 264 y 269, donde esta proteína tiene una Metionina (M), mientras que las otras presentan Isoleucina (I), Leucina (L) y Fenilalanina (F), y en la segunda la

mayoría presenta el aminoácido Lisina (K) incluyendo estas proteínas excepto la dh1 de *C.nucifera* que presenta el aminoácido Metionina (M).

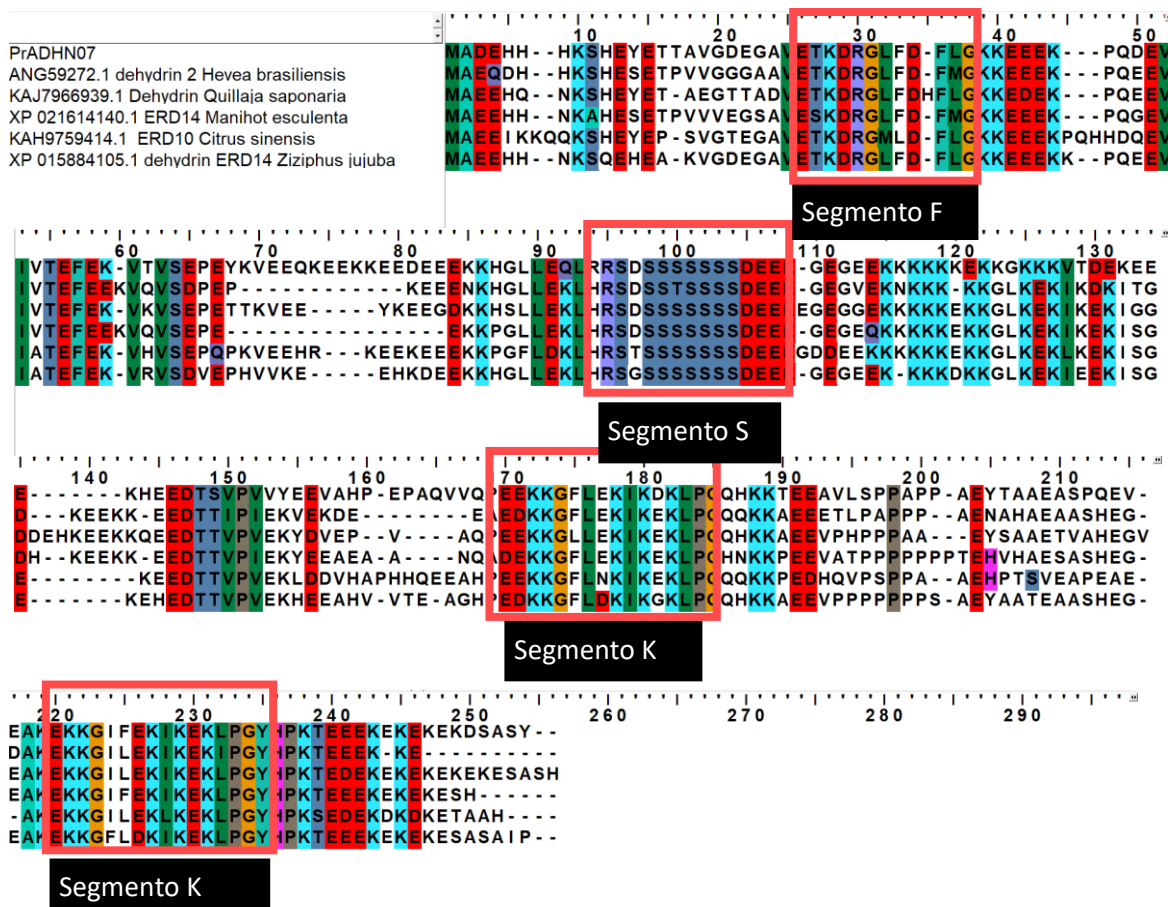


Figura 3.8. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN7. Enmarcadas en color rojo se detalla el segmento F, S, K1 y K2.

PrADHN7 presentó cuatro segmentos conservados, el primer segmento es el F, este segmento mostró pocas diferencias con las proteínas homólogas, existe una zona en la posición 35 donde no se alineó la proteína con sus homólogos, el segmento S no presentó mayor diferencia de aminoácidos, solo en una zona donde no hubo identidad en la posición 97 y esta proteína presentó un aminoácido ácido aspártico (D), mientras que las homólogas también presentaron este aminoácido pero además Treonina (T) y Glicina (G), en el primer

segmento K no hubieron mayores diferencias a excepción del remplazo de un aminoácido en la posición 181 de ácido glutámico (E) a ácido aspártico (D), mientras que en el segundo segmento K existe un cambio de aminoácidos en la posición 225 de Leucina (L) a Fenilalanina (F), pero este cambio también lo presenta la proteína ERD10 de *C.sinensis*.

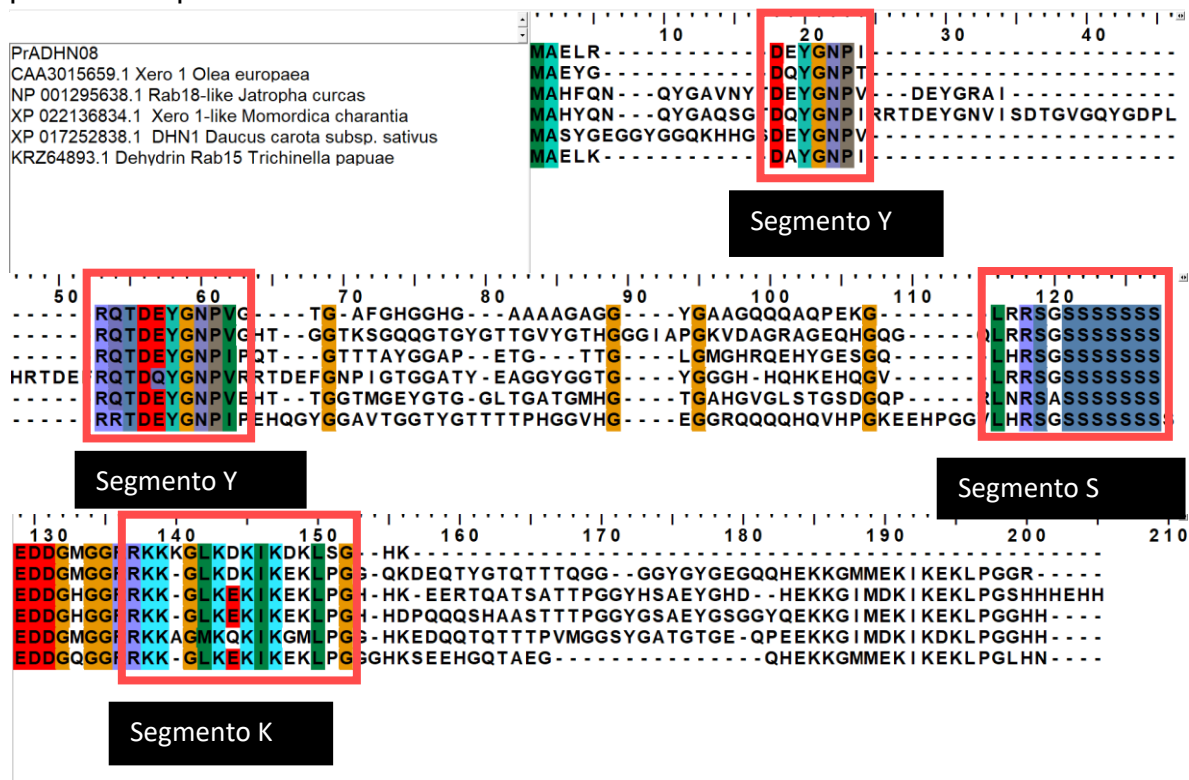


Figura 3.9. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN8. Enmarcadas en rojo se detallan el segmento Y1, Y2, S y K.

