



UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE CIENCIAS FORESTALES
INGENIERÍA EN BIOTECNOLOGÍA VEGETAL

ARN NO CODIFICANTE DE CADENA LARGA EN RESPUESTA A ESTRÉS
ABIÓTICO EN HÍBRIDOS DE EUCALYPTUS NITENS x EUCALYPTUS
GLOBULUS

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

POR: David Alfredo Lisboa Jacobsen

Profesor Guía: Marta Fernández

Concepción, Chile 2023

© 2023

David Alfredo Lisboa Jacobsen

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

ARN NO CODIFICANTE DE CADENA LARGA EN RESPUESTA A ESTRÉS
ABIÓTICO EN HÍBRIDOS DE EUCALYPTUS NITENS x EUCALYPTUS
GLOBULUS

Profesor Guía



Marta Fernández Reyes.
Profesora Asociada
Bióloga, Dra.

Profesor Guía



Rodrigo Hasbún Zaror.
Profesor Asociado
Ingeniero Forestal, Dr.

Calificación de la Habilitación Profesional:

Marta Fernández R. :

Rodrigo Hasbún Z. :

TABLA DE CONTENIDO

I. INTRODUCCIÓN.....	1
II. METODOLOGÍA	9
2.1. Anotación del transcriptoma del híbrido de eucalipto.....	9
2.2 Filtro por ORF.....	9
2.3 Potencial codificante.....	9
2.4 Masking.....	10
2.5 Filtro por homología.....	10
2.6 Filtro estructural.....	10
2.7 Búsqueda de ARNs no codificantes de cadena larga.	10
2.8 Clasificación e identificación de lncRNA putativos de interés.....	11
2.9 Anotación funcional de las proteínas compañeras.	12
III. RESULTADOS	13
IV. DISCUSIÓN.....	22
4.1 Filtro por Marco de lectura abierto (ORF).....	22
4.2 Softwares usados para la identificación de lncRNA.	22
4.3 lncRNA putativos identificados y sus proteínas blanco en respuesta a estrés por frío.	29
V. CONCLUSIONES	33
VI. BIBLIOGRAFÍA	34

ÍNDICE DE TABLAS

Tabla 1: lncRNA conocidos que están involucrados en la respuesta a frío en especies vegetales	4
Tabla 2: Resumen de los identificadores asociados a los lncRNA putativos identificados	17
Tabla 3: Clasificación y caracterización de cada lncRNA de acuerdo a FEELnc.....	18
Tabla 4: Resultados del análisis de ontología de genes: proceso biológico de las proteínas compañeras	19
Tabla 5: Resultados del análisis de ontología de genes: función molecular asociada a las proteínas compañeras	20
Tabla 6: Dominios proteicos identificados en las proteínas compañeras	20
Tabla 7: Propiedades de los marco de lectura abierto (ORF)	22
Tabla 8: Resumen de las características principales de cada software usado para la identificación de los lncRNA	28

ÍNDICE DE ILUSTRACIONES

Figura 1: Arquetipos de mecanismos de acción de los lncRNA	6
Figura 2: Línea de trabajo empleada para identificar ARNs no codificantes de cadena larga.....	16
Figura 3: Expresión diferencial de las proteínas compañeras	21

RESUMEN

Los retos que presentan las organizaciones productivas en cuanto a investigación y desarrollo para la producción de madera van en incremento. Los progresivos aumentos en la demanda de las comunidades, el mercado y los desafíos que propone el cambio climático exigen mayores esfuerzos en la búsqueda de eficacia y eficiencia en los procesos de cultivo orientados a la producción de madera. Ante los múltiples factores limitantes para el mejoramiento genético, provenientes del desconocimiento de los procesos moleculares involucrados en el desarrollo de *Eucalyptus*, se hace necesario proponer nuevas metodologías para revelar la maquinaria molecular involucrada en la respuesta de estrés abiótico. El presente estudio buscó revelar la influencia de ARNs no codificante de cadena larga (lncRNAs) en la respuesta a estrés por baja temperatura. La metodología propuesta buscó complementar las fortalezas y debilidades de diferentes herramientas bioinformáticas para dilucidar, en base a la expresión de lncRNA en respuesta a aclimatación a baja temperatura, el rol de los lncRNAs en la respuesta a estrés por baja temperatura en un híbrido de *Eucalyptus globulus x Eucalyptus nitens*. Como resultado, se identificaron 22 lncRNAs que podrían participar en la modulación de la expresión de genes de respuesta a estrés por baja temperatura en el híbrido. Dichas proteínas podrían estar involucradas en diferentes procesos de respuesta a estrés por frío. En base a los resultados y la información asociada con análisis en otras especies vegetales se concluye que, en base a la abundancia de transcritos de los lncRNA y la expresión *in silico* de los genes objetivos putativos, existiría regulación de la expresión génica en respuesta a estrés por bajas temperaturas por parte de los lncRNAs.

ABSTRACT

The challenges presented by productive organizations in terms of research and development to produce wood are increasing. The progressive increases in the demand of the communities, the market and the challenges posed by climate change demand greater efforts in the search for effectiveness and efficiency in the cultivation processes oriented to the production of wood. Given the multiple limiting factors for genetic improvement, coming from the ignorance of the molecular processes involved in the development of *Eucalyptus*, it is necessary to propose new methodologies to reveal the molecular machinery involved in the abiotic stress response. The present study sought to reveal the influence of long-chain non-coding RNAs (lncRNAs) on the response to low-temperature stress. The proposed methodology sought to complement the strengths and weaknesses of different bioinformatics tools to elucidate the role of lncRNAs in the response to low-temperature stress in a *Eucalyptus globulus* x *Eucalyptus nitens* hybrid. As a result, 22 lncRNAs were identified that could participate in the modulation of the cold responsive genes associated in the hybrid. Based on the results and the information associated with analyzes in other plant species, it is concluded that, based on the abundance of lncRNA transcripts and the *in silico* expression of the putative target genes, there would be regulation of gene expression in response to stress by low temperatures by the lncRNAs.

I. INTRODUCCIÓN

Las especies de árboles madereros, como el eucalipto, desempeñan un papel fundamental al proporcionar materias primas para las industrias del papel y la madera. La creciente demanda de papel ha impulsado un aumento constante en la expansión de las plantaciones de *Eucalyptus globulus* en la región del Biobío de Chile, alcanzando un total de 247,967 hectáreas (Gysling *et al.* 2018). Sin embargo, las plantas jóvenes de esta especie son altamente susceptibles al estrés causado por bajas temperaturas, lo que puede comprometer su crecimiento y desarrollo (Costa e Silva *et al.* 2008). Este fenómeno, junto con la creciente necesidad de árboles más resistentes a las bajas temperaturas, ha llevado a la exploración de la hibridación entre *Eucalyptus globulus* y *Eucalyptus nitens* en busca de clones que combinen el rápido crecimiento de la primera con la tolerancia al frío de la segunda.

La respuesta de las plantas al estrés ambiental es un proceso complejo que involucra cambios fisiológicos, bioquímicos y moleculares profundos. Se ha demostrado que los ARN no codificantes de cadena larga (lncRNA) desempeñan un papel esencial en la regulación de la respuesta al estrés abiótico y biótico en diversas especies (Di *et al.* 2014, Qin *et al.* 2017). Estos lncRNA, definidos como ARN con al menos 200 nucleótidos que no codifican proteínas, están implicados en una variedad de mecanismos moleculares, incluida la regulación de la expresión génica, el splicing alternativo y la modificación de la cromatina (Zhang *et al.* 2010, Hansen *et al.* 2013).

En estas plantaciones, los individuos juveniles son más susceptibles a sufrir daño por las bajas temperaturas, por lo que la tolerancia al estrés por frío puede ser un factor determinante en el éxito del cultivo (Costa e Silva *et al.* 2008). El estrés por bajas temperaturas (bajo los 20 °C) y el estrés por congelamiento (bajo los 0 °C) ocasionan desajustes a las plantas, afectando la homeostasis de las plantas y aumentando el estrés

hídrico en estos individuos y como consecuencia el crecimiento y desarrollo se ve afectado negativamente. Algunos ejemplos de desregulación del metabolismo son: inhibición de la captación de agua debido a la osmosis provocada por el hielo extracelular (Kazemi y Maali 2018, Chinnusamy *et al.* 2007, Warren 1998), desajuste en la absorción de nutrientes, alteración de la conformación de las proteínas y desequilibrio en el metabolismo celular (Chinnusamy *et al.* 2007).

En Chile se ha usado la especie *Eucalyptus nitens* (con 104.569 ha en la región del Biobío) para establecer plantaciones en zonas con mayor frecuencia de heladas, en las que en *E. globulus* no es capaz de sobrevivir, ya que *E. nitens* más tolerante a bajas temperaturas y tiene un crecimiento más rápido (Ipinza *et al.* 2014, Gysling *et al.* 2018). La hibridación entre estas 2 especies busca obtener clones con el rápido crecimiento de *E. globulus* y la tolerancia de *E. nitens*.

La respuesta de las plantas al estrés ambiental es un proceso complejo que involucra cambios fisiológicos, bioquímicos y moleculares profundos. Se ha demostrado que los ARN no codificantes de cadena larga (lncRNA) desempeñan un papel esencial en la regulación de la respuesta al estrés abiótico y biótico en diversas especies (Di *et al.* 2014, Qin *et al.* 2017). Estos lncRNA, definidos como ARN con al menos 200 nucleótidos que no codifican proteínas, están implicados en una variedad de mecanismos moleculares, incluida la regulación de la expresión génica, el splicing alternativo y la modificación de la cromatina (Zhang *et al.* 2010, Hansen *et al.* 2013).

