Universidad Michoacana de San Nicolás de Hidalgo
Facultad de Agrobiología “Presidente Juárez”

Tesis

Identificación de adaptación en *Pinus pinceana*

Que para obtener el grado académico de
Maestra en Ciencias Biológicas
PRESENTA:

Biól. Laura Alicia Figueroa Corona

Directora de tesis: Dra. Patricia Delgado Valerio

Co Director: Dr. Daniel Piñero Dalmau

Morelia, Mich Febrero, 2016

Índice

Índice de tablas y figuras

Resumen

Este trabajo comprende la descripción, procesamiento y análisis del transcriptoma de *Pinus pinceana,* una conífera endémica de México. Es un esfuerzo para identificar los genes que se expresan y vincular la variación genética de dos distintas regiones geográficas con caracteres morfológicos y cambios climáticos que concurren en la distribución de la especie, involucra el uso de herramientas bioinformáticas como el ensamble *de novo* y los mapeos de referencia, así como, la caracterización de genes y el análisis de expresión genética.

Se detectó la expresión de 35,916 genes de distintos tejidos y estadios de crecimiento que conforman una estimación global y general de las respuestas del genotipo ante procesos de crecimiento y respuesta a patógenos.

La distribución de *P. pinceana* es restringida, las características climáticas generan un gradiente ambiental que aumenta las condiciones áridas en la proporción Norte, la precipitación y la aridez son los factores que divergen de manera más importante entre las dos regiones.

Existen variaciones en la morfología foliar en las cubiertas cerosas y en la organización de los estomas en las caras de la hoja a lo largo del gradiente ambiental. Siendo la cubierta cerosa más abundante en las acículas de la región Sur, mientras en la región Norte en la cara abaxial carece frecuentemente de estomas.

Se analizaron las diferencias genéticas entre las dos zonas, a nivel de expresión donde se pudieron establecer diferencias en la expresión de 421 transcritos donde destacan cambios en proteínas de respuesta a factores climáticos, dehidrinas y proteínas de choque térmico. Y a nivel estructural, con la detección de polimorfismos en donde se encontraron cambios en genes asociados a la tasa de fotorespiración y al inicio del proceso de germinación.

Las diferencias en algunos caracteres morfológicos, genéticos y la variación climática a lo largo de la distribución, pueden ser planteados como marcadores candidatos potenciales en futuros estudios de asociación y/o ensayos de expresión. Reconocer y corroborar estas asociaciones genotípicas y fenotípicas ayudarán a identificar procesos evolutivos, de adaptación, estrés y resistencia de las coníferas en condiciones áridas y semi-áridas.

Los resultados obtenidos son una contribución a la caracterización del genoma de las coníferas, y gimnospermas, destacando la importancia de las nuevas tecnologías de secuenciación y de herramientas bioinformáticas para la caracterización transcriptomas y mayor conocimiento de genomas grandes y complejos.

Palabras clave: *Pinus pinceana*, transcriptoma, RNAseq, marcadores candidatos, adaptación.

Abstract

The aim of this study was to identify the expressed genes on *Pinus pinceana*, a Mexican endemic conifer and link the genetic variation found in two distinct geographical regions with morphological and climatic variables across its distribution. We made the description, processing and analysis of the transcriptome making use of bioinformatics tools like assemblages de novo and mapping references to characterize genes and make a gene expression analysis.

We identified 35,916 genes in different tissues obtained from different plant tissues through several growth stages which allowed us to obtained a estimate genotype responses to processes especific metabolic proceses like the response to pathogens.

Pinus pinceana has a restricted distribution due to the soil and the prevailing climatic conditions, which create a gradient increasing aridity gradient in the northern distribution.

We found variations in the leaf morphology, in the wax cover and the organization of stomata along the environmental gradient. Leaves (needles) from the southern distribution showed a more abundant wax coating, while leaves from the northern distribution lack stomata in the abaxial face.

We detected differences at the expression level 421 transcripts, which highlight changes in protein response to climatic factors, dehydrins and heat-shock proteins. At the structural level, with the detection of SNPs we found changes in genes associated with photorespiration rate and with the start of the germination process.

Differences found in morphological and in the genetic characteristics according to the climatic variation along the distribution can be used as potential markers for future studies of association and / or expression assays.