PrADHN8 presenta cuatro segmentos conservados, con pocas modificaciones en el segmento Y respecto a sus homólogos, teniendo diferencias en la posición 19 donde esta proteína mostró ácido glutámico (E) al igual que la Rab18-like de *J. curcas* y con la DHN1 de *D. carota*; mientras que en la posición 24 presenta Isoleucina (I) al igual que la Xero 1-like de *M. charantia* y la Rab15 de *T. papuae*. El segundo segmento Y no presenta mayores diferencias con sus homólogos, es casi completamente igual a excepción de algunos aminoácidos de sus

homólogos. El segmento S presenta dos sitios sin identidad en las posiciones 117 y 120; en la primera PrADHN8 presenta Arginina (R) mientras sus homólogos tienen Histidina (H) y Asparagina (N), diferenciándose de proteínas homologas como Rab18-like de *J. curcas* y DHN1 de *D. carota* respectivamente. Otro cambio es de Glicina (G) a Alanina (A) de la DHN1 de *D. carota*, finalmente el segmento K presenta dos cambios importantes en las posiciones 144 y 151, remplazando el aminoácido ácido glutámico por ácido aspártico en dos oportunidades; además, se cambia Prolina (P) por Serina (S).

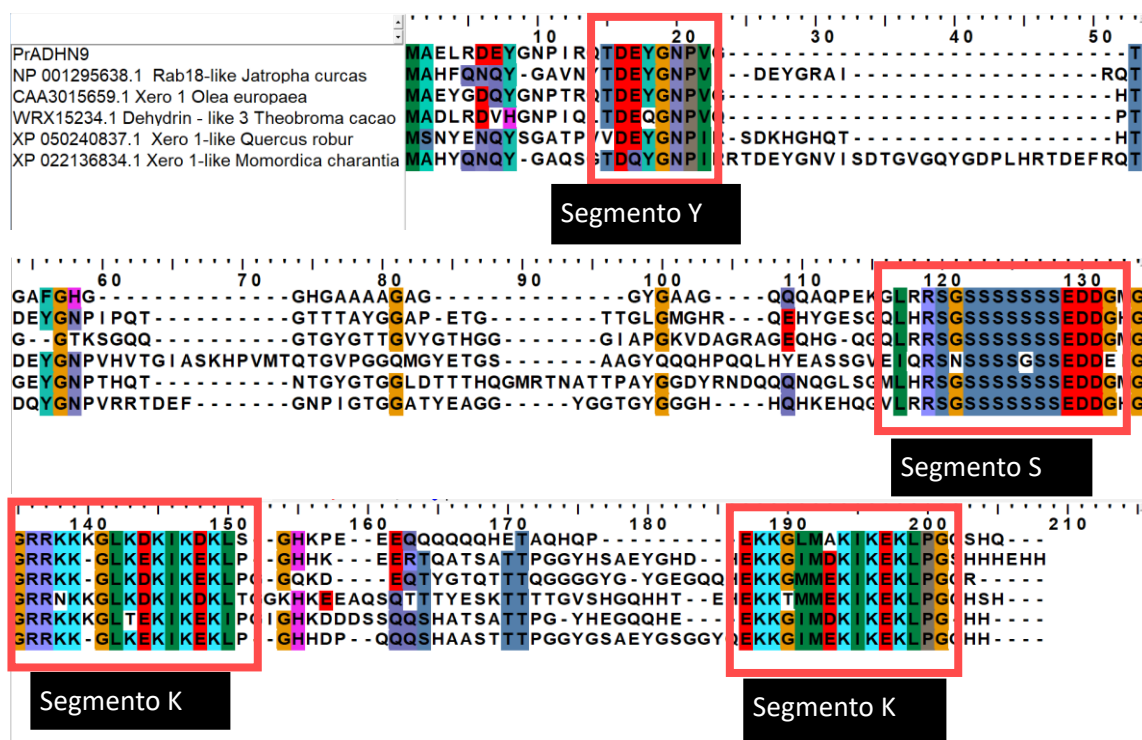


Figura 3.10. Alineamiento de proteínas ortólogas de la secuencia deducida de PrADHN9. Enmarcadas en rojo se detallan el segmento Y, S y K.

PrADHN9 presenta cuatro segmentos, el primer segmento es el Y el cual está bien conservado ya que no existe diferencias con sus homólogos, en cuanto al segmento S existe un aminoácido que no presenta identidad y está en la posición

118 como es el caso del aminoácido Arginina (R) que está en Xero1 de *O.europea* y Xero1-like de *M.charantia*, el primer segmento K es prácticamente igual a sus homólogos excepto en el aminoácido en la posición 140 donde tres homólogos no alinearon pero dos proteínas presentan el mismo aminoácido, en cuanto al segundo segmento K este presenta una leve diferencia con sus homólogos debido a que estos últimos presentan un aminoácido ácido aspártico (D) o ácido glutámico (E) en su posición 193 pero esta proteína presenta un aminoácido Alanina (A).

c. Identificación de firmas dehidrina y sitios de interacción

En Prosite se identificaron dos firmas dehidrina, cuyos códigos se identificaron como: PS00315 y PS00823, presentando variaciones debido a diferentes modificaciones. La firma consenso para PS00315 es: S(4)-[SD]-[DE]-x-[DE]-[GVE]-x(1,7)-[GE]-x(0,2)-KR, mientras que la firma consenso para PS00823 es: [KR]-[LIM]-K-[DE]-K-[LIM]-P-G. Al eliminar la opción de excluir los motivos con alta probabilidad de ocurrencia, obtuvimos todos los posibles sitios de interacción de estas proteínas. Entre estos sitios se encuentran los sitios de: amidación, fosforilación de la caseína quinasa II, fosforilación de la proteína quinasa C, glicosilación, miristoilación y fosforilación de tirosina quinasa. Cada sitio y firma se describen detalladamente a continuación:

- PrADHN1 presentó una firma dehidrina 1 (PS00315) y una firma dehidrina 2 (PS00823), en las posiciones 66-82 y 116-123, respectivamente. Identificando cinco sitios de fosforilación de la caseína quinasa II (PS00006) en las posiciones 9-12, 68-71, 69-72, 70-73 y 98-101, así como dos sitios de amidación (PS00009) en las posiciones 22-25 y 76-79.
- PrADHN2 posee una región rica en Lisina entre los residuos 127-179, identificando una firma dehidrina 1 (PS00315) y dos dehidrina 2 (PS00823) en las posiciones 115-135 y 140-147, 166-173, respectivamente. Presenta dos sitios de fosforilación de la proteína quinasa C (PS00005) en las posiciones 10-12 y 158-160, y tres sitios de fosforilación de la caseína quinasa II (PS00006) en las posiciones 117-120, 118-121 y 119-122. Se observaron tres sitios de amidación (PS00009) en los rangos 18-21, 125-128 y 130-133, junto con un sitio de glicosilación (PS00001) en la posición 111-114.
- PrADHN3 presentó una firma dehidrina 1 (PS00315) y una firma dehidrina 2 (PS00823) en las posiciones 81-102 y 129-136, respectivamente. Posee tres sitios de amidación (PS00009) en las posiciones 18-21, 91-94 y 97-100, y cuatro sitios de fosforilación de la caseína quinasa II (PS00006) en las posiciones 83-86, 84-87, 85-88 y 117-120, incluyendo un sitio de fosforilación de la proteína quinasa C (PS00005) en la posición 10-12.
- PrADHN4 presentó tres firmas dehidrina 2 (PS00823) entre las posiciones 35-42, 72-79 y 109-116, respectivamente. Posee un sitio de fosforilación

de la proteína quinasa C (PS00005) entre los residuos 4-6 y dos sitios de amidación (PS00009) entre las posiciones 9-12 y 115-118.

- PrADHN5 presentó una firma dehidrina 1 (PS00315) y una dehidrina 2 (PS00823) en las posiciones 97-118 y 146-153, respectivamente. Posee 3 sitios de amidación (PS00009) entre las posiciones 18-21, 107-110 y 113-116, así como 4 sitios de miristoilación (PS00008) entre las posiciones 31-36, 68-73, 87-92 y 130-135. Se identificó un sitio de fosforilación de la tirosina quinasa (PS00007) entre las posiciones 75-82, y 4 sitios de fosforilación de la caseína quinasa II (PS00006) en las posiciones 83-86, 99-102, 100-103 y 101-104. Además, se identificaron 2 sitios de glicosilación (PS00001) en las posiciones 93-96 y 94-97.
- PrADHN6 reveló dos firmas dehidrina 2 (PS00823) entre las posiciones 35-45 y 72-79, junto con un sitio de fosforilación de la proteína quinasa C (PS00005) entre las posiciones 4-6. Presentó dos sitios de amidación (PS00009) en las posiciones 9-12 y 78-81, e identificando cinco sitios de miristoilación (PS00008): 25-30, 42-47, 45-50, 53-58 y 56-61.
- PrADHN7 presentó una región rica en lisina y ácido glutámico, así como una firma dehidrina número 1 (PS00315) y dos firmas número 2 (PS00823) entre las posiciones 93-112 y 162-169, 209-216, respectivamente. Presentó un sitio de miristoilación (PS00008) entre las posiciones 21-26, y 9 sitios de fosforilación de la caseína quinasa II (PS00006) entre las posiciones 49-52, 55-58, 57-60, 95-98, 96-99, 97-100, 189-192, 194-197 y

221-224. Adicionalmente, se incluyeron dos sitios de amidación (PS00009) entre las posiciones 34-37 y 115-118.