Se ha descubierto que los lncRNAs son reguladores importantes en múltiples vías en diferentes especies, incluyendo animales y plantas. Se han descrito diferentes mecanismos biológicos en el que lncRNAs median la expresión génica: imitando el gen endógeno objetivo para adherirse a ciertos miARN, resultando en un incremento o

disminución de la expresión de ciertos ARNms; participación en trans-splicing, cambiando los patrones del splicing, teniendo como consecuencia una isoforma alternativa y alteradas; adherencia a proteínas, cambiando la translocación de las proteínas chaperonas objetivo y modulación de la modificación de cromatinas mediante la marca represivas de la deposición de histonas, modificando sus dominios o interactuando con cromatinas para modular su estructura nuclear (Zhang *et al.* 2010, Hansen *et al.* 2013, Csorba *et al.* 2014 y Shichino *et al.* 2014).

En el contexto de las especies vegetales, se han identificado lncRNA en diversas especies, como *Arabidopsis thaliana*, *Oryza sativa* y *Zea mays*, y se ha demostrado su participación en procesos clave como la transición floral y el desarrollo de tejidos (Ding *et al.* 2012, Li *et al.* 2014). Incluso se han identificado lncRNA expresados diferencialmente en el xilema de madera en el modelo *Populus*, sugiriendo un papel importante en la formación de la madera (Chen *et al.* 2015).

Se han identificado algunos lncRNA en respuesta a frío en plantas, incluyendo *A. thaliana*, *Vitis vinifera*, *Manihot esculenta* y *Medicago truncatula* (Tabla 1). COOLAIR, COLDAIR y COLDWRAP participan en la regulación del proceso de vernalización a través del silenciamiento del gen FLC (Flowering Locus C). Mientras que en *M. esculenta*, los ncM9574 gene, NcP12248 y NcP456, están involucrados en la regulación de los genes OXS3 (Oxidative Stress 3), codificante para proteína de la familia de la policétido sintasa enoil reductasa y WRKY DNA-bindingprotein 33, respectivamente. En *V. vinifera*, los lncRNA VIT_216s0100n00030, VIT_202s0025n00100 y VIT_200s0225n00020 están involucrados en la regulación del factor de transcripción CBF4, factor de transcripción WRKY 41 y Subunidad 5 de la NADH deshidrogenasa. En *M. truncatula*, el lncRNA MtCIR1 participa en la regulación de proteínas de unión a repetición/DRE (CBF) los cuales son factores reguladores de la transcripción

fundamentales de genes de respuesta al frío (COR) durante el tratamiento con frío en múltiples especies de plantas (Thomashow 1999).

Tabla 1: lncRNA reportados por otros estudios en planta y que han sido involucrados en la respuesta a baja temperatura en especies vegetales.

LncRNA	Organismo	Mecanismo de regulación	Expresión	Referencia
COLDAIR	<i>A. thaliana</i>	Modificación de histonas	Inducido	Heo & Sung 2011
COOLAIR	<i>A. thaliana</i>	Modificación de histonas	Inducido	Csorba <i>et al.</i> 2014
COLDWRAP	<i>A.thaliana</i>	Formación de loops en cromatina	Inducido	Kim & Sung 2017
VIT_216s0100n00030	<i>V. vinifera</i>	Elemento regular en cis	Inducido	Wang <i>et al.</i> 2019
VIT_202s0025n00100	<i>V. vinifera</i>	Elemento regular en cis	Inducido	Wang <i>et al.</i> 2019
VIT_200s0225n00020	<i>V. vinifera</i>	Elemento regular en trans	Reprimido	Wang <i>et al.</i> 2019
NcM9574	<i>M. esculenta</i>	Elemento regular en cis	Inducido	Suksamran <i>et al.</i> 2020
NcP12248	<i>M. esculenta</i>	Elemento regular en cis	Reprimido	Suksamran <i>et al.</i> 2020
NcP456	<i>M. esculenta</i>	Elemento regular en trans	Reprimido	Suksamran <i>et al.</i> 2020
MtCIR1	<i>M. truncatula</i>	Elemento regular en cis	Inducido	Zhao <i>et al.</i> 2020

Recientemente, se ha avanzado en la identificación de lncRNA en *Eucalyptus grandis* y se ha explorado su función en la regulación de genes durante la biosíntesis de celulosa y en respuesta al estrés por frío (Lin *et al.* 2019, Zhan *et al.* 2021, Chen *et al.* 2021). Estos estudios abren nuevas perspectivas sobre el papel de los lncRNA en *Eucalyptus spp.*

La biogénesis de los lncRNA puede ser dividida en 5 categorías, de acuerdo con el sitio de transcripción de la Pol II: lncRNA-sentido que son transcritos en la misma hebra que exones, lncRNA-antisentido que son transcrito en la hebra contraria que los exones, lncRNA-intrónicos transcritos en intrones, lncRNA-intergénicos ubicados entre 2

distintos genes y lncRNA-promotores que provienen de una región promotora de genes codificantes a proteínas (Waititu *et al.* 2020). Se han descrito diferentes mecanismos de como los lncRNA participan en la regulación: remodelamiento de la proteína, supresión de la transcripción, empalme del ARN y potenciador transcripcional (Chen & Penfield 2018, Kim & Sung 2013, Marquardt *et al.* 2014, Song *et al.* 2019, Bardou *et al.* 2014). Múltiples lncRNA han sido identificados en *Arabidopsis* bajo estrés por sequía, frío, salinidad y calor (Di *et al.* 2014). A pesar de que los lncRNA son pobremente conservados, se puede encontrar una baja conservación de secuencias a nivel de transcrito entre subsets de lncRNAs entre especies (Zhao *et al.* 2021).

Se han propuestos diferentes arquetipos para explicar cómo los lncRNA ejercen su función, mostrando que las estructuras secundarias y terciarias de los lncRNA tienen un rol en el mecanismo de acción de estos (Ramakrishnaiah *et al.* 2020).

Arquetipo 1: el lncRNA puede actuar como enlaces uniendo riboproteínas (RBP) y factores de transcripción en diferentes loci en el ADN en la región promotora para comenzar la transcripción (Figura 1).

Arquetipo 2: lncRNA puede vincular proteínas como los complejos de factor de remodelación de cromatina PRC2 y LSD1 a los loci que desempeñan un papel crucial en la regulación epigenética (Figura 1).

Arquetipo 3: lncRNA puede interactuar con miRNAs que están involucrados en el proceso postraduccional actuando como una esponja o señuelo (Figura 1).

Arquetipo 4: los lncRNA pueden formar una estructura de ácido nucleico de tres cadenas llamada R Loop, que es un mecanismo común que regula la expresión génica (Figura 1).

Arquetipo 5: Los transcritos naturales antisentido son un tipo de LncRNA que mantienen la estabilidad de su transcripción de codificación correspondiente después de la transcripción (Figura 1).

Arquetipo 6: LncRNA forma una triple hélice de estructura triplex con el ADN, esta estructura interactúa con proteínas que pueden estabilizarla o desenrollarla (Figura 1).

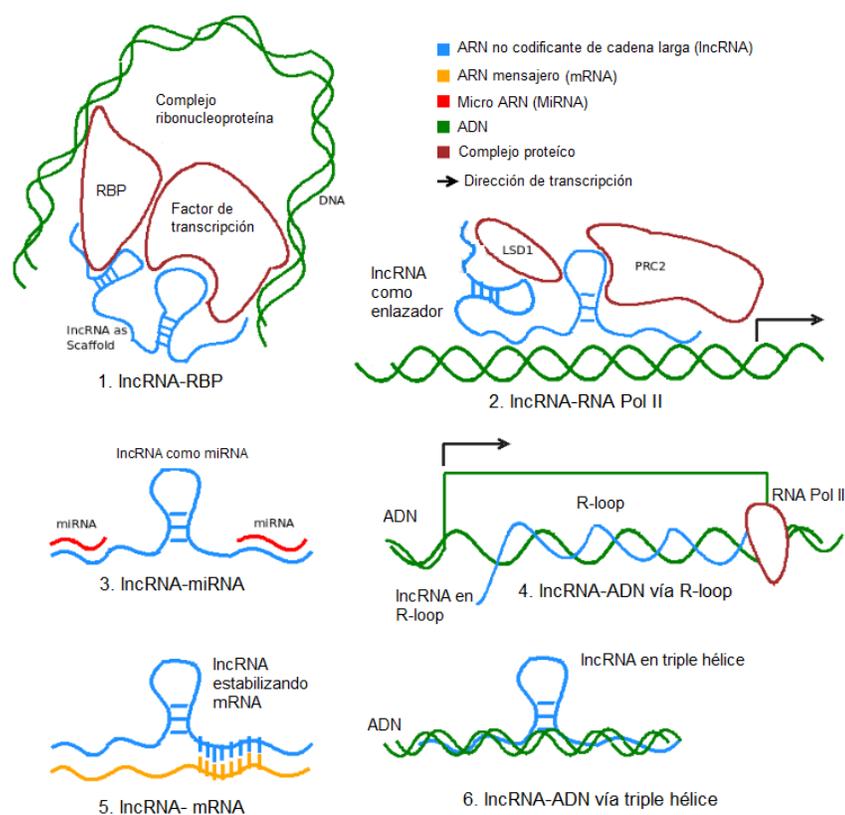


Figura 1: Arquetipos de diferentes mecanismos de acción en los cual los lncRNA podrían ejercer su función. Modificado de Ramakrishnaiah *et al.* 2020.