The recognition of these genotypic and phenotypic associations may help to identify evolutionary processes, adaptation and stress resistance of conifers under arid and semi-arid conditions. The results are a contribution to the characterization of the genome of conifers and gymnosperms, highlighting the importance of new sequencing technologies and bioinformatics tools for transcriptome characterization for a better understanding of large and complex genomes.

Introducción

El análisis del genoma se ha convertido en una herramienta útil para la comprensión de los cambios evolutivos. En la última década la innovación de tecnologías, metodologías eficientes y de bajo costo han permitido una explosión y diversificación de esta disciplina para caracterizar de los componentes genéticos de una célula, tejido u organismo.

Los estudios genómicos abordan dos perspectivas generales, i) la estructural, donde es utilizada para entender la evolución de los genes y genomas, el proceso básico de cambio en la secuencia de DNA, acceder a los polimorfismos que expliquen cambios morfológicos o fisiológicos, y determinar eventos filogenéticos (Labrou *et al.*, 2015; Paudel *et al.*, 2015; Escaramís *et al.*, 2015), y ii) la funcional, se utiliza para entender procesos evolutivos en familias de genes específicos, presiones de selección, susceptibilidad a enfermedades y patrones de expresión genética (Oleksiak *et al.*, 2002; Karhu *et al.,* 1996).

La caracterización de transcriptomas representa un esfuerzo descriptivo de la variación genómica funcional en los organismos, utilizada para el desarrollo de nuevos marcadores moleculares, construir microarreglos o perfiles de expresión génica y explicar patrones que permitan entender la variación fenotípica, composiciones y procesos ecológicos y evolutivos, como la adaptación y la especiación (Feder Mitchell-Olds, 2003; Zhou *et al*., 2004; Cañas *et al.,* 2015).

La dinámica de cambio, que explica cómo surgen mecanismos adaptativos y cómo es que permanecen en la población y/o que producen su aislamiento, ha sido abordada por la genética de poblaciones, la genética cuantitativa, la ecología de los genes (Turessw,1923) y recientemente por la genómica. A pesar de los avances, la identificación de las mutaciones, genes y vías bioquímicas implicadas, el reconocimiento de los rasgos fenotípicos responsables de la adaptación, sigue siendo laborioso y difícil, puesto que la detección de huellas de selección depende de la naturaleza y la fuerza de los eventos de selección, la escala evolutiva a la que sucedieron y la sensibilidad del método de detección (Mitton *et al.*, 2002; Guo *et al.,* 2015).

Los estudios de asociación (genotipo y fenotipo) puede ayudar a entender si las especies presentan cambios en su adecuación, el mecanismo y efecto de la Selección Natural y así, comprender el mantenimiento y dinámica ecológica, procesos de evolución de genes, los patrones de cambio y aislamiento, con lo que en perspectiva permitiría platear mejores estrategias de migración asistida y conservación (Frichot *et al.,* 2015; Nair *et al.,* 2014; LESC, 2012).

Reconocer rasgos adaptativos parten de la caracterización de la variación genética funcional, para la descripción de la estructura y organización de la expresión de transcritos, la evolución de genes e identificar genes candidatos para el estudio de la adaptación. Además hace posible el descubrimiento de nuevos marcadores genéticos a gran escala para construir perfiles de variación que se puedan extrapolar para tratar de entender la variación fenotípica e inferir procesos de diferenciación dado el efecto de distintos procesos evolutivos.

Pinus pinceana (Pinaceae; Gordon & Glend) es un pino piñonero endémico de México, que se distribuye en suelos calcáreos, rocosos en condiciones de extrema aridez. Se conocen pocas poblaciones, pequeñas, dispersas y discontinuas, en la Sierra Madre Oriental (Figura 1).

La especie está amenazada por su distribución limitada, el aumento en la erosión del suelo, el sobrepastoreo e incendios inducidos, está dentro de la lista de especies amenazadas de la legislación nacional en la categoría de especies sujetas a protección especial (*Pr*) (NOM-SEMARNAT-059-2010) y en el índice internacional de la Unión Internacional para la Conservación de la Naturaleza (UICN) se encuentra catalogada como especie casi amenazada (*Nt*).

Estudios genéticos previos en *P. pinceana* (Escalante, 2001; Ledig *et al.* 2001; Molina-Freaner *et al.* 2001) han mostrado a partir de diferentes marcadores, que existe una alta diferenciación genética entre sus poblaciones en comparación con otras especies de pinos (*e. g.*, Delgado *et al.*, 1999; Cuenca, 2003; Karhu *et al.*, 2006).