- PrADHN8 reveló una firma dehidrina número 1 (PS00315) entre las posiciones 63-79. Presentó sitios de miristoilación (PS00008) entre las posiciones 19-24, 30-35, 33-38, 40-45, 46-51, 55-60 y entre la posición 60-65, así como 3 sitios de fosforilación de la caseína quinasa II (PS00006) entre las posiciones 65-68, 66-69 y 67-70. Adicionalmente, se incluyó un sitio de amidación (PS00009) entre las posiciones 73-76.
- PrADHN9 reveló dos firmas dehidrina, la número 1 (PS00315) y la número 2 (PS00823), en las posiciones 63-79 y 119-126, además de encontrar sitios de miristoilación (PS00008) en las posiciones 19-24, 30-35, 33-38, 40-45, 46-51, 55-60 y 60-65. Presentó sitios de fosforilación de la caseína quinasa II (PS00006) en las posiciones 65-68, 66-69 y 67-70, así como también sitio de amidación (PS00009) desde la posición 73 a la 76.

Al analizar los segmentos conservados de las dehidrina deducidas se encontró la presencia de seis segmentos conservados, tres de ellos pertenecientes al segmento K (logo 1, 3 y 6 en Figura 3.11). Para los segmentos restantes solo existe una figura consenso, en donde, el segmento S se representa con el logo 2, el segmento Y se representa con el logo 4, y el segmento F se representa con el logo 5.

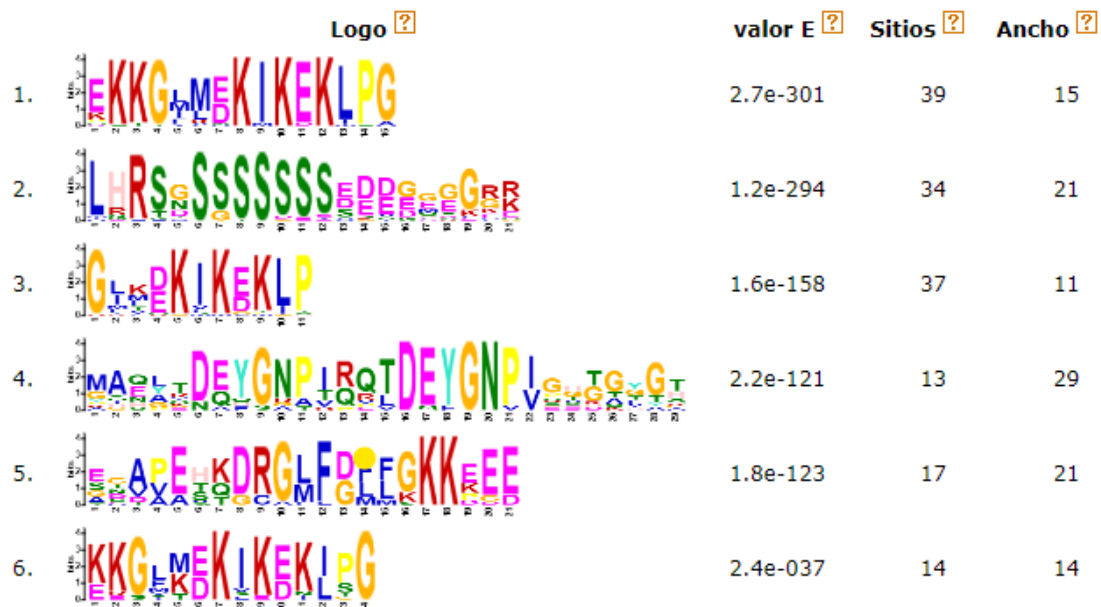


Figura 3.11. Logo de los motivos conservados encontrados en las proteínas deducidas junto a sus homólogos, además aparece el valor de significancia, numero de sitios donde se encontró el motivo y el ancho de este.

d. Estructura secundaria de las proteínas

En cuanto a las estructuras conformacionales que poseen las secuencias deducidas, los resultados demuestran que todos los segmentos K poseen una alfa hélice en su formación. A continuación, se describe cada secuencia y su estructura resultante:

- PrADHN1 (figura 3.12) presentó una estructura de hebra beta, simbolizado con una flecha en los residuos posicionados en los residuos 57 al 60; y

dos alfa hélices entre las posiciones 84-91 y 113-120, con una confiabilidad de puntaje 8 para ambas alfas hélices.

- PrADHN2 (figura 3.13) presentó una sola estructura alfa hélice entre la posición 163-170, con una confiabilidad de puntaje 7.
- PrADHN3 (figura 3.14) presentó una estructura alfa hélice en su extremo C-terminal entre la posición 126 y 133, con una confiabilidad de puntaje 7.
- PrADHN4 (figura 3.15) presentó dos estructuras alfa hélice entre las posiciones 32-39 y 106-113, con una confiabilidad neutra, la primera de estas presenta un puntaje de 7, mientras que la segunda un puntaje de 6.
- PrADHN5 (figura 3.16) presentó una estructura alfa hélice en su extremo terminal entre las posiciones 143-150, con una confiabilidad de puntaje 7.
- PrADHN6 (figura 3.17) presentó dos estructuras alfa hélice entre las posiciones 32-39 y 69-76, ambas con una confiabilidad de puntaje 7.
- PrADHN7 (figura 3.18) presentó tres estructuras de hebra beta entre las posiciones 11-13, 54-55, 137-141, con confiabilidad neutra (representado por el color verde), con cinco estructuras alfa hélice entre las posiciones 63-70, 73-78, 160-165, 206-212 y 223-233, con un puntaje 6.
- PrADHN08 (figura 3.19) presentó dos estructuras alfa hélice entre las posiciones 5-6 y 81-88, pero una solo contiene tres residuos y posee una confiabilidad neutra, mientras que la segunda estructura alfa hélice presenta más residuos en su composición y con una confiabilidad de puntaje 7.

- PrADHN9 (figura 3.20) presentó dos estructuras alfa hélice entre las posiciones 81-88 y 115-123, ambas con una confiabilidad de puntaje 6.

Se llevaron las secuencias de proteínas deducidas a la aplicación *Bioedit* la cual permitió mostrarnos la secuencia enumerando sus aminoácidos para así lograr representar de mejor manera los sitios de fosforilación, amidación, glicosilación, miristoilación y motivos conservados resultantes del análisis en *Prosit* y el software *MEME*, donde se encuentran representados desde la figura 3.12 a la 3.20.



Figura 3.12. Representación de PrADHN1. En color rosado están enmarcados los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados verdes, y los sitios de fosforilación de la caseína quinasa II están enmarcados negros, además de la estructura secundaria bajo los aminoácidos que la componen.



Figura 3.13. Representación de PrADHN2. Coloreado en rosado están enmarcados los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, mientras que los sitios de fosforilación de la caseína quinasa II están enmarcados en color negro, los sitios de fosforilación de la quinasa C están enmarcados con la línea negra discontinua y el sitio de glicosilación se representa enmarcado en color naranja, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.



Figura 3.14. Representación de PrADHN3, en rosado están enmarcados los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están en verde, mientras que los sitios de fosforilación de la caseína quinasa II están en negro, y la estructura secundaria está bajo los aminoácidos que la componen.



Figura 3.15. Representación de PrADHN4. Coloreado en rosado están enmarcados los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, los sitios de fosforilación de la quinasa C están enmarcados con la línea negra discontinua, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.