Además, Se ha mostrado que algunos lncRNA son capaces de dividir la región nuclear en diferentes compartimentos y participar en la organización de regiones multicromosomales ya que tienen afinidad a cromosomas a través de factores de matriz

nuclear, por lo que proveen una ventaja favorable para lncRNA para interactuar con elementos funcionales del ADN relacionados con la regulación transcripcional (Engreitz *et al.* 2016). Por lo que lncRNAs no solo pueden interferir con la expresión de genes codificantes cercanos, sino que también se pueden esparcir en el núcleo, cerca de sitios de afinidad espacial y regular la expresión de genes en cromosomas (West *et al.* 2014, Engreitz *et al.* 2013).

Los lncRNAs pueden alcanzar su función vía reclutamiento, inhibición o actuando indirectamente a través de la organización del genoma y transcripción (Long *et al.* 2017). Un ejemplo es el lncRNA ASCO (Alternative Splicing Competitor) que es regulador de empalmes alternativos en *Arabidopsis* mediante interacciones con factores de empalme (Rigo *et al.* 2020).

La comprensión de los mecanismos de biosíntesis y crecimiento secundario es fundamental para los programas de mejoramiento genético en especies leñosas, dada su importancia en la producción de productos forestales. Para abordar el desafío de distinguir entre transcritos codificantes y no codificantes, se han desarrollado varios programas de bioinformática, como CNCI, CPC2, FEElnc y LncFinder (Sun *et al.* 2013, Kang *et al.* 2017, Wucher *et al.* 2017, Han *et al.* 2018), cada uno con sus propias ventajas y requerimientos de datos.

Dada la relación de los lncRNAs con el crecimiento y desarrollo de las plantas, la presente investigación busca responder si existe una modulación en la expresión de ARNs no codificantes en la respuesta a aclimatación al frío en el híbrido *Eucalyptus nitens* x *Eucalyptus globulus*. Para responder esta pregunta se plantea la siguiente hipótesis de trabajo: Existe una modulación en la expresión de ARNs no codificantes de

cadena larga en la respuesta a aclimatación al frío en plantas juveniles de un híbrido de *Eucalyptus nitens x Eucalyptus globulus*.

En base a esto se planteó como objetivo general: Revelar asociación los ARNs no codificantes de cadena larga y la respuesta a estrés por frío en librerías de expresión en el híbrido de *Eucalyptus nitens x Eucalyptus globulus* con el estado de aclimatación al frío de las plantas. Además, se plantearon los siguientes objetivos específicos: Discriminar transcritos codificantes para lncRNA y sus respectivos genes blanco, mediante criterios estructurales en un transcriptoma de referencia de un híbrido *E. nitens x E.globulus* expuesto a bajas temperaturas; evaluar expresión *in silico* de los genes blanco putativos de los lncRNA en condición no-aclimatado y aclimatado en un híbrido *E.nitens x E.globulus* expuesto a bajas temperaturas de genes blanco y asociar abundancia relativa de lncRNA con genes blanco diferencialmente expresados en condición no-aclimatado y aclimatado en un híbrido *E.nitens x E.globulus* expuesto a bajas temperaturas.

II. METODOLOGÍA

2.1. Anotación del transcriptoma del híbrido de eucalipto.

Se usó transcriptoma el ensamblado por Apablaza *et al.* (2021), quienes para generar el transcriptoma global en respuesta a la baja temperatura de *E. nitens x E. globulus* quienes realizaron un control de calidad de lectura se utilizando FastQC. (Andrés 2010). Las lecturas de baja calidad, ADN no codificante fragmentos, adaptadores, barcodes y las secuencias enriquecidas con homopolímeros se filtraron y las 25 pb de 5 ' extremos se recortaron utilizando el software Prinseq-lite ([http:// prinseq.sourceforge.net](http://prinseq.sourceforge.net)). El diseño del estudio consideró la construcción para transcriptoma completo (NA, AAH, ADH) en respuesta a la aclimatación al frío. Se utilizó la aproximación estándar del ensamblaje de novo para obtener todos los transcriptomas utilizando el ensamblador Trinity (Grabherr *et al.* 2011). En este estudio, se usó el transcriptoma generado con un kmer=32 y pares de base igual o mayores 200.

2.2 Filtro por ORF.

Se usó el software TransDecoder (disponible en <https://github.com/TransDecoder/TransDecoder>) para determinar el largo del marco de lectura (ORF, por sus siglas en inglés) de cada contig del transcriptoma. Se consideraron los que tuviesen un ORF menor a 100 aa.

2.3 Potencial codificante.

Se determinó el potencial codificante con los softwares CNCI y CPC2 (Sun *et al.* 2013, Kang *et al.* 2017). Los que presentaron potencial codificante fueron descartados.

2.4 Masking.

Se usó el software RepeatMasker (disponible en <https://www.repeatmasker.org/RepeatMasker/>) para detectar repeticiones intercaladas y secuencias de baja complejidad.

2.5 Filtro por homología.

Se filtraron los contigs sin potencial codificante por homología a través de la herramienta BlastX (disponible en <https://blast.ncbi.nlm.nih.gov/Blast.cgi>), usando las bases proteicas no redundantes (disponible en <https://www.ncbi.nlm.nih.gov/refseq/>) y de proteínas de *E. grandis* (disponible en <https://phytozome-next.jgi.doe.gov/>). Se filtraron por e-value e^{-3} .

2.6 Filtro estructural

Los contigs restantes fueron filtrados por su estructura, usando el software desarrollado por Wheeler (2013) HMMER y la base de datos Pfam-A 32.0 (disponible en <https://www.ebi.ac.uk/interpro/>). Fueron considerados los contigs que no posean dominios proteicos (con un e-value menor a e^{-3}) y estos serán los lncRNA putativos.

2.7 Búsqueda de ARNs no codificantes de cadena larga.

Se identificaron los lncRNA putativos usando se usó el paquete de R “lncFinder” (Han *et al.* 2019). Para identificar se requirieron diferentes archivos de entrada: secuencias a identificar (obtenidas de la función run_RNAfold), archivo de frecuencias (obtenidas de la función make_frequencies) y un modelo svm (generado por la función build_model).

Para obtener las secuencias a identificar se usaron los lncRNA putativos obtenidos en el paso 2.6.

Para obtener el archivo de frecuencias se usaron las secuencias de transcritos de *E. grandis* v2.0 (disponible en <https://phytozome-next.jgi.doe.gov/>) y lncrna identificados en *E. grandis* por Linn *et al.* (2019).

Para obtener el archivo de modelo svm requerido para el análisis se usaron los lncrna identificados en *E. grandis* por Linn *et al.* (2019) y la anotación de secuencias CDS *E. grandis* v2.0 (disponible en <https://phytozome-next.jgi.doe.gov/>) y transcritos de *E. grandis* v2.0 (disponible en <https://phytozome-next.jgi.doe.gov/>). Se consideraron las secuencias “no codificantes” y las que tuviesen un ORF igual o menor a 100 aa.

2.8 Clasificación e identificación de lncRNA putativos de interés.

Para la clasificación se requirió mapear las secuencias no codificantes contra el genoma de *E. grandis* v2.0 disponible en Phytozome con el software G-MAP (Wu y Watanabe 2005). Para convertir archivo de salida de formato gff3 a gtf, se usó el software Gffread (Disponible en <https://github.com/gpertea/gffread>). Luego se usó el script Feelnc classifier para clasificar el tipo de interacción de los lncRNAs con su proteína compañera putativa en: intergénicos (divergentes, convergentes o misma hebra) en exónicos o intrónicos o en lncRNAs génicos (sobrepuestos, contenedores o anidados) río arriba o río abajo (Wucher *et al.* 2017). Esta clasificación depende según la orientación y la ubicación relativa de los dos tipos de ARN, entregando así información sobre la relación espacial entre el lncRNA y su proteína compañera (o proteína blanco).

Estos resultados se aunaron con el análisis de expresión diferencial de proteínas generado empleando el paquete de R DEXSeq (Anders *et al.* 2012), donde se evaluaron las comparaciones NA ADH, NA AAH y AAH ADH, los resultados se filtraron con un FDR < 0.05 (Apablaza *et al.* 2021); y con la tabla de abundancia generada usando edgeR, filtradas por pvalue y fdr 0.05 (Conejera 2022, McCarthy *et al.* 2012).

Los transcritos clasificados que estaban expresados diferencialmente, y que tenían variación en la abundancia en cada condición (NA, ADH, AAH) fueron evaluados mediante Blast con las bases de datos NCBI, RNACentral. Paralelamente, la proteína compañera de estos fue comparada con la base de datos Uniprot.

Se consideraron como lncRNA putativos de interés los que cumplieron con las siguientes condiciones: presentes en el transcriptoma (abundancia > 0) y que su proteína compañera estuviera diferencialmente expresada. Se generó un mapa de calor con las proteínas compañeras diferencialmente expresadas.

2.9 Anotación funcional de las proteínas compañeras.

Las proteínas compañeras fueron anotadas usando la herramienta DAVID desarrollada por Huang *et al.* (2021) para conocer función molecular, proceso biológico y dominio funcional.

III. RESULTADOS

1. Ensamble de novo

El ensamble *de novo* generado contiene 265.985 transcritos con secuencias mayores o iguales a 200pb. Las estadísticas del ensamble son las siguientes:

-Recuentos de transcritos.

Gen total de la trinidad': 129565

Transcripciones totales de Trinity: 265985

Porcentaje GC: 42,64

-Estadísticas basadas en todos los contigs de transcritos:

Contig N10: 5609

Contig N20: 4414

Contig N30: 3649

Contig N40: 3053

Contig N50: 2523

Longitud mediana del contig: 702

Contig promedio: 1323.39

Total bases montadas: 352002483

-Estadísticas basadas en sólo la isoforma más larga por gen:

-Contig N10: 5169.

-Contig N20: 3912.

-Contig N30: 3043.

- Contig N40: 2280.
- Contig N50: 1543.
- Longitud mediana del contig: 385.
- Contig promedio: 798,95.
- Total bases montadas: 103515708.

