

# Ensamblado del transcriptoma de novo y análisis de expresión diferencial de genes en respuesta al estrés hídrico en *Nothofagus alpina*

Rita M. Lopez Laphitz<sup>1</sup>, Verónica Arana<sup>2</sup>, Santiago Varela<sup>2</sup>, Leandro Becker<sup>3</sup>, Carolina Soliani<sup>2</sup>, María M. Azpilicueta<sup>2</sup>, Paula Marchelli<sup>2</sup>, Nicolás Bellora<sup>4</sup>

**Palabras Clave:** Cambio climático, Patagonia, bosques.

## Introducción

Los bosques cumplen un rol de importancia en la regulación de las temperaturas y el ciclo del agua a través de la evapotranspiración (ET) el cual está siendo afectado por escenarios de recurrentes sequías (Diao et al. 2020). Debido a los cambios de escenarios climáticos (e.g. sequías, incremento de temperatura), las poblaciones han estado bajo procesos de adaptación, migraciones a otros ambientes, y extinciones locales. La variabilidad de los rasgos fenotípicos en poblaciones de árboles, frecuentemente se correlaciona con factores climáticos y geográficos, sugiriendo que la variación genética y/o genómica podría seguir este patrón y evidenciar factores selectivos (McKown et al. 2013). Por lo tanto, el estudio de las bases genéticas de la variación en rasgos fenotípicos, además de cómo estos rasgos se distribuyen a través de la especie, es indispensable en predicciones futuras de respuestas a distintos escenarios de cambio climático (Bellard et al. 2012; Razgour et al. 2019). El bosque templado subantártico ocupa una gran extensión desde los 34° 50' S hasta el sur del archipiélago Tierra del Fuego (55° S), y constituye un spot de gran biodiversidad donde el género *Nothofagus* es dominante. El presente estudio reporta el primer transcriptoma de *N. alpina* en condiciones de estrés hídrico, siendo el objetivo último caracterizar el perfil transcriptómico en respuesta a la sequía. Basado en este análisis se identificarán genes claves en vías metabólicas en respuesta a este estrés abiótico. Estudios recientes en algunas de las especies del género han aportado información genómica sobre las respuestas de procesos moleculares, que sirven de base para evidenciar respuestas adaptativas y así poder incluirlas en el manejo y la conservación de estos bosques (Torales et al. 2012; El Mujtar et al. 2014, Es-

travis-Barcala et al. 2021).

## Materiales y Métodos

**Ensayos de estrés hídrico:** Se colectaron plantas juveniles de *N. alpina* ("raulí") provenientes de una misma población de la cuenca Lácar (Parque Nacional Lanín, Neuquén) de la zona Quila Quina. Las plantas se cultivaron en macetas de 3 litros realizándose dos tratamientos: inducción de estrés hídrico mediante la suspensión de riego, y control con condiciones de riego periódico. El estado fisiológico de las plantas se monitoreó a través de mediciones periódicas del contenido de agua en maceta, del potencial hídrico de la planta en pre-alba, y de la conductancia estomática. El ensayo se repitió durante dos estaciones de crecimiento, 2014-2015 y 2015-2016, para obtener réplicas. Se colectaron hojas de estos ensayos para realizar una secuenciación completa del transcriptoma, tanto de la situación control (T0;  $\Psi$  (pd) < -0,8 MPa) como de las plantas sometidas a estrés hídrico (estrés máximo ( $\Psi$  (pd) < -1,5 MPa; T2), totalizando ocho muestras.

**Secuenciación, ensamblado y anotación de transcriptomas:** El material proveniente de los ensayos de estrés hídrico se utilizó para analizar la expresión diferencial de genes. Sin genomas de referencia de la especie estudiada, se realizaron ensamblados de novo de los transcriptomas completos en estudio. Se realizó control de calidad y limpieza de las lecturas con FastQC, Rcorrector y Trimmomatic, y se obtuvo un ensamblado de referencia con el programa Trinity. Asimismo, se realizaron ensamblados para las condiciones por separado (raulí control y raulí estrés), con el objetivo de identificar y estudiar potenciales isoformas únicas de cada condición biológica. A continuación se anotaron los transcriptos (secuencias ensambladas) con la información disponible en el proteoma de *Arabidopsis thaliana*, en la secuenciación parcial