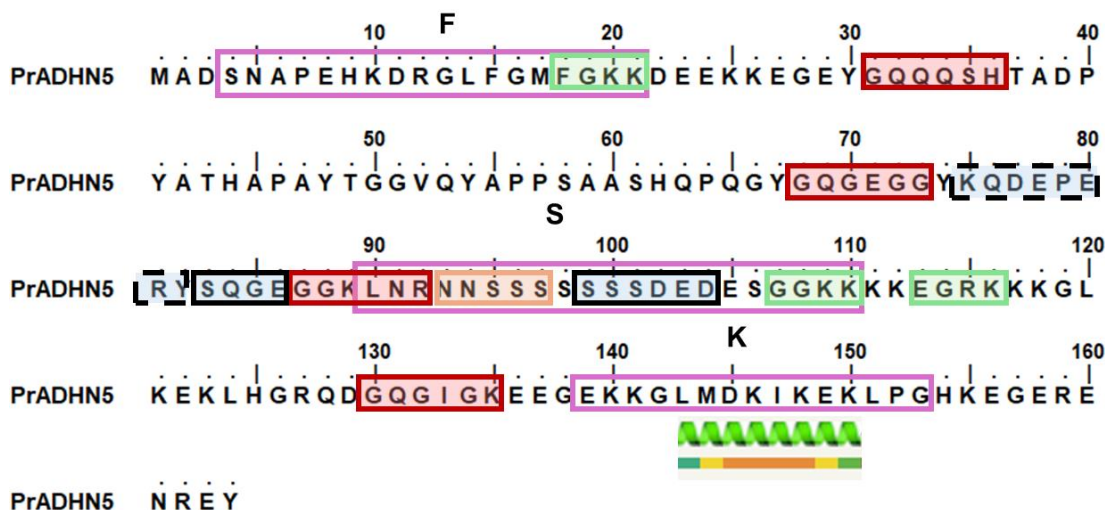


Figura 3.16. Representación de PrADHN5. Coloreado en rosado están enmarcados los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, mientras que los sitios de fosforilación de la caseína quinasa II están enmarcados en color negro, los sitios de fosforilación de la tirosina quinasa están enmarcados con la línea negra discontinua, los sitios de miristoilación están enmarcados en color rojo y finalmente el sitio de glicosilación enmarcado en color naranja, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.

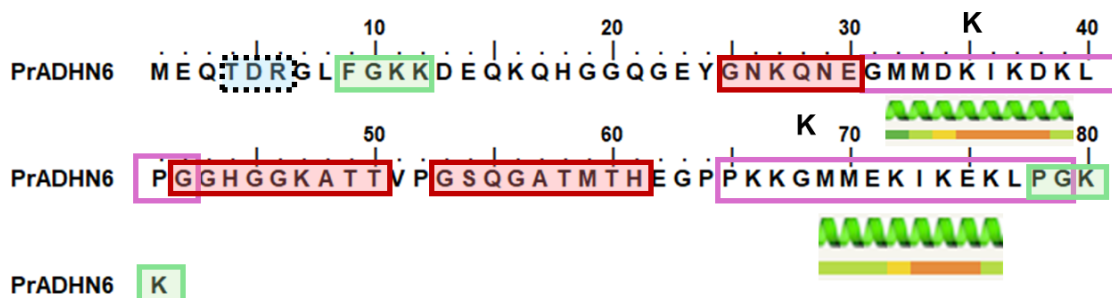


Figura 3.17. Representación de PrADHN6. Coloreado en rosado están enmarcados los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, mientras que el sitio de fosforilación de la quinasa C está enmarcado con la línea negra discontinua y los sitios de miristoilación están enmarcados en color rojo, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.

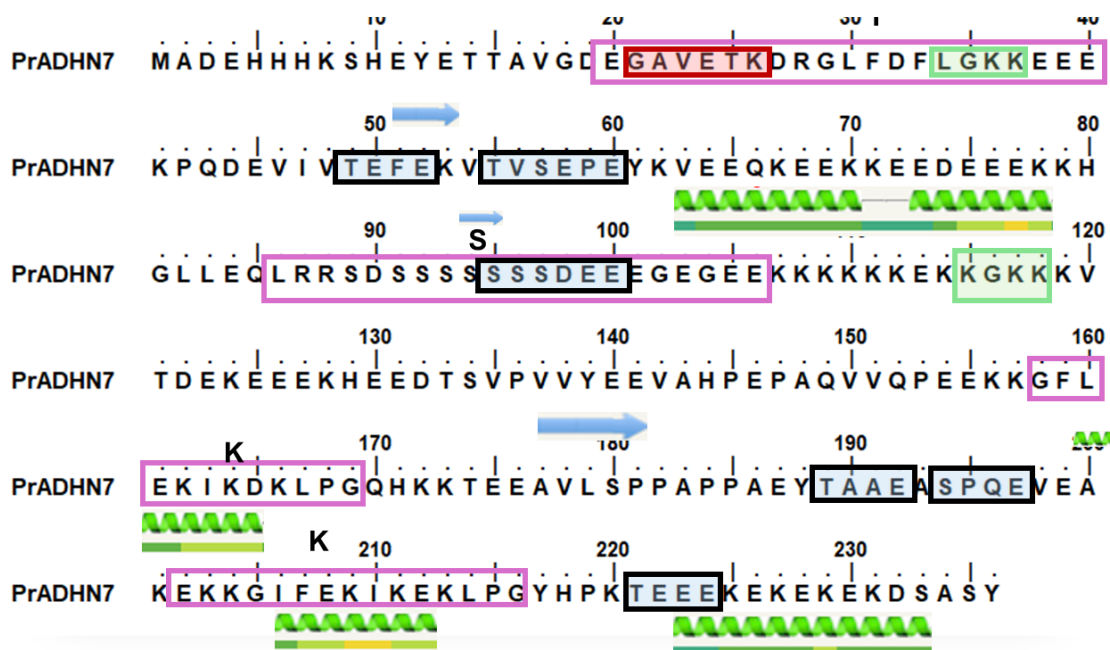


Figura 3.18. Representación de PrADHN7. Coloreado en rosado están enmarcados los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, mientras que los sitios de fosforilación de la caseína quinasa II y los sitios de miristoilación están enmarcados en color rojo, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.

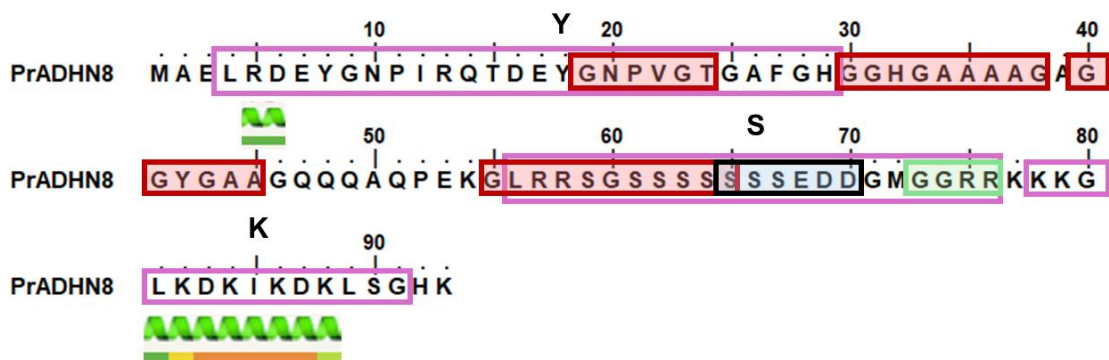


Figura 3.19. Representación de PrADHN8. Coloreado en rosado están enmarcados los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, mientras que los sitios de fosforilación de la caseína quinasa II y los sitios de miristoilación están enmarcados en color rojo, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.



Figura 3.20. Representación de PrADHN9. Coloreado en rosado están enmarcados los segmentos conservados con su respectiva letra en la parte superior, los sitios de amidación están enmarcados en color verde, mientras que los sitios de fosforilación de la caseína quinasa II y los sitios de miristoilación están enmarcados en color rojo, además la estructura secundaria se encuentra bajo los aminoácidos que la componen.

e. Caracterización bioquímica de las proteínas

Este análisis reveló una gran cantidad de aminoácidos hidrofílicos, como residuos de glicina y lisina. Todas las puntuaciones GRAVY fueron negativas, confirmando una alta hidrofilia, ya identificado en otras dehidrina vegetales (Sun *et al.* 2021). Estos resultados se resumen en la tabla 3.2:

Tabla 3.2. Resumen de las características bioquímicas de las secuencias.