2. Filtro ORF:

Luego de que el filtro por ORF fuera aplicado, quedaron 132.273 transcritos con un ORF igual o menor a 100.

3. Filtro por potencial codificante:

El potencial codificante de cada secuencia fue determinado por los softwares CNCI y CPC2 por separado. Se conservaron las secuencias en común, quedando así 124.646 secuencias.

4. Filtro por homología:

Se descartaron las secuencias con un e-value menor a e^{-10} , quedando así 82.854 secuencias.

5. Filtro estructural:

Se descartaron las secuencias que tuvieran un dominio proteico (considerando un e-value menor a e^{-3}). Bajo este criterio, 91971 secuencias no presentaron dominios proteicos

6. Identificación y asociación a una proteína blanco:

El modelo estadístico generado por lncFinder identificó 31.702 potenciales lncRNA, de los cuales 790 fueron clasificados por FEELnc y asociados a una proteína blanco (Figura 2).

De estos contigs, 22 presentaron abundancia para los lncRNA putativos, y al mismo tiempo sus transcritos compañeros (o proteínas blanco) estuvieron diferencialmente expresados (Tabla 2). La descripción del tipo de interacción con su proteína blanco se encuentra en la Tabla 3

7. Anotación

Sus respectivos transcritos compañeros fueron anotados por ontología (Tabla 4, 5 y 6).

Figura 2. Línea de trabajo empleada para identificar ARNs no codificantes de cadena larga en muestra de hoja del híbrido *Eucalyptus nitens* x *Eucalyptus globulus*.

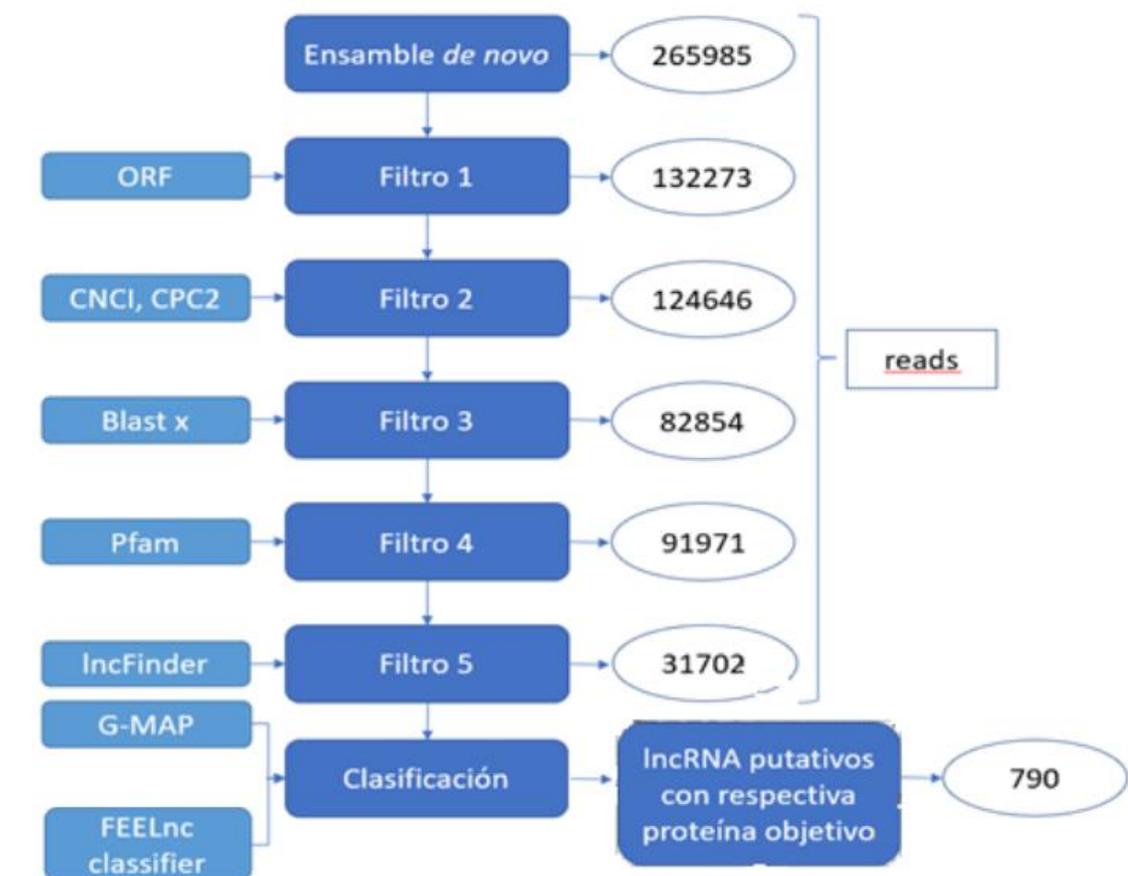


Tabla 2: Tabla resumen de los identificadores asociados a los lncRNA putativos.

lncRNA putativo	Gen del transcrito objetivo	Transcrito ARN objetivo	Mejor hit en Arabidopsis para el transcrito ARN objetivo
Contig 1	Eucgr.K02431	Eucgr.K02431.1	AT2G36810.1
Contig 2	Eucgr.C00095	Eucgr.C00095.1	AT4G29900.1
Contig 3	Eucgr.F01070	Eucgr.F01070.1	AT2G20900.3
Contig 4	Eucgr.J00363	Eucgr.J00363.1	AT5G41040.1
Contig 5	Eucgr.C03713	Eucgr.C03713.1	AT3G05890.1
Contig 6	Eucgr.B03162	Eucgr.B03162.1	AT1G78610.1
Contig 7	Eucgr.I01013	Eucgr.I01013.1	AT2G23610.1
Contig 8	Eucgr.H03473	Eucgr.H03473.1	NA
Contig 9	Eucgr.K02224	Eucgr.K02224.2	AT1G68490.1
Contig 10	Eucgr.H03473	Eucgr.H03473.1	NA
Contig 11	Eucgr.H03473	Eucgr.H03473.1	NA
Contig 12	Eucgr.B00609	Eucgr.B00609.1	AT1G74160.1
Contig 13	Eucgr.H03473	Eucgr.H03473.1	NA
Contig 14	Eucgr.K02224	Eucgr.K02224.1	AT5G16110.1
Contig 15	Eucgr.H03473	Eucgr.H03473.1	NA
Contig 16	Eucgr.B02640	Eucgr.B02640.1	AT1G59960.1
Contig 17	Eucgr.D02142	Eucgr.D02142.1	AT4G17140.3
Contig 18	Eucgr.D02142	Eucgr.D02142.2	AT4G17140.3
Contig 19	Eucgr.D02142	Eucgr.D02142.3	AT4G17140.3
Contig 20	Eucgr.D02142	Eucgr.D02142.2	AT4G17140.3
Contig 21	Eucgr.B01915	Eucgr.B01915.2	AT3G05500.1
Contig 22	Eucgr.K01030	Eucgr.K01030.1	AT4G03560.1

Tabla 3: Clasificación y caracterización de cada lncRNA de acuerdo a FEEInc.

Gen lncRNA putativo	dirección	tipo	distancia	subtipo	ubicación
Contig 1	antisentido	génico	0	anidado	exónico
Contig 2	antisentido	génico	0	superposición	exónico
Contig 3	sentido	génico	0	anidado	intronic
Contig 4	sentido	intergénico	2545	misma hebra	Corriente abajo
Contig 5	antisentido	génico	0	superposición	exónico
Contig 6	antisentido	intergénico	6495	convergente	corriente abajo
Contig 7	antisentido	génico	0	anidado	exónico
Contig 8	sentido	intergénico	29583	misma hebra	corriente arriba
Contig 9	sentido	intergénico	139	misma hebra	corriente abajo
Contig 10	antisentido	intergénico	34824	divergente	corriente arriba
Contig 11	antisentido	génico	0	contenido	intrónico
Contig 12	sentido	intergénico	4619	misma hebra	corriente arriba
Contig 13	sentido	intergénico	29583	misma hebra	corriente arriba
Contig 14	sentido	intergénico	139	misma hebra	corriente abajo
Contig 15	sentido	intergénico	41031	misma hebra	corriente arriba
Contig 16	antisentido	intergénico	3370	convergente	corriente abajo
Contig 17	antisentido	génico	0	anidado	exónico
Contig 18	sentido	intergénico	2593	misma hebra	corriente arriba
Contig 19	antisentido	génico	0	anidado	exónico
Contig 20	sentido	intergénico	2593	misma hebra	corriente arriba
Contig 21	sentido	génico	0	contenido	intrónico
Contig 22	antisentido	génico	0	superposición	exónico

Tabla 4: Resultados del análisis de ontología de genes correspondiente al proceso biológico asociado a cada proteína compañera.