1 IDEAUS (Instituto de Diversidad y Evolución Austral-CONICET). Contacto: rlaphitz@cenpat-conicet.gob.ar.

2 EEA-Bariloche- IFAB (Instituto de Investigaciones Forestales y Agropecuarias Bche.; INTA-CONICET).

3 IDEAUS (Instituto de Diversidad y Evolución Austral-CONICET).

4 INTECNUS (Institute of Nuclear Technologies for Health-CONICET).

del transcriptoma de raulí (Torales et al. 2012) y en otras bases de datos nucleotídicas y proteicas, para caracterizar familias génicas y vías metabólicas.

**Análisis de expresión diferencial y de enriquecimiento funcional:** Los genes diferencialmente expresados (DEGs) se identificaron con los programas Salmon y DESeq2. El análisis de enriquecimiento funcional incluyó análisis de sobrerrepresentación (overrepresentation analysis, ORA) y de enriquecimiento de conjunto de genes (gene set enrichment analysis, GSEA) realizados con el programa de R clusterProfiler. Se identificaron ontologías génicas (Gene Ontology terms, GOs): Proceso Biológico (Biological Process, BP), Función Molecular (Molecular Function, MF) y Componente Celular (Cellular Component, CC), así como vías KEGG (Kyoto Encyclopedia of Genes and Genomes). Se utilizaron los programas Cytoscape, STRING y cytohubba para la visualización de redes de interacciones proteína-proteína.

## Resultados

**Ensayo de estrés hídrico:** Las respuestas fisiológicas específicas para cada tratamiento (i.e. control y estrés) se graficaron para cada una de las variables monitoreadas (contenido de agua, potencial agua en pre-alba y conductancia estomática de la hoja). La figura 1 muestra dos períodos de crecimiento (réplicas biológicas), observando que los patrones de potencial agua de la hoja son similares para ambos ensayos. Las plantas bajo estrés severo en T2 muestran una disminución abrupta del potencial agua que coincidió con una disminución del contenido de agua de la maceta y de la conductancia estomática de la hoja (gráficos no mostrados).

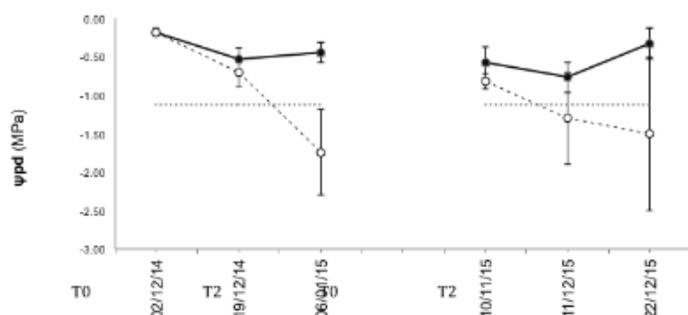


Figura 1: Potencial agua en pre-alba ( $\Psi_{pa}$ ). A la izquierda se grafica el primer ensayo 2014-2015, y a la derecha el segundo ensayo 2015-2016 (réplicas biológicas). Líneas continuas: control; líneas segmentadas: estrés.

## Secuenciación, ensamblado y anotación de transcriptomas

Se secuenciaron dos conjuntos de hojas de raulí, las cuales provenían de dos estados hídricos contrastantes (control vs. estrés) como se describió en la sección anterior. Se obtuvieron un total de 34.153.157 lecturas, las cuales se ensamblaron en 179.620 transcritos. Se eliminó la redundancia obteniéndose un total de 105.052 transcritos. La anotación funcional contra los proteomas de *A. thaliana*, *Populus trichocarpa*, *Juglans regia*, y *refseq\_rna* resultó en 57.497 transcritos anotados representando el 54.73%.

**Análisis de expresión diferencial.** Se encontraron un total de 2.879 genes expresados diferencialmente (DEGs) entre los dos tratamientos (i.e. control vs. estrés). La Figura 2 es una representación gráfica de los genes que poseen expresión diferencial significativa, segregándose en activados o sobre-expresados ( $n=1.113$ ) y suprimidos o sub-expresados ( $n=1.766$ ).