Dehidrinas	TIPO	N°AA	PI	PW	GRAVY	%G Y K
PrADHN1	FSK ₂	127	6.87	13411.74kDa	-1.305	14,2% 14,2%
PrADHN2	FSK ₂	181	7,00	19520,32 kDa	-1,495	14,4% 13,8%
PrADHN3	FSK ₂	144	8,62	15606,26 kDa	-1.516	13,9% 17,4%
PrADHN4	K ₃	118	9,67	12507,35 kDa	-1.340	20,3% 17,8%
PrADHN5	FSK	164	7.03	17999.48 kDa	-1.734	16.5% 14.0%
PrADHN6	K ₂	81	9,52	8765,01 kDa	-1.441	19,8% 18,5%
PrADHN7	FSK ₂	236	4,96	27058,73 kDa	-1,596	5,1% 18,2%
PrADHN8	YSK	93	9,52	9531,37 kDa	-1.151	21,5% 9,7%
PrADHN9	YSK ₂	130	9,13	13846,13 kDa	-1,397	16,9% 10,8%

N°AA: número de residuos o aminoácidos en la secuencia; **PI:** punto isoeléctrico teórico; **PW:** peso molecular; **GRAVY:** índice de hidropaticidad; **% G Y K:** porcentaje de aminoácidos, como glicina y lisina.

Las secuencias obtenidas presentaron un alto grado de desorden en su estructura, detallados en la tabla 3.3.

Tabla 3.3. Porcentaje del índice de desordenamiento de la estructura secundaria de las secuencias deducidas.

Secuencia dehidrina	IDP
PrADHN1	94%
PrADHN2	95%
PrADHN3	93%
PrADHN4	92%
PrADHN5	95%
PrADHN6	89%
PrADHN7	99%
PrADHN8	66%
PrADHN9	85%

En cuanto a la configuración de cada dehidrina deducida, estas se representan en la tabla 3.4:

Tabla 3.4. Resumen del tipo de configuración de cada proteína.

Proteínas	Tipo
PrADHN1	FSK2
PrADHN2	FSK2
PrADHN3	FSK2
PrADHN4	K3

PrADHN5	FSK
PrADHN6	K2
PrADHN7	FSK2
PrADHN8	YSK
PrADHN9	YSK2

f. Análisis filogenético

El árbol filogenético presentó dos grandes grupos o clados, con valores de BOOSTRAPs bajos, los cuales oscilan entre 56 y 48%, respectivamente. Se encontró una rama que no pertenece a ningún clado, disminuyendo así el porcentaje de Bootstrap en la raíz del árbol filogenético.

El porcentaje es el grado de confianza de la agrupación de las proteínas dentro del mismo clado, en la figura 3.21 el clado que se encuentra en la parte inferior contiene a PrADHN7, agrupándola junto a *Quillaja saponaria* con un porcentaje de confianza de 96%, siendo *Citrus sinensis* y *Ziziphus jujuba* los siguientes ancestros más cercanos.

El siguiente clado posee su primer nodo con un porcentaje de confianza de 56% con un mayor número de especies, en el cual se observa un 100% de agrupación entre PrADHN8 y PrADHN9, como también entre PrADHN4 y PrADHN6, para este último el ancestro en común más cercano es *Cynara cardunculus*, mientras que para el primer grupo es *Theobroma cacao*.

Para las proteínas deducidas PrADHN5, PrADHN3 y PrADHN2 *Cryptomeria japonica* es el ancestro en común más cercano, PrADHN5 y PrADHN3 pertenecen al mismo nodo con un 89% de confianza, mientras que PrADHN2 está en un nodo anterior a estas últimas proteínas. Finalmente, PrADHN1 esta agrupada dentro del género *Pinus*, con especies como *P. tabuliformis*, *P. hwangshanensis* y *P. mugo*.

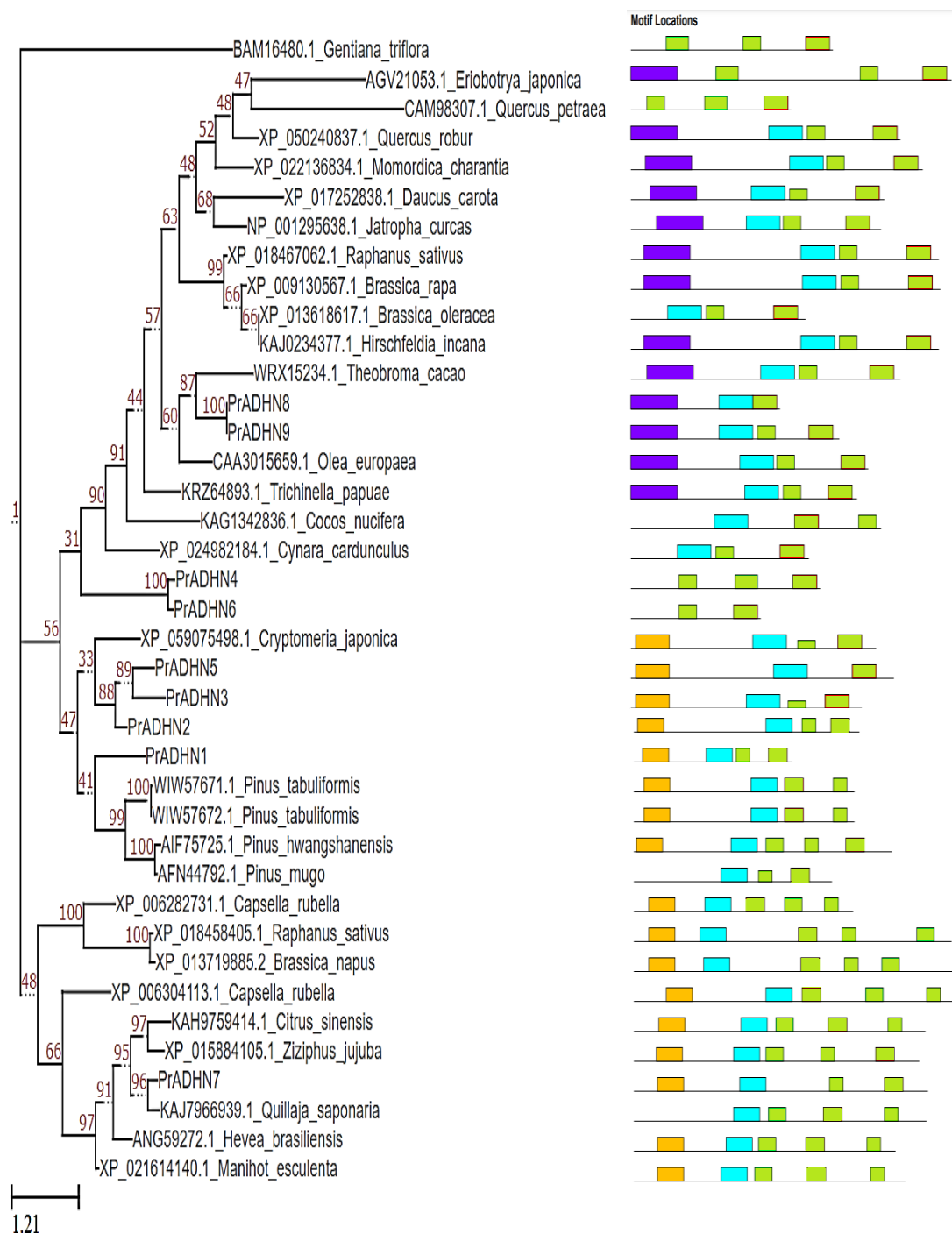


Figura 3.21. Árbol filogenético de proteínas dehidrina homólogas a las proteínas deducidas en *Prumnopytis andina*. Al lado derecho de la imagen están los segmentos conservados, donde el color morado representa el segmento Y, el amarillo el segmento F, el celeste el segmento S y el verde el segmento K. Los segmentos se alinean con su respectivo código y especie de la proteína.

IV. DISCUSIÓN

Gracias a los resultados obtenidos en esta investigación, identificamos la presencia de nueve genes que codifican proteínas dehidrina en la biblioteca de secuenciación de *P. andina*. La caracterización de estas dehidrina permitió obtener diferentes configuraciones y funciones, describiendo tanto los sitios conservados como los de interacción de las proteínas. Así mismo, el árbol filogenético demostró la cercanía evolutiva de las dehidrininas a las dehidrina putativas. Por lo tanto, se acepta la hipótesis planteada, ya que se encontraron diferentes configuraciones de dehidrina en la biblioteca de expresión de *P. andina*, donde cuatro proteínas dehidrina presentaron la configuración FSK2, siendo esta configuración de proteínas la que expresa una respuesta al estrés abiótico por frío, ya que sus segmentos conservados presentan características de protección enzimática y protección a la membrana. Sin embargo, se encontraron configuraciones del tipo YnSKn las cuales son típicas de angiospermas, adicionalmente, el segmento Y está involucrado en la defensa contra la deshidratación en plantas que poseen sus semillas recubiertas.