ID	Proceso biológico
Eucgr.K02431	Gravitropismo, organización amiloplasto, gravitropismo negativo
Eucgr.G00843	Proceso biosíntesis suberina, proceso biosíntesis de pectina en pared celular
Eucgr.B00609	Crecimiento celular unidimensional, regulación de crecimiento celular monopolar
Eucgr.L01633	Respuesta a frío
Eucgr.B02645	Respuesta a sustancia inorgánica
Eucgr.B01915	Almacenamiento lipídico, organización de partículas lipídicas, regulación positiva de crecimiento, crecimiento de desarrollo vegetativo, regulación positiva de respuesta a privación de agua
Eucgr.C00095	Morfogénesis de inflorescencia, disparar sistema de desarrollo, transporte transmembrana ión calcio
Eucgr.F01070	Respuesta de defensa, vía de señalización del receptor acoplado a proteína G activadora de proteína quinasa C, respuesta a estímulo de temperaturas, respuesta al agua, respuesta a hongos, respuesta a estrés salino, respuesta a ácido abscísico, regulación de respuesta de defensa, respuesta de defensa a bacteria, respuesta al compuesto nitrógeno, respuesta celular al compuesto que contienen oxígeno
Eucgr.B03162	Transporte aniones, respuesta a frío, respuesta a privación de agua, respuesta a heridas, respuesta a ácido abscísico, crecimiento de desarrollo, transporte transmembrana
Eucgr.I01013	Proceso metabólico de ácido jasmónico, proceso metabólico de ácido salicílico
Eucgr.D02142	Proteína dirigida a vacuola, retención de proteína en aparato de Golgi, división celular
Eucgr.K02224	Regulación de desarrollo post embrionario
Eucgr.K01030	Transporte ión calcio, germinación semilla, regulación de movimiento estomático, señalización mediada por calcio, regulación de transporte ión membrana, regulación del proceso de biosíntesis de ácido jasmónico
Eucgr.K02224	Respuesta a intensidad lumínica, crecimiento del desarrollo

Tabla 5: Resultados del análisis de ontología de genes correspondiente a la función molecular asociada de cada proteína compañera.

ID	Función molecular
Eucgr.G00843	Actividad transferasa, transferir grupos acilo distintos de los grupos aminoacil, actividad hidroxicinamoil transferasa, actividad Omega-hidroxipalmitato O-sinapoil transferasa
Eucgr.B02645	Actividad alditol:NADP+ 1-oxidoreductasa, actividad D-treo-aldosa 1-deshidrogenasa
Eucgr.B01915	Actividad del factor de elongación de traslación, unión de proteínas
Eucgr.C00095	Actividad ATPasa transportadora de calcio, unión de calmodulina, unión de ATP, actividad ATPasa transportadora de cationes, unión de iones metálicos
Eucgr.F01070	Actividad NAD+ kinasa, actividad diacilglicerol kinasa, unión de ATP
Eucgr.B03162	Actividad del canal iónico activado mecánicamente
Eucgr.I01013	Actividad hidrolasa actuando en enlaces ester, actividad de metil indol-3-acetato esterasa, actividad de la esterasa de salicilato de metilo, actividad de metiljasmonato esterasa
Eucgr.K01030	Actividad de los canales de calcio dependientes de voltaje, actividad de los canales de calcio, unión de iones de calcio, unión de proteínas idénticas,

Tabla 6: Dominios proteicos identificados en las proteínas compañeras

ID	Dominio
Eucgr.K02431	KW-0677~Repeat,
Eucgr.L01633	KW-0812~Transmembrane,KW-1133~Transmembrane helix,
Eucgr.C00095	KW-0812~Transmembrane,KW-1133~Transmembrane helix,
Eucgr.B03162	KW-0812~Transmembrane,KW-1133~Transmembrane helix,
Eucgr.K01030	KW-0677~Repeat,KW-0812~Transmembrane,KW-1133~Transmembrane helix,

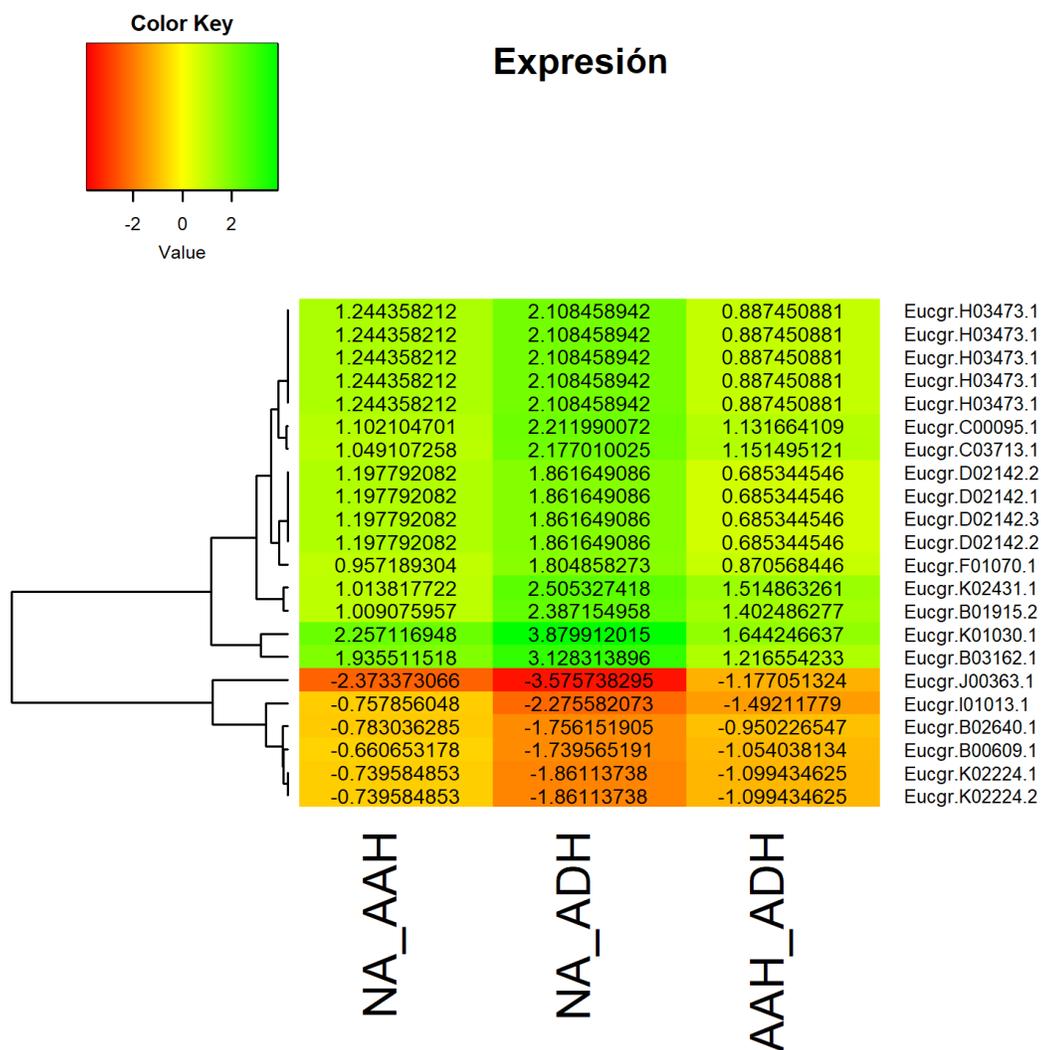


Figura 3: Expresión diferencial de las proteínas compañeras. NA: Condición control. AAH: Aclimatado Antes Helada. ADH: Aclimatado Después de Helada.

IV. DISCUSIÓN

4.1 Filtro por Marco de lectura abierto (ORF).

Se descartaron los contigs con marco de lectura mayor a 100 codones con objetivo de descartar aquellos con ORFs canónicos que pudiesen ser transcritos a ARNm. Este criterio, si bien es arbitrario, tiene como base la clasificación realizada por Couso y Patraquinn (2017) (citado en 2022). Ya que esta clasificación no ha sido verificada en especies vegetales, pero si en otras especies (Tabla 7), se optó por parámetro menos estricto (descartar > 100), teniendo en cuenta que esta clasificación no ha sido verificada en especies vegetales y que el tamaño promedio de lnc ORF es 24.

Tabla 7: Propiedades de ORFs en moscas de la fruta y mamíferos. Las clases fueron asignadas en base a ORFs extraídos de transcriptomas de *Drosophila melanogaster*, *Mus musculus* y *Homo sapiens*.

Clase de ORF	Tamaño promedio (codones)	Tipo ARN	Traducción	Conservación
ORF intergénico	22	ninguno	No	No
uORF	22	5'UTRs de ARNm	Baja	No
lnc ORF	24	lncRNA	Baja	No
CDS corto	79	ARNm corto	Alta	Clase
Isoformas cortas	79	ARNm empalmado	Alta	Reino
ORF canónicos	491	ARNm	Alta	Reino

4.2 Softwares usados para la identificación de lncRNA.

El principal criterio para seleccionar la herramienta computacional para identificar los potenciales lncRNA fue de que el software fuera funcional para una especie no modelo. Se seleccionaron 4 softwares, teniendo en cuenta sus ventajas y desventajas: CNCI, CPC2, FEELnc y lncFinder (Tabla 8). Tanto CNCI como CPC2 usan el modelo de aprendizaje SVM, pero usan como input características de la secuencia diferentes. Con

el objetivo de complementar las características, el set de datos filtrado por ORF fue analizado por separado por cada software, y se continuó con la intersección de ambos resultados. A continuación, se refinaron los resultados integrando la información de CDS, transcritos y lncRNA de *E. grandis*. Las diferencias entre los transcriptomas de *E. grandis* y *E. globulus* x *E. nitens* es el principal inconveniente, seguido por la escasa información de lncRNA en *Eucalyptus* en las bases de datos actuales. Dado que LncFinder usa un conjunto de características más amplio (incluyendo información de la estructura y propiedades fisicoquímicas), e integra diferentes modelos de aprendizaje (LR, SVM, RF, ELM y ANN), se espera que esta combinación óptima de características y de clasificadores, suponga una ventaja ya que a diferencia de CNCI y CPC2, si puede integrar información sobre *Eucalyptus*. Sin embargo, se desconoce la calidad del desempeño de LncFinder en esta especie, y el modelo generado fue personalizado para este trabajo y con la limitada información disponible.

Luego, se usó el script FEELnc_classifier.pl que clasificó los ARN largos no codificantes (lncRNA) en función de su relación con los ARN mensajeros (ARNm) anotados. El script analizó la ubicación y dirección de los lncARN y los ARNm en el genoma para determinar su tipo de interacción. Por ejemplo, si encuentra que un lncARN está transcribiendo en la misma dirección que un ARNm cercano, lo clasificó como una interacción "convergente".