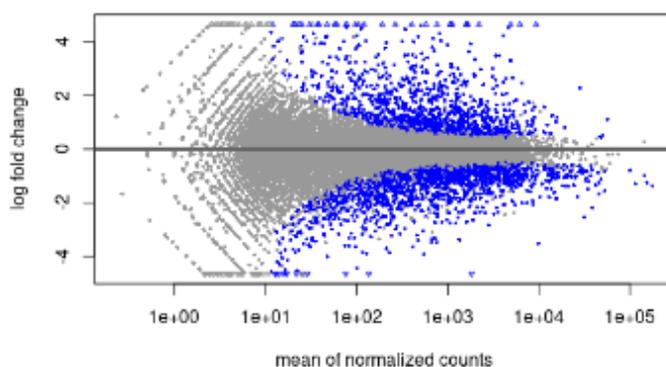


Figura 2: MA-plot indicando en eje Y en escala logarítmica el cambio ("log fold change") en la expresión de cada gen en tratamiento versus control. El eje X indica el conteo promedio de lecturas normalizado ("mean of normalized counts") para cada gen. Los puntos azules indican los genes expresados diferencialmente.

**Análisis de enriquecimiento.** Los 20 términos de Procesos Biológicos más significativos obtenidos con el análisis de enriquecimiento de conjunto de genes se representan en la Fig. 3. Entre ellos, figuran activados o sobre-expresados varios en la categoría "procesos catabólicos", así como "respuesta a ácido abscísico" y "respuesta a hipoxia". Como suprimidos o sub-expresados aparecen varios en las categorías "procesos metabólicos", de "biogénesis" y "biosíntesis", algunos de estos últimos relacionados a sustancias de la vía metabólica de los lípidos, así como algunos relacionados a la fotosíntesis, entre los cuales están "cadena de trans-

porte de electrones” y “fotosíntesis, reacción a la luz”.

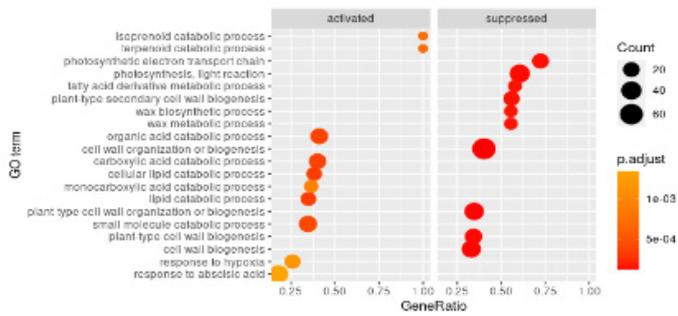


Figura 3: Gráfico de puntos (dotplot) de genes sobre-expresados (“activated”) y sub-expresados (“suppressed”) con los 20 términos de ontología de genes (“GO term”) más significativos en su relación al total de genes (“GeneRatio”) en el transcriptoma de la hoja de *N. alpina* en estado de estrés hídrico.

**Redes de interacción entre proteínas:** Los nodos de la red definen a los genes sobrerrepresentados en el análisis de enriquecimiento y aquellos coloreados son éstos de mayor tráfico o repetidos en varias vías metabólicas. En la figura 4 el cluster “a” incluye genes “hub” tales como ABI1 y LTI65 que fueron clasificados dentro del metabolismo de regulación de la vía metabólica del ácido abscísico, y por lo tanto especifica aquellos genes para el término GO “respuesta a ácido abscísico”. A modo de ejemplo, se destaca en la Figura 4 un grupo de genes entre corchetes relacionados a proteínas “heat-shock”, anteriormente observadas en el transcriptoma de *N. pumilio* como chaperonas a respuestas abióticas tal como las altas temperaturas (Estravis-Barcalá et al. 2021).