En la base de datos NCBI, existen solo 726 dehidrina descritas para coníferas, pertenecientes únicamente a *P. pinaster* y *P. halapensis*. Por otro lado, en UniProt se han descrito 611 dehidrina en coníferas. Sin embargo, actualmente no

existen dehidrina descritas para *P. andina*. Por lo que esta investigación da el primer paso para estudiar estas proteínas, cruciales para los procesos de adaptación al estrés en coníferas, un desafío significativo debido al cambio climático.

Las dehidrina se subclasifican en cinco arquitecturas según la presencia y el orden de los tres principales motivos conservados (Close 1996), En esta investigación obtuvimos las arquitecturas Kn, YSKn y FSKn, donde cada segmento presenta ciertas funciones y características específicas. Sin embargo, (Perdiguero *et al.* 2015) postula que existe un sitio conservado en los primeros 23 aminoácidos de la región N-terminal con la secuencia consenso MAEEAPEHQDRGMFGLFGKKKED en *Pinus pinaster* el cual coincide con la secuencia consenso del segmento F, curiosamente este fragmento conservado se encuentra en gimnospermas como también en angiospermas del tipo SKn (Perdiguero *et al.* 2014)

(Hughes y Graether 2011, Atkinson *et al.* 2016)) mencionan que el segmento K parece una hélice alfa anfipática, aunque se sugiere que es débilmente helicoidal en ausencia de ligando. En nuestro análisis todas las dehidrina presentaron una estructura helicoidal en alguno de los segmentos K de las proteínas., (Malik *et al.* 2017) describen que existen dehidrina con segmentos K donde ninguno coincide exactamente con el segmento K canónico, por lo que en el análisis en MEME se

optó por denominar tres de los seis logos arrojados como segmentos K, ya que no existía mayor divergencia entre los aminoácidos relacionados entre los tres logos, donde PrADHN4 y PrADHN6 pertenecen a esta configuración, pudiendo influenciar en la tolerancia al estrés abiótico por frío, ya que se ha descrito que las dehidrina Kn se regulan al alza en frío y sequía (Malik *et al.* 2017).

En cuanto al segmento S, este se reconoce como un sitio de fosforilación, necesitando una región ácida C-terminal para ser fosforilado (Mehta *et al.* 2009), además una vez fosforilado provoca la translocación desde el citoplasma al núcleo (Goday *et al.* 1994), aumentando la capacidad de unión al calcio (Alsheikh *et al.* 2003). Los resultados demostraron que todos los segmentos S de las proteínas deducidas contenían uno o más sitios de fosforilación de la caseína quinasa II (CKII), para las dehidrina SKn, (Malik *et al.* 2017) identificaron pocos cambios en la expresión bajo condiciones de estrés. Las dehidrina evolucionaron para lidiar con los múltiples residuos de serina, o que las múltiples señales de fosforilación sirvan como algún tipo de señal consenso (Rust y Thompson 2011), utilizándolo, así como una medida del nivel de estrés celular.

Para la estructura secundaria, (Strimbeck 2017) propuso que el segmento F puede formar una hélice anfipática capaz de unirse a membranas o proteínas, sin embargo, para estas proteínas en la predicción de la estructura secundaria esto no fue así, no obstante, se encontró este tipo de estructura en el segmento K, en

donde, también es una característica definitoria la existencia de un dominio anfipático formador de una hélice alfa (Close 1996, Malik *et al.* 2017).

Además, en gimnospermas existen otros dos motivos conservados que no han aparecido en nuestros análisis de las proteínas putativas, estos corresponden a los segmentos A y E, los cuales tiene como secuencia consenso EAASYYP y GHGHEGQLTPEEAEQQKH respectivamente, las proteínas con estos segmentos se acumulan en condiciones normales y de estrés por lo que se cree tengan funciones fisiológicas constitutivas, pero no se han demostrado funciones fisiológicas específicas (Perdiguero *et al.* 2015)

En cuanto a la filogenia, (Riley *et al.* 2019) sugieren que el DHN ancestral pertenece a un grupo Kn o SKn, y que YSKn e YKn surgieron directamente en angiospermas, de la misma manera se demostró que las DHN de angiospermas con segmentos F e Y pertenecen a grupos ortólogos diferentes. Un estudio de filogenia no identificó DHN del tipo FSKn en cuatro genomas de briofitas, pero si se encontraron en tres gimnospermas, *P. abies*, *P. glauca* y *G. biloba* (Karas *et al.* 2024). En angiospermas existen las DHN: FSK2 y FKS3, aunque FSK3 solo se encuentra en monocotiledóneas y eudicotiledónea, divergente temprana *Nelumbo nucifera*, así como angiospermas basales *A. trichopoda* y *N. colarata*, sugiriendo que FSK2 podría haber surgido de FSK3 (Melgar y Zelada 2021). Sin embargo, esta especie pertenece a las gimnospermas, donde si existen DHN de

la configuración FSK2, además las pocas DHN que solo poseen segmento K o la combinación KS, se posicionan dentro de las ramas F o Y, sugiriendo que estas DHN pertenecen a cualquiera de estos grupos (Melgar y Zelada 2021), esto también se evidenció en el caso de estas DHN que poseen solo segmento K o la configuración SK, las cuales se posicionan justo en el centro del árbol filogenético. Todas las DHN con estructura Kn y SK en realidad pertenecen a los grupos F o Y, indicando una clasificación basada en la composición de segmento, provocando la agrupación de DHNs con propiedades fisicoquímicas diferentes (Malik *et al.* 2017).

Las nueve dehidrina descritas en la colección de secuencias de *P. andina* representan las primeras proteínas de este tipo caracterizadas en esta especie. Estas proteínas fueron estudiadas para identificar sus posibles funciones biológicas, particularmente en la adaptación al estrés por frío, reconociendo sus sitios de actividad y segmentos conservados. Además, se reconoce a las dehidrina como una potente defensa contra el cambio climático, lo que las posiciona como un recurso valioso para mejorar cultivos alimenticios y forestales. Por lo tanto, estos genes son candidatos prometedores para futuras investigaciones en la ingeniería genética de plantas. En la práctica, se podrían realizar ensayos de estrés en plantas de *P. andina* para la extracción de ARN, con el fin de obtener transcritos de proteínas dehidrina y observar si su cantidad aumenta en presencia del estrés o no se modifica. Es importante resaltar que las

secuencias obtenidas provienen de proteínas ya secuenciadas, lo que limita esta investigación a las proteínas de los códigos utilizados para el alineamiento con la biblioteca de datos SRA.

V. CONCLUSIONES

La hipótesis de esta investigación es aceptada ya que se identificaron nueve proteínas dehidrinas en la colección lecturas de secuenciación de transcritos, de las cuales se encontraron cuatro con la configuración de tipo FSK2 y una con la configuración FSK, las cuales poseen características asociadas a estrés abiótico por bajas temperaturas, además se encontraron otras configuraciones como K3, K2, YSK e YSK2, mostrando una diversidad de seis tipos diferentes de proteína dehidrina en la colección de lecturas de secuenciación de *P. andina*.

La dehidrina relacionada con más secuencias putativas fue COR47-like de *Cryptomeria japonica*, la cual está relacionada a las proteínas PrADHN2, PrADHN3 y PrADHN5, la COR47-like es una proteína asociada a bajas temperaturas o aclimatación al frío. *P. andina*, como especie de hoja perenne, enfrenta el desafío de tolerar el estrés causado por las condiciones de crecimiento, incluyendo las heladas, las cuales provocan deshidratación en las células vegetales. Esto indica que la función de estas proteínas en la especie está vinculada tanto a la resistencia al frío como a contrarrestar la deshidratación durante el invierno.

Este estudio plantea la necesidad de investigar las proteínas que se encuentran en *P. andina*, especialmente las proteínas dehidrina, por lo cual, esta investigación propone su continuación futura mediante el diseño de cebadores o

sondas específicas para las dehidrina de *P. andina*, con la finalidad de evaluar directamente las hojas o frutos de esta especie, detectando la presencia de secuencias que codifican a proteínas dehidrinas, las cuales desempeñan roles fundamentales en la fisiología vegetal.