Los fundamentos en los que se basa este proceso de clasificación se derivan de la biología molecular y la genómica estructural. Se basa en la ubicación relativa y la dirección de la transcripción de lncARN y ARNm en el genoma para determinar cómo interactúan. Estas interacciones pueden ser importantes para comprender la función de los lncARN y su relación con la expresión génica (Wucher *et al.* 2017).

El script clasifica los lncRNA en función de su ubicación y orientación con respecto a los genes codificantes de proteínas (ARNm) en el genoma. El objetivo principal es asignar categorías a los lncRNA en función de su relación espacial con los genes codificantes de proteínas. Esta clasificación es esencial para identificar y caracterizar los lncRNA y sus posibles funciones reguladoras.

Los criterios de clasificación son los siguientes (Wucher *et al.* 2017):

- Superposición sentido: Si una transcripción de lncRNA se superpone (comparte coordenadas genómicas) con un ARNm transcrito en la misma cadena (orientación de sentido), puede clasificarse como un "lncRNA de sentido superpuesto".

- Superposición antisentido: Si una transcripción de lncRNA se superpone con un mRNA transcrito en la cadena opuesta (orientación antisentido), puede clasificarse como un "lncRNA antisentido".

- Bidireccional: si una transcripción de lncRNA se encuentra dentro de una determinada distancia genómica (p. ej., 1 kb) y en la orientación opuesta a un sitio de inicio de transcripción de mRNA (TSS) vecino, puede clasificarse como "lncRNA bidireccional". "

- Intergénico: Si una transcripción de lncRNA no se superpone con ninguna transcripción de mRNA o está ubicada lejos de los genes que codifican proteínas, puede clasificarse como un "lncRNA intergénico".

La clasificación de ARN largos no codificantes (lncRNA) en función de su ubicación genómica y orientación en relación con los genes codificadores de proteínas (mRNA) es una estrategia valiosa para identificar y caracterizar los lncRNA y descubrir sus posibles funciones reguladoras. He aquí por qué esta clasificación es útil para los siguientes fines (Wucher *et al.* 2017):

1. Identificación de supuestos lncRNA regulatorios:

- lncRNA superpuestos con sentido: Cuando un lncRNA se superpone con un mRNA en la misma cadena (orientación con sentido), sugiere que el lncRNA puede estar involucrado en la regulación de la expresión de ese mRNA. Puede funcionar como un regulador que actúa en cis, influyendo en la expresión de genes cercanos. Estos lncRNA pueden actuar como potenciadores, promotores o coreguladores del mRNA asociado.

- lncRNA antisentido: Los lncRNA en orientación antisentido a los mRNA pueden potencialmente formar dúplex de ARN-ARN con secuencias complementarias en el mRNA. Esta interacción puede conducir a la degradación del ARNm, la represión traduccional u otros efectos reguladores. Los lncRNA antisentido pueden modular la expresión de sus genes codificadores de proteínas correspondientes.

-lncRNA bidireccionales: Los promotores bidireccionales a menudo impulsan la transcripción tanto de genes codificadores de proteínas como de lncRNA bidireccionales. Estos lncRNA pueden compartir elementos reguladores con sus mRNA vecinos. Su expresión puede estar coordinada y es posible que hayan compartido vías regulatorias.

2. Anotaciones funcionales:

Al clasificar los lncRNA en estas categorías, se puede hipotetizar sobre las funciones potenciales de estas transcripciones. Por ejemplo, los lncRNA con sentido superpuesto podrían estar involucrados en la regulación de genes cercanos relacionados con los mismos procesos biológicos.

3. Perspectivas sobre los mecanismos regulatorios:

Comprender la clasificación de los lncRNA puede proporcionar información sobre los mecanismos reguladores subyacentes. Por ejemplo, los lncRNA antisentido pueden ejercer sus efectos reguladores a través de interacciones de emparejamiento de bases con el mRNA, mientras que los lncRNA sentido superpuestos pueden funcionar a través de la proximidad física al promotor del gen objetivo u otros elementos reguladores.

4. Predicción del objetivo: La clasificación ayuda a predecir posibles genes objetivo o vías reguladas por lncRNA específicos. A futuro se podrían realizar estudios funcionales en estos genes blanco para descubrir funciones reguladoras específicas.

5. Análisis de expresión diferencial: La clasificación de los lncRNA en función de su relación con los mRNA permite la identificación de lncRNA expresados diferencialmente en diversas condiciones. Esta información se puede utilizar para priorizar los lncRNA para estudios funcionales.

6. Validación experimental: Una vez clasificados los lncRNA, se podrían diseñar experimentos para validar sus funciones reguladoras. Por ejemplo, pueden realizar estudios de eliminación o sobreexpresión para investigar el impacto de lncRNA específicos en la expresión de RNAm cercano.

7. Conocimientos biológicos: Esta clasificación proporciona conocimientos sobre el panorama regulatorio del genoma. Ayuda a dilucidar cómo se integran los lncRNA en

las redes reguladoras de genes y cómo contribuyen a la complejidad de la expresión genética.

En resumen, clasificar los lncRNA en función de su ubicación genómica y orientación en relación con los mRNA es un paso crucial para comprender sus funciones reguladoras y ayudar a dilucidar el rol de los lncRNA en la regulación genética y diversos procesos biológicos.

Tabla 8: Tabla resumen de las características principales de cada software usado para la identificación de los lncRNA.

Herramienta (año)	Modelo de aprendizaje	Conjunto de características	Desventajas	Ventajas
CNCI (2013)	SVM	-Largo y S-score del dominio codificante de la secuencia más probable -Porcentaje largo de distancia -Codón-bias	-Baja heterogeneidad de características -Manejo inadecuado de transcripciones de longitud parcial -Baja tolerancia a ensamblajes débiles	-Apta para especies no-modelos -Puede ser usado para estudiar la evolución de las características de las secuencias
CPC2 (2017)	SVM	Secuencia: -Puntuación Fickett -TESTCODE -Largo ORF -Integridad ORF -Punto isoeléctrico	-Baja tolerancia a ensamblajes débiles -Manejo inadecuado de transcripciones de longitud parcial	-Agnóstico de especies
FEElnc (2017)	RF	Secuencia: -Frecuencias k-mer, -Cobertura ORF	-ARNms débilmente expresados podrían ser clasificados erróneamente -Bajo desempeño en el caso de ensamblajes con alta tasa de error	-Agnóstico de especies -Selección de características suele ser suficiente para capturar todas las clasificaciones de lncRNAs
LncFinder (2019)	LR, SVM, RF, ELM y ANN	Secuencia: -Distancia logarítmica del hexámero de un ORF, largo del ORF más largo, cobertura del ORF más largo. Estructura: -Mínimo de energía libre, Frecuencia UP de secuencia emparejada-no emparejada - logaritmo-distancia de la secuencia acguD -logaritmo-distancia de la secuencia acgu-ACGU Físicoquímico: -Señal en posición 1/3 -SNR -Cuantil estadísticos (Q1, Q2, mín y máx)	-No considera interacción, anotación ni características relacionadas con la evolución. -Desempeño varía entre especies	-Integra múltiples clasificadores -Algoritmo es desarrollado basado en la combinación óptima de características y de los clasificadores más apropiados

4.3 LncRNA putativos identificados y sus proteínas blanco en respuesta a estrés por frío. Contig 1 interactúa con la proteína Eucgr.K02431 que está involucrada en el proceso de gravitropismo, mediante la modulación de la dinámica de la membrana vacuolar (en las células sensibles a la gravedad (por ejemplo, las células endodérmicas) durante el proceso de sedimentación del amiloplasto, el cual promueve el crecimiento de la raíz (Hashigushiet *al.* 2014). Esta proteína se encontró regulada positivamente frente a estrés por frío.

Contig 2 interactúa con la proteína Eucgr.C00095, anotada como ATPasa 10 transportadora de calcio en la membrana plasmática. Esta enzima dependiente de magnesio cataliza la hidrólisis de ATP junto con la translocación de calcio desde el citosol hacia el retículo endoplásmico. Esta proteína se encontró regulada positivamente frente a estrés por frío.

Contig 3 interactúa con la proteína Eucgr.F01070.1 anotada por similitud como diacilglicerol quinasa 5, la cual está encargada de fosforilar el segundo mensajero, el diacilglicerol, para generar ácido fosfatídico, otra importante molécula de señalización. Se ha mostrado que esta conversión es crítica para la tolerancia a congelación en *A. thaliana* (Tan *et al.* 2018). Esta proteína se encontró regulada positivamente frente a estrés por frío.

Contig 4 interactúa con la proteína Eucgr.J00363.1 anotada como Omega-hidroxi palmitato O-feruloil transferasa y está involucrado en la síntesis de compuestos aromáticos del polímero suberina (Gouet *al.* 2009, Molina *et al.* 2009). Esta proteína se encontró regulada negativamente frente a estrés por frío.

Contig 5 interactúa con la proteína Eucgr.C03713.1 anotada como proteína hidrofóbica RCI2B, o también conocida como Proteína sensible a la sal y a baja temperatura LTI6B y podrían desempeñar un papel en el mantenimiento de la función y/o integridad de la membrana en situaciones de estrés hídrico provocadas por bajas temperaturas, congelamiento u otras condiciones ambientales que reducen la disponibilidad de agua (Capel *et al.* 1997). Esta proteína se encontró regulada positivamente frente a estrés por frío.