### Discusión y conclusiones

Este estudio es pionero en el ensamblado de un transcriptoma de novo de raulí (*N. alpina*) en condiciones de estrés hídrico. El 54,73% de los transcritos ensamblados fue exitosamente anotado. Estos últimos resultados fueron identificados en vías metabólicas relacionadas a la vía del ácido abscísico, producción de metabolitos secundarios (como las sustancias reactivas al oxígeno; ROS), metabolismo de lípidos, y además se identificaron factores de transcripción (TFs). Este estudio de secuenciación de próxima generación contribuye a la identificación de procesos biológicos y funciones moleculares, tanto como vías metabólicas que responden al estado de un estrés hídrico severo. La información generada permitirá continuar trabajando sobre regiones específicas del genoma relacionadas al estrés en

respuestas a factores abióticos, así como proyectar estudios de poblaciones naturales, buscando genes candidatos.

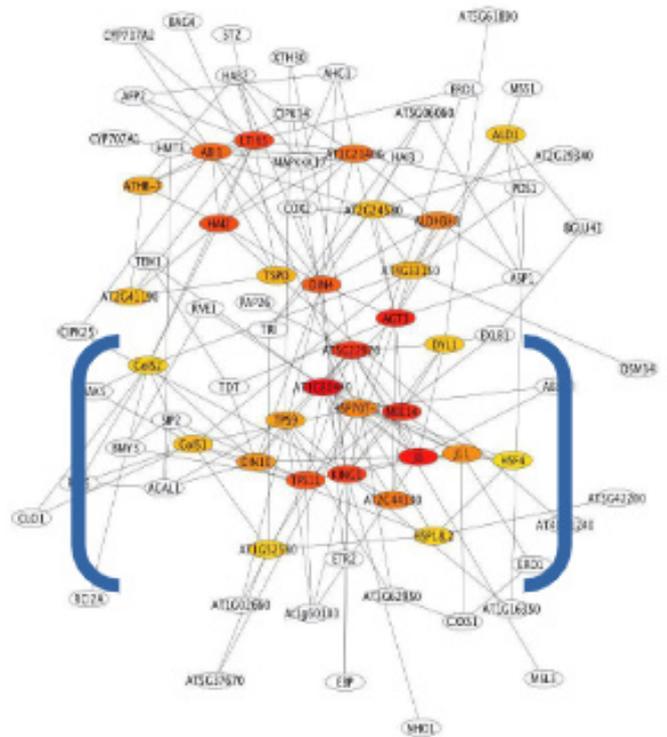


Figura 4: Red de interacciones proteína-proteína (Protein-Protein Interaction, PPI network) de genes sobre-expresados con valores de  $\log_2FC > 1,5$ . Con color anaranjado/rojizo se indican los “hub genes” (genes con alto tráfico en distintas vías metabólicas) obtenidos con el programa “cytohubba”.

### Bibliografía Citada

- Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F. 2012. Impacts of climate change on the future of biodiversity. *Ecology Letters* 15: 365–377.
- Diao H, Wang A, Yang H, Yuan F, Guan D, Wu J. 2021. Responses of evapotranspiration to droughts across global forests: a systematic assessment. *Can J Forest Res* 51: 1-9.
- El Mujtar VA, Gallo L, Lang T, Garnier-Géré P. 2014. Development of genomic resources for *Nothofagus* species using next-generation sequencing data. *Mol. Ecol. Res.* 14: 1281–1295.
- Estravis-Barcala M, Heer K, Marchelli P, Ziegenhagen B, Arana MV, Bellora N. 2021. Deciphering the transcriptomic regulation of heat stress responses in *Nothofagus pumilio*. *PLoS One* 16: e0246615.
- McKown AD, Guy RD, Klápště J, Gerald A, Friedman M, Cronk QCB, El-Kassaby YA, Mansfield SD,

Douglas CJ. 2014. Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*. *New Phytol* 201: 1263-1276.

Razgour O, Forester B, Taggart JB, Bekaert M, Juste J, Ibanez C, et al. 2019. Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *PNAS* 116:10418–10423.

Torales S, Rivarola M, Pomponio MF, Fernández P, Acuña C, Marchelli P, Gonzalez S, Azpilicueta MM, Hopp HE, Gallo L, Paniago N, Marcucci Poltri S. 2012. Transcriptome survey of Patagonian southern beech *Nothofagus nervosa* (= *N. alpina*): assembly, annotation and molecular marker discovery. *BMC Genomics* 13:291-303.