VI. GLOSARIO

- **BLASTn:** Algoritmo que compara secuencias nucleotídicas contra una base de datos de secuencias nucleotídicas.
- **BLASTp:** Algoritmo que compara secuencias de proteínas contra una base de datos de secuencias de proteínas.
- **BLASTx:** Algoritmo que compara una secuencia nucleotídica traducida en todas las fases contra una base de datos de secuencias de proteínas.
- **BIOEDIT:** Software utilizado para la edición y análisis de secuencias de ADN y proteínas.
- **BOOSTRAPs:** Método estadístico usado para estimar la precisión de los árboles filogenéticos generados en análisis evolutivos.
- **CLUSTALOMEGA:** Herramienta bioinformática que permite realizar alineamientos múltiples de secuencias con mayor precisión y escalabilidad.
- **CLUSTALW:** Algoritmo utilizado para alinear múltiples secuencias de ADN o proteínas.
- **CRIOPROTECCIÓN:** Técnica que permite la preservación de células o tejidos a temperaturas extremadamente bajas, protegiéndolos de daños durante el proceso de congelación.
- **DIOICO:** Término que describe plantas que tienen flores masculinas y femeninas en individuos separados.
- **DHN (Dehydrins):** Grupo de proteínas que se acumulan en plantas bajo condiciones de estrés hídrico y frío, protegiendo las células de daños.
- **EMBRIOGÉNESIS TARDÍA ABUNDANTE (LEA):** Familia de proteínas asociadas con la resistencia de plantas al estrés hídrico durante las últimas fases de desarrollo del embrión.
- **FASTA:** Formato de archivo utilizado para representar secuencias de aminoácidos o nucleótidos.

- **GRAVY (Grand Average of Hydropathy):** Índice que calcula la hidrofobicidad promedio de una proteína a partir de sus aminoácidos.
- **IDP (Proteínas Intrínsecamente Desordenadas):** Proteínas o regiones de proteínas que carecen de una estructura tridimensional fija y bien definida.
- **ILLUMINA:** Plataforma tecnológica utilizada para secuenciación masiva de ADN, basada en el método de secuenciación por síntesis.
- **LIOFILIZACIÓN:** Proceso de secado de muestras que consiste en congelarlas y luego reducir la presión para eliminar el agua por sublimación.
- **LDH (Lactato Deshidrogenasa):** Enzima que cataliza la conversión de lactato a piruvato, importante en el metabolismo energético.
- **MIRISTOILACIÓN:** Modificación postraducciona de proteínas que implica la adición de un ácido mirístico (ácido graso) a una proteína para anclarla a la membrana.
- **MONOICO:** Término que describe plantas que tienen flores masculinas y femeninas en el mismo individuo.
- **NCBI (National Center for Biotechnology Information):** Centro de información biotecnológica que proporciona acceso a bases de datos de secuencias genéticas, publicaciones científicas y otras herramientas bioinformáticas.
- **POIQUILOTERMOS:** Organismos cuya temperatura corporal varía con la temperatura del entorno.
- **PROTEÍNA PUTATIVA:** Proteína predicha o hipotética, cuya función aún no ha sido completamente confirmada.
- **PROTPARAM:** Herramienta bioinformática que permite calcular las propiedades fisicoquímicas de una proteína a partir de su secuencia de aminoácidos.
- **RNASEQ:** Técnica utilizada para analizar el transcriptoma completo de un organismo a partir de la secuenciación masiva de ARN.

- SACRCOTESTA: Capa externa carnosa de la semilla en algunas especies de plantas.
- SRA (Sequence Read Archive): Base de datos de secuencias de lectura corta, alojada por el NCBI, que almacena secuencias obtenidas de experimentos de secuenciación masiva.
- TEMPERATURA DE ALINEAMIENTO: Temperatura óptima a la que las secuencias de ADN o ARN se alinean para formar estructuras estables en ensayos de hibridación.

VII. BIBLIOGRAFÍA

1. Abarzúa A.M., A. Martel-Cea, V. Lobos. 2020. Vegetation–Climate–Megafauna Interactions During the Late Glacial in Pilauco Site, Northwestern Patagonia. pp. 157-173. En: Pilauco: A Late Pleistocene Archaeo-paleontological Site.
2. Alsheikh M.K., B.J. Heyen, S.K. Randall. 2003. Ion binding properties of the dehydrin ERD14 are dependent upon phosphorylation. *J Biol Chem* 278(42): 40882-40889.
3. Álvarez C., C. Le Quesne, M. Rojas-Badilla, V. Rozas, Á. González-Reyes. 2021. Dendrochronological potential of *Prumnopitys andina* (Podocarpaceae) at the southern limit of its range in the Chilean Andes. *New Zealand Journal of Botany* 59(4): 423-439.
4. Amara I., I. Zaidi, K. Masmoudi, M.D. Ludevid, M. Pagès, A. Goday, F. Brini. 2014. Insights into Late Embryogenesis Abundant (LEA) Proteins in Plants: From Structure to the Functions. *American Journal of Plant Sciences* 05(22): 3440-3455.
5. Arora R., K. Taulavuori. 2016. Increased Risk of Freeze Damage in Woody Perennials VIS-À-VIS Climate Change: Importance of Deacclimation and Dormancy Response. *Frontiers in Environmental Science* 4.
6. Atkinson J., M.W. Clarke, J.M. Warnica, K.F. Boddington, S.P. Graether. 2016. Structure of an Intrinsically Disordered Stress Protein Alone and Bound to a Membrane Surface. *Biophys J* 111(3): 480-491.
7. Benoit. 1989. Libro rojo de la flora terrestre de Chile. p.
8. Cano M.J., A.D. Twyford, P.M. Hollingsworth. 2022. Population and conservation genetics using RAD sequencing in four endemic conifers from South America. *Biodiversity and Conservation* 31(13-14): 3093-3112.
9. Carvalho L.M., G. Borelli, A.P. Camargo, M.A. de Assis, S.M.F. de Ferraz, M.B. Fiamenghi, J. José, L.S. Mofatto, S.T. Nagamatsu, G.F. Persinoti, N.V. Silva, A.A. Vasconcelos, G.A.G. Pereira, M.F. Carazzolle. 2019. Bioinformatics applied to biotechnology: A review towards bioenergy research. *Biomass and Bioenergy* 123: 195-224.
10. Cheng Z T.J., Huang X and Wu R. 2002. Wheat LEA genes, PMA80 and PMA1959, enhance dehydration tolerance of transgenic rice (*Oryza sativa* L.). *Molecular Breeding*.
11. Christenhusz M.J.M., J.W. Byng. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa* 261(3).

12. Close T.J. 1996. Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. *Physiologia Plantarum* 97.
13. Donoso. 2006. Las especies arbóreas de los bosques templados de Chile y Argentina. P.
14. Drira M., W. Saibi, F. Brini, A. Gargouri, K. Masmoudi, M. Hanin. 2013. The K-segments of the wheat dehydrin DHN-5 are essential for the protection of lactate dehydrogenase and beta-glucosidase activities in vitro. *Mol Biotechnol* 54(2): 643-650.
15. Dure L., Crouch, M., Harada, J., Ho, T. H. D., Mundy, J., Quatrano, R., ... & Sung, Z. R. 1989. Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Mol Biol*.
16. Dyson H.J., & Wright, P. E. 2002. Coupling of folding and binding for unstructured proteins. *Current opinion in structural biology* 12: 54-60.
17. Ewart K.V., Lin, Q., & Hew, C. L. 1999. Structure, function and evolution of antifreeze proteins. *Cellular and Molecular Life Sciences CMLS* 55: 271-283.
18. Farjon A. 2018. The Kew Review: Conifers of the World. *Kew Bulletin* 73(1).
19. Galau G.A., D.W. Hughes, and L. Dure. 1986. Absciscic acid induction of cloned cotton late embryogenesis abundant (LEA) mRNAs. *Plant Mol*.
20. García N.C.O. 2008. Árboles Nativos de Chile. P.
21. Goday, Anders B. Jensen, Francisco A. Culiáñez-Macia, M. Mar Alba, Mercè Figueras, Joan Serratos, Margarita Torrent, a.M. Pagès. 1994. The Maize Absciscic Acid-Responsive Protein Rab17 1s Located in the Nucleus and Interacts with Nuclear Localization Signals. *American Society of Plant Physiologists*.
22. Gordon-Kamm W.J., P.L. Steponkus. 1984. Lamellar-to-hexagonal phase transitions in the plasma membrane of isolated protoplasts after freeze-induced dehydration. *Proc Natl Acad Sci U S A* 81(20): 6373-6377.
23. Hale M.G., & Orcutt, D. M. . 1987. The physiology of plants under stress. P.
24. Hara M. 2010. The multifunctionality of dehydrins: an overview. *Plant Signal Behav* 5(5): 503-508.
25. Hara M., T. Endo, K. Kamiya, A. Kameyama. 2017. The role of hydrophobic amino acids of K-segments in the cryoprotection of lactate dehydrogenase by dehydrins. *J Plant Physiol* 210: 18-23.