Contig 6 interactúa con la proteína Eucgr.B03162.1, anotada como proteína 6 del canal iónico mecanosensible, la cual se activa directamente a través de cambios en la tensión de la membrana, como la presión de turgencia o la temperatura (actuando como una válvula osmótica para prevenir la ruptura celular bajo un shock hipo-osmótico extremo) (Haswell 2007). Esta proteína se encontró regulada positivamente frente a estrés por frío.

Contig 7 interactúa con la proteína AT2G23610.1, anotada como metilesterasa 3 que tiene actividad carboxilesterasa, actividad esterasa de ácido metil indol-3-acético y actividad esterasa de metiljasmonato, y está involucrada en el metabolismo de lípidos y biosíntesis de hormonas y oxipilina (Yang *et al.* 2008). Esta proteína se encontró regulada negativamente frente a estrés por frío.

Contig 9 interactúa con la proteína Eucgr.K02224.2, anotada como una translocasa subunidad secA. De acuerdo con el análisis por ontología, está involucrado en el proceso de regulación de desarrollo post embrionario. Esta proteína se encontró regulada negativamente frente a estrés por frío.

Contig 12 interactúa con la proteína Eucgr.B00609.1, anotada como proteína longifolia y está involucrada en el proceso de crecimiento celular unidimensional y regulación de crecimiento celular monopolar (Lee *et al.* 2018). Esta proteína se encontró regulada negativamente frente a estrés por frío.

Contig 14 interactúa con la proteína Eucgr.K02224.1, anotada como Proteína no caracterizada T21H19_3. De acuerdo con el análisis por ontología, está involucrada en el proceso de respuesta a intensidad lumínica y crecimiento del desarrollo. Además, se ha mostrado que esta proteína podría desempeñar un papel importante en la señalización de la respuesta al estrés abiótico o en los mecanismos generales de aclimatación (Luhua *et al.* 2013) Esta proteína se encontró regulada negativamente frente a estrés por frío.

Contig 16 interactúa con la proteína Eucgr.B02640.1, anotada como Proteína de la superfamilia de oxidoreductasa ligada a NAD(P). De acuerdo con el análisis por ontología, está involucrada en el proceso de respuesta a materia inorgánica. Esta proteína se encontró regulada negativamente frente a estrés por frío.

Contig 17, Contig 18, Contig 19 y Contig 20 interactúan con la proteína anotada como una proteína que contiene el dominio de homología de Pleckstrin en 3 isoformas diferentes. De acuerdo con el análisis por ontología, es posible que participe en los procesos: Proteína dirigida a vacuola, retención de proteína en aparato de Golgi y división celular. Esta proteína se encontró regulada positivamente frente a estrés por frío.

Contig 21 interactúa con la proteína Eucgr.B01915.2 anotada como proteína tipo REF/SRPP At3g05500 que tiene un rol dual, ya que participa en los procesos biológicos tanto de crecimiento y desarrollo de tejidos, y de respuesta a estrés por sequía (Kim *et al.* 2016). Esta proteína se encontró regulada positivamente frente a estrés por frío y congelación. Se puede inferir que el rol que está cumpliendo esta proteína es el de respuesta a estrés por sequía, debido a la baja disponibilidad del agua durante las bajas temperaturas.

Contig 22 interactúa con la proteína Eucgr.K01030.1 anotada como proteína de canal de calcio de dos poros 1, la cual participa en la regulación de la germinación y el movimiento estomático, actuando como el principal canal de Ca²⁺ sensible a ROS y es el posible objetivo de la inhibición dependiente de aluminio (Al) (Peiter *et al.* 2005).

En la presente investigación se descubrió que los lncRNAs presentes podrían modular positivamente la expresión de proteínas involucradas de involucrada en los procesos de: de gravitropismo, señalización, mantenimiento de la función y/o integridad de la membrana, prevención de la ruptura celular bajo un shock hipo-osmótico extremo, retención de proteínas en aparato de Golgi, respuesta a estrés por sequía y regulación de la germinación y el movimiento estomático; además de que podrían modular negativamente la expresión de proteínas involucradas en los procesos de síntesis de compuestos aromáticos del polímero suberina, metabolismo de lípidos y biosíntesis de hormonas y oxipilina, regulación de desarrollo post embrionario, crecimiento celular unidimensional y regulación de crecimiento celular monopolar, proceso de respuesta a intensidad lumínica y crecimiento de desarrollo, señalización de la respuesta al estrés abiótico y en los mecanismos generales de aclimatación y respuesta a materia inorgánica.

V. CONCLUSIONES

En conclusión, el estudio se propuso como objetivo general revelar la asociación entre los ARNs no codificantes de cadena larga y la respuesta al estrés por frío en librerías de expresión en el híbrido de *Eucalyptus nitens* x *Eucalyptus globulus*, considerando el estado de aclimatación al frío de las plantas. Para lograr este propósito, se establecieron objetivos específicos que incluyeron la discriminación de transcritos codificantes para lncRNA y sus respectivos genes blancos, la evaluación *in silico* de la expresión de los genes blanco putativos de los lncRNA en condiciones no aclimatadas y aclimatadas, y la asociación de la abundancia relativa de lncRNA con los genes blanco diferencialmente expresados en estas mismas condiciones. Este enfoque integrado proporciona una comprensión más profunda de los mecanismos moleculares subyacentes a la aclimatación al frío en este híbrido de *Eucalyptus*, abriendo nuevas perspectivas para la investigación en la mejora genética y la adaptación de estas especies a ambientes adversos.

Los resultados de esta investigación sugieren una posible regulación de la expresión génica mediada por lncRNA en respuesta a estrés por baja temperatura en hojas de un híbrido de eucalipto. Se encontró una posible relación entre las “proteínas compañeras” o “proteínas blanco” diferencialmente expresados en condiciones no aclimatadas y aclimatadas y la función biológica de estas.

Las posibles direcciones futuras de investigación van desde la validación experimental de los lncRNA identificados, estudios funcionales para comprender su rol exacto en la respuesta al frío y la aplicación de estos conocimientos en programas de mejora genética para incrementar la tolerancia al frío en especies de *Eucalyptus*.

VI. BIBLIOGRAFÍA

Anders, S., Reyes, A. y Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Research*, 22, 2008-17.

Apablaza, H., Solís, M., Conejera, D., Fonseca, A., Cid, J., Tarifeño-Saldivia, E., Valenzuela, S., Emhart, V. y Fernández, M. (2021). bHLH Transcription Factors Undergo Alternative Splicing During Cold Acclimation in a Eucalyptus hybrid. *Plant Molecular Biology Reporter*, 40, 310-326.

Bardou, F., Ariel, F., Simpson, C.; Romero-Barrios, N., Laporte, P., Balzargue, S., Brown, J., Crespi, M. (2014). Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. *Developmental Cell*, 30, 166-176.

Capel, J., Jarillo, J., Salinas, J., Martinez-Zapater, J. (1997). Two homologous low-temperature-inducible genes from Arabidopsis encode highly hydrophobic proteins. *Plant Physiology*, 115, 569-576.

Chen, J., Quan, M., Zhang, D. (2015) Genome-wide identification of novel long non-coding RNAs in Populus tomentosa wood, opposite wood and normal wood xylem by RNA-seq. *Planta*, 241, 125-43.

Chen, M., Penfield, S. (2018). Feedback regulation of COOLAIR expression controls seed dormancy and flowering time. *Science*, 360, 1014-1017.

Chinnusamy, V., Zhu, J., Zhu, J. (2007). Cold stress regulation of gene expression in plants. *Trends in Plant Science*, 12, 444-451.

Conejera, D. (2022). Análisis transcriptómico y comparativo de genes de síntesis de cutícula en aclimatación a estrés por bajas temperaturas en Eucalyptus spp [Tesis de magíster no publicada]. Universidad de Concepción

Costa e Silva, F., Shvaleva, A., Broetto, F., Ortuno, M. F., Rodrigues, M. L., Almeida, M., Pereira, J. (2008). Acclimation to short-term low temperatures in two Eucalyptus globulus clones with contrasting drought resistance. *Tree Physiology*, 29, 77–86.

Csorba, T., Quest, J., Sun, Q., Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proceedings of the National Academy of Sciences*, 111, 16160-5.

Di, C., Yuan, J., Wu, Y., Li, J., Lin, H., Hu, L., Zhang, T., Qi, Y., Gerstein, M., Guo, Y., Lu, Z. (2014). Characterization of stress-responsive lncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. *Plant Journal*, 80, 848-861.

Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., Xu, C., Li, X., Xiao, J., Zhang, Q. (2012). A long noncoding RNA regulates photoperiod-sensitive male sterility, and essential component of hybrid rice. *Proceedings of the National Academy of Sciences*, 109, 2654-2659.

Engreitz, J., Ollikainen, N., Guttman, M. (2016). Long non-coding RNAs: Spatial amplifiers that control nuclear structure and gene expression. *Nature Reviews Molecular Cell Biology*, 17, 756-770.

Engreitz, J., Pandya-Jones, A.; McDonel, P., Shishkin, A., Sirokman, K., Surka, C.; Kadri, S., Xing, J., Goren, A., Lander, E. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*, 55, 791-802.

Gaete-Loyola, J., Lagos, C., Beltrán, M., Valenzuela, S., Emhart, V., Fernández, M. (2017). Transcriptome profiling of Eucalyptus nitens reveals deeper insight into the molecular mechanism of cold acclimation and deacclimation process. *Tree Genetics & Genomes*, 13, 37.

- Gou, J., Yu, X., Liu, C. (2009). A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 18855-18860.
- Grabherr, M., Haas, B., Yassour, M., Levin, J., Thompson, D., 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., Nusbaum, C., Lindblad-Toh, K., Friedman, N. y Regev, A. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Natural Biotechnology*, 29, 644-652.
- Gysling, A., Álvarez, V., Soto, D., Pardo, E. y Poblete, P. (2018). Anuario Forestal 2018. *Boletín Estadístico*, 163, 27-28.
- Hansen, T., Jensen, T. Clausen, B., Bramsen, J. Finsen, B., Damgaard, C. y Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature*, 495, 384.
- Hashiguchi, Y., Yano, D., Nagafusa, K., Kato, T., Saito, C., Uemura, T., Ueda, T., Nakano, A., Tasaka, M., Terao Morita, M. (2014). A unique HEAT repeat-containing protein SHOOT GRAVITROPISM6 is involved in vacuolar membrane dynamics in gravity-sensing cells of Arabidopsis inflorescence stem. *Plant and Cell Physiology*, 55, 811-822.
- Hashiguchi, Y., Yano, D., Nagafusa, K., Kato, T., Saito, C., Uemura, T., Ueda, T., Nakano, A., Tasaka, M., Terao, M. (2014). A unique HEAT repeat-containing protein SHOOT GRAVITROPISM6 is involved in vacuolar membrane dynamics in gravity-sensing cells of Arabidopsis inflorescence stem. *Plant Cell Physiology*, 55, 811-822.
- Haswell, E. (2007). MscS-like proteins in plants. *Current Topics in Membranes*, 58, 329-359.
- Heo, J. B., and Sung S. (2011), Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*, 331, 76-79.

- Ipinza, R., Barros, S., Gutierrez, B. y Brand, M. (2014). *Mejoramiento genético de Eucaliptos historia, avances y tendencias* (1a. ed.). INFOR-FIA.
- Kang, Y., Yang, D., Kong, L., Hou, M., Meng, Y., Wei, M. y Gao, G. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Research*, 45, 12-16.
- Kawano, T., Kadono, T., Fumoto, K., Lapeyrie, F., Kuse, M., Isobe, M., Furuichi, T., Muto, S. (2004). Aluminum as a specific inhibitor of plant TPC1 Ca(2+) channels. *Biochemical and Biophysical Research Communications*, 324, 40-45.
- Kazemi, S., Maali, R. (2018). Global insights of protein responses to cold stress in plants: Signaling, defence, and degradation. *Journal of Plant Physiology*, 226, 123–135.
- Kim, D., Sung, S. (2017). Vernalization-triggered intragenic chromatin loop formation by long noncoding RNAs. *Developmental Cell*, 40, 302-12.e4.
- Kim, D., Sung, S. (2013). Coordination of the vernalization response through a VIN3 and FLC gene family regulatory network in Arabidopsis. *Plant Cell*, 25, 454-469.
- Kim, E., Park, K., Seo, Y. y Kim, W. (2016). Arabidopsis Small Rubber Particle Protein Homolog SRPs Play Dual Roles as Positive Factors for Tissue Growth and Development and in Drought Stress Responses. *Plant Physiology*, 170, 2494-2510.
- Lee, Y., Rhee, J., Lee, S., Chung, G., Park, S., Segami, S., Maeshima, M. y Choi, G. (2018). Functionally redundant LNG3 and LNG4 genes regulate turgor-driven polar cell elongation through activation of XTH17 and XTH24. *Plant Molecular Biology*, 97, 23-36.
- Li, L., Ma, W., Wang, J., Geng, B., Yang, J. y Cui, Q. (2015). Genome wide discovery and characterization of maize long non-coding RNAs. *Genome Biology*, 15, 40.
- Li, W. y Godzik A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22, 1658-9.

- Lin, Z., Long, J., Yin, Q., Wang, Q., Li, H., Luo, J., Wang, H. C. y Wu, A. (2019). Identification of novel lncRNAs in *Eucalyptus grandis*. *Industrial Crops & Products*, 129, 309-317.
- Long, Y., Wang, X., Youmans, D. y Cech, T. (2017). How do lncRNAs regulate transcription?. *Science Advance*, 3, eaao2110.
- Luhua, S., Hegie, A., Suzuki, N., Shulaev, E., Luo, X., Cenariu, D., Ma, V., Kao, S., Lim, J., Gunay, M., Oosumi, T., Lee, S., Harper, J., Cushman, J., Gollery, M., Girke, T., Bailey-Serres, J., Stevenson, R., Zhu, J. y Mittler, R. (2013). Linking genes of unknown function with abiotic stress responses by high-throughput phenotype screening. *Physiology Plantarum*, 148, 322-333.
- Marquart, S., Raitskin, O., Wu, Z., Liu, J., Sun, Q. y Dean, C. (2014). Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription. *Molecular Cell*, 54, 156-165.
- McCarthy, D., Chen, Y. y Smyth, K. (2012). Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. *Nucleic Acids Research*, 40, 10.
- Molina, I., Li-Beisson, Y., Beisson, F., Ohlrogge, J. y Pollard, M. (2009). Identification of an *Arabidopsis* feruloyl-coenzyme A transferase required for suberin synthesis. *Plant Physiology*, 151, 1317-1328.
- Peiter, E., Maathuis, F., Mills, L., Knight, H., Pelloux, J., Hetherington, A. y Sanders, D. (2005). The vacuolar Ca²⁺-activated channel TPC1 regulates germination and stomatal movement. *Nature*, 434, 404-408.
- Qin, T., Zhao, H. Cui, P., Albeshar, N., y Xiong, L. (2017). A nucleus-localized long non-coding RNA enhances drought and salt stress tolerance. *Plant Physiology*, 175, 1321-1336.

Ramakrishnaiah, Y., Kuhlmann, L. y Tyahi, S. (2020). Towards a Comprehensive Pipeline to Identify and Functionally Annotate Long Noncoding RNA (lncRNA). *Computers in Biology and Medicine*, 127, 104028.

Rigo, R., Bazin, J., Romero-Barrios, N., Maison, M., Lucero, L., Christ, A., Benhamed, M., Blein, T. Huguet, S. y Charon, C. (2020). The Arabidopsis lncRNA ASCO modulates the transcriptome through interaction with splicing factors. *EMBO reports*, 21, e48977.

Shichino, Y., Tamashita, A. y Yamamoto, M. (2014). Meiotic long non-coding meiRNA accumulate as a dot at its genetic locus facilitated by Mmi1 and plays as a decoy to lure Mmmi1. *Open Biology*, 4, 140022.

Shriram, V., Kumar, V., Devarumath, R. M., Khare, T. S., y Wani, S. H. (2016). MicroRNAs As Potential Targets for Abiotic Stress Tolerance in Plants. *Frontiers in Plant Science*, 817, 1-18.

Song, Y. Xuan, A., Bu, C., Ci, D., Tian, M., Zhang, D. (2019). Osmotic stress-responsive promoter upstream transcripts (PROMPTs) act as carriers of MYB transcription factors to induce the expression of target genes in *Populus simonii*. *Plant Biotechnology*, 17, 164-177.

Suksamran, R., Saithong, T., Thammamongtham, C. y Kalapanulak, S. (2020). Genomic and transcriptomic analysis identified novel putative cassava lncRNAs involved in cold and drought stress. *Genes*, 11, 366.

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., Liu, Y., Chen, R. y Zhao, Y. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research*, 41, 1-18.

Tan, W., Yang, Y., Zhou, Y., Huang, L., Xu, L., Chen, Q., Yu, L. y Xiao, S. (2018). DIACYLGLYCEROL ACYLTRANSFERASE and DIACYLGLYCEROL KINASE

modulate triacylglycerol and phosphatidic acid production in the plant response to freezing stress. *Plant Physiology*, 177, 1303-1318.

Waititu, J, Zhang, C., Liu, J. y Wang, H. (2020). Plant Non-Coding RNAs: Origin, Biogenesis, Mode of Action and Their Roles in Abiotic Stress. *International Journal of Molecular Sciences*, 21, 8401.

Wang, H. Niu, Q., Wu, H., Liu, J., Ye, J., Yu, N. y Chua, N. (2015). Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. *Plant Journal*, 84, 404-416.

Wang, H., Chung, P., Liu, J., Jang, I., Kean, M., Xu, J. y Chua, N. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. *Genome Research*, 24, 444-453.

Wang, M., Wu, H., Fang, J., Chu, C. y Wang, X. (2017). A long nocoding RNA involved in rice reproductive development by negatively regulating osa- miR160. *Science Bulletin*, 62(7), 470-475.

Wang, W., Qin, Q., Sun, F., Wang, Y., Xe, D. y Li, Z. (2016). Genome-wide differences in DNA methylation changes in two contrasting rice genotypes in response to drought conditions. *Frontiers in Plant Science*, 7, 1675.

Warren, G. J. (1998). Cold stress: manipulating freezing tolerance in plants. *Current Biology*, 8(15), 514-516.

West, J., Davis, C., Sunwoo,H., Simon, M., Sadreyev, R., Wang, P., Tolstorukov, M., Kingston, R. (2014). The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Molecular Cell*, 55, 791-802.

Yang, Y., Xu, R., Ma, C., Vlot, A., Klessig D. y Pichersky, E. (2008). Inactive methyl indole-3-acetic acid ester can be hydrolyzed and activated by several esterases belonging to the AtMES esterase family of Arabidopsis. *Plant Physiology*, 147, 1034-1045.

Zhang, L., Lu, H., Xin, D., Cheng, H. y Zhou, R. (2010). A novel ncRNA gene from mouse chromosome 5 trans-splices with Dmrt1 on chromosome 19. *Biochemical and Biophysical Research Communications*, 400(4), 696-700.

Zhao, L., Wang, J., Li, Y.; Song, T., Wu, Y., Fang, S., Bu, D., Li, H., Sun, L. y Pei, D. (2021). NONCODEV6: An updated database dedicated to long non-coding RNA annotation in both animals and plants. *Nucleic Acids Research*, 49, 165-171.