26. Hechenleitner V.P., Gardner, M. F., Thomas, P. I., & Brownless, P. 2005. Las Plantas Amenazadas del Centro-Sur de Chile. Distribución, conservación y propagación. P.
27. Hernández Camacho S., Pérez Molphe Balch, E. M., Alpuche Solís, Á. G., & Morales Domínguez, J. F. 2017. Identification and evolutionary relationships of partial gene sequences from dehydrin group in three species of cacti.
28. Hernandez P. 2012. Estructuración genética en tres poblaciones marginales de *Prumnopitys andina*, una especie amenazada.
29. Hughes S., S.P. Graether. 2011. Cryoprotective mechanism of a small intrinsically disordered dehydrin protein. *Protein Sci* 20(1): 42-50.
30. Hughes S.L., V. Scharf, J. Malcolmson, K.A. Hogarth, D.M. Martynowicz, E. Tralman-Baker, S.N. Patel, S.P. Graether. 2013. The importance of size and disorder in the cryoprotective effects of dehydrins. *Plant Physiol* 163(3): 1376-1386.
31. Jarillo J.A., Capel, J., Leyva, A., Martínez-Zapater, J. M., & Salinas, J. . 1994. Two related low-temperature-inducible genes of *Arabidopsis* encode proteins showing high homology to 14-3-3 proteins, a family of putative kinase regulators. *Plant molecular biology* 25: 693-704.
32. Karas M., D. Veselenyiova, E. Boszoradova, P. Nemecek, Z. Gersi, J. Moravcikova. 2024. Comparative Analysis of Dehydrins from Woody Plant Species. *Biomolecules* 14(3).
33. Khan R., R.S. Hill, V.M. Dorken, E. Biffin. 2022. Detailed seed cone morpho-anatomy of the *Prumnopityoid* clade: an insight into the origin and evolution of *Podocarpaceae* seed cones. *Ann Bot* 130(5): 637-655.
34. Malik A.A., M. Veltri, K.F. Boddington, K.K. Singh, S.P. Graether. 2017. Genome Analysis of Conserved Dehydrin Motifs in Vascular Plants. *Front Plant Sci* 8: 709.
35. Maszkowska J., J. Debski, A. Kulik, M. Kistowski, M. Bucholc, M. Lichocka, M. Klimecka, O. Sztatelman, K.P. Szymanska, M. Dadlez, G. Dobrowolska. 2019. Phosphoproteomic analysis reveals that dehydrins ERD10 and ERD14 are phosphorylated by SNF1-related protein kinase 2.10 in response to osmotic stress. *Plant Cell Environ* 42(3): 931-946.
36. Mehta P.A., K.C. Rebala, G. Venkataraman, A. Parida. 2009. A diurnally regulated dehydrin from *Avicennia marina* that shows nucleo-cytoplasmic localization and is phosphorylated by Casein kinase II in vitro. *Plant Physiol Biochem* 47(8): 701-709.

37. Melgar A.E., A.M. Zelada. 2021. Evolutionary analysis of angiosperm dehydrin gene family reveals three orthologues groups associated to specific protein domains. *Sci Rep* 11(1): 23869.
38. MMA M.d.M.A. 2012. FICHA DE ANTECEDENTES DE ESPECIE: *Prumnopitys andina*.
39. MMA M.d.M.A. 2018. Biodiversidad de chile. P.
40. Nylander M. 2001. Stress-induced accumulation and tissue-specific localization of dehydrins in *Arabidopsis thaliana*. *Plant molecular biology* 45.
41. Ohkubo T., A. Kameyama, K. Kamiya, M. Kondo, M. Hara. 2020. F-segments of *Arabidopsis* dehydrins show cryoprotective activities for lactate dehydrogenase depending on the hydrophobic residues. *Phytochemistry* 173: 112300.
42. Palmer S.R., R. De Villa, S.P. Graether. 2019. Sequence composition versus sequence order in the cryoprotective function of an intrinsically disordered stress-response protein. *Protein Sci* 28(8): 1448-1459.
43. Perdiguero P., C. Collada, A. Soto. 2014. Novel dehydrins lacking complete K-segments in Pinaceae. The exception rather than the rule. *Front Plant Sci* 5: 682.
44. Perdiguero P., Á. Soto, C. Collada. 2015. Comparative analysis of *Pinus pinea* and *Pinus pinaster* dehydrins under drought stress. *Tree Genetics & Genomes* 11(4).
45. Riley A.C., D.A. Ashlock, S.P. Graether. 2019. Evolution of the modular, disordered stress proteins known as dehydrins. *PLoS One* 14(2): e0211813.
46. Rust H.L., P.R. Thompson. 2011. Kinase consensus sequences: a breeding ground for crosstalk. *ACS Chem Biol* 6(9): 881-892.
47. Shi J., M. Liu, Y. Chen, J. Wang, C. Lu. 2015. Heterologous expression of the dehydrin-like protein gene AmCIP from *Ammopiptanthus mongolicus* enhances viability of *Escherichia coli* and tobacco under cold stress. *Plant Growth Regulation* 79(1): 71-80.
48. Smith M.A., S.P. Graether. 2022. The Disordered Dehydrin and Its Role in Plant Protection: A Biochemical Perspective. *Biomolecules* 12(2).
49. Steponkus P.L., Uemura, M., Joseph, R. A., Gilmour, S. J., & Thomashow, M. F. 1998. Mode of action of the COR15a gene on the freezing tolerance of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* 95: 14570-14575.
50. Strimbeck G. 2017. Hiding in plain sight: the F segment and other conserved features of seed plant SK(n) dehydrins. *Planta* 245(5): 1061-1066.

51. Sun Z., S. Li, W. Chen, J. Zhang, L. Zhang, W. Sun, Z. Wang. 2021. Plant Dehydrins: Expression, Regulatory Networks, and Protective Roles in Plants Challenged by Abiotic Stress. *Int J Mol Sci* 22(23).
52. Thomashow M.F. 1998. Role of cold-responsive genes in plant freezing tolerance. *Plant physiology* 118(1): 1-8.
53. Thomashow M.F. 2001. So what's new in the field of plant cold acclimation? Lots! *Plant physiology* 125(1): 89-93.
54. Tunnacliffe A., M.J. Wise. 2007. The continuing conundrum of the LEA proteins. *Naturwissenschaften* 94(10): 791-812.
55. Wei H., Y. Yang, M.E. Himmel, M.P. Tucker, S.Y. Ding, S. Yang, R. Arora. 2019. Identification and Characterization of Five Cold Stress-Related Rhododendron Dehydrin Genes: Spotlight on a FSK-Type Dehydrin With Multiple F-Segments. *Front Bioeng Biotechnol* 7: 30.
56. West J., M. Bhattacharya. 2016. Intelligent financial fraud detection: A comprehensive review. *Computers & Security* 57: 47-66.
57. Wisniewski M., R. Webb, R. Balsamo, T.J. Close, X.M. Yu, M. Griffith. 2002. Purification, immunolocalization, cryoprotective, and antifreeze activity of PCA60: A dehydrin from peach (*Prunus persica*). *Physiologia Plantarum* 105(4): 600-608.
58. Yu X.M., M. Griffith. 2001. Winter rye antifreeze activity increases in response to cold and drought, but not abscisic acid. *Physiol Plant* 112(1): 78-86.
59. Zamecnik J. 1992. Interaction of antifreeze proteins from cold hardened cereal seedlings with ice nucleation active bacteria. *Cryobiology* 29: 718-719.