Escalante (2001) determinó que esta diferenciación está correlacionada con la separación espacial entre las poblaciones y la evidencia del DNA del cloroplasto (Figueroa, *en preparación*; Figura 2) sugiere que los dos grupos genéticos conformados por las poblaciones del Norte y del Sur de esta especie, se han mantenido aisladas y sin flujo génico.

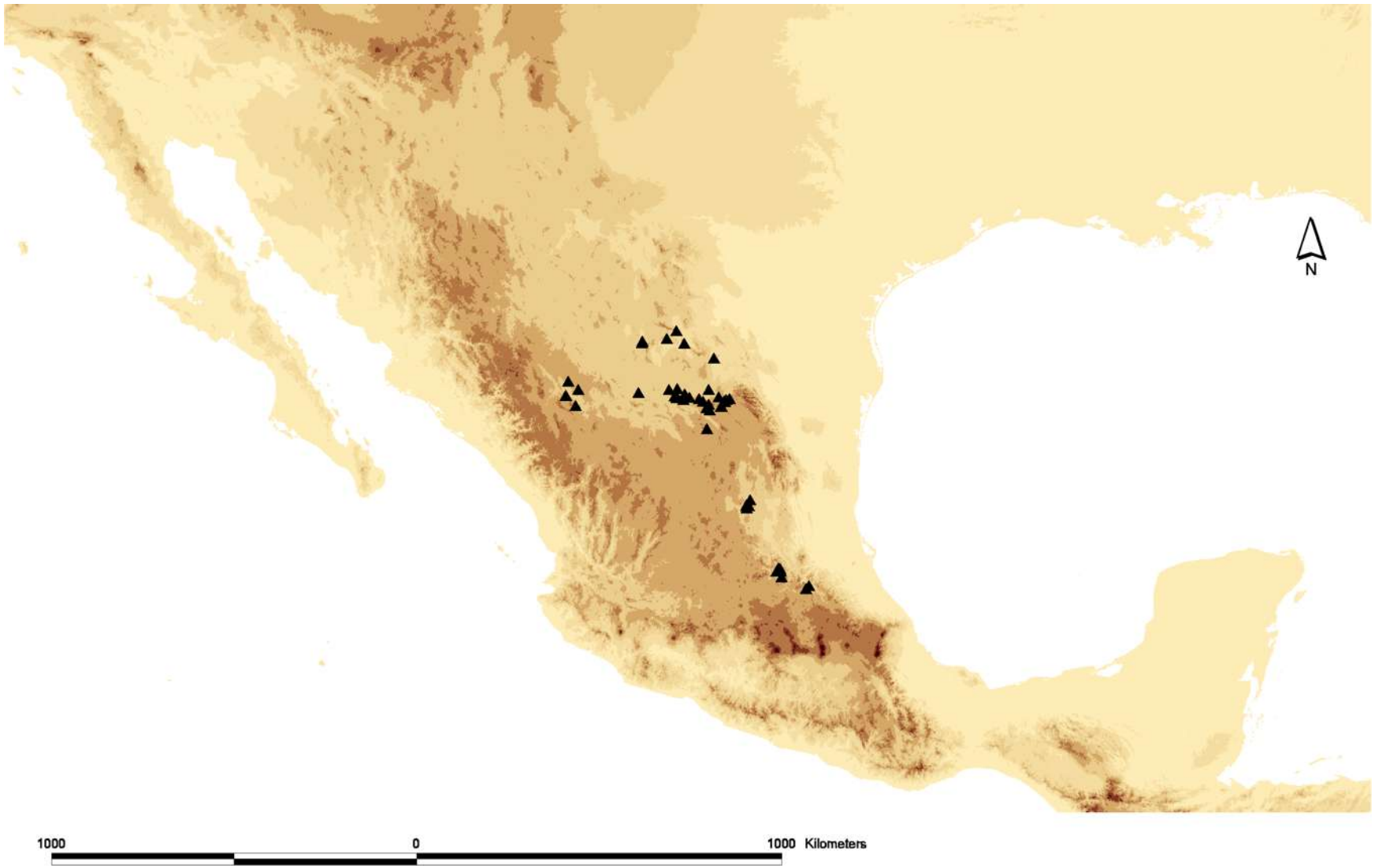


Figura 1. Mapa de las localidades de *Pinus pinceana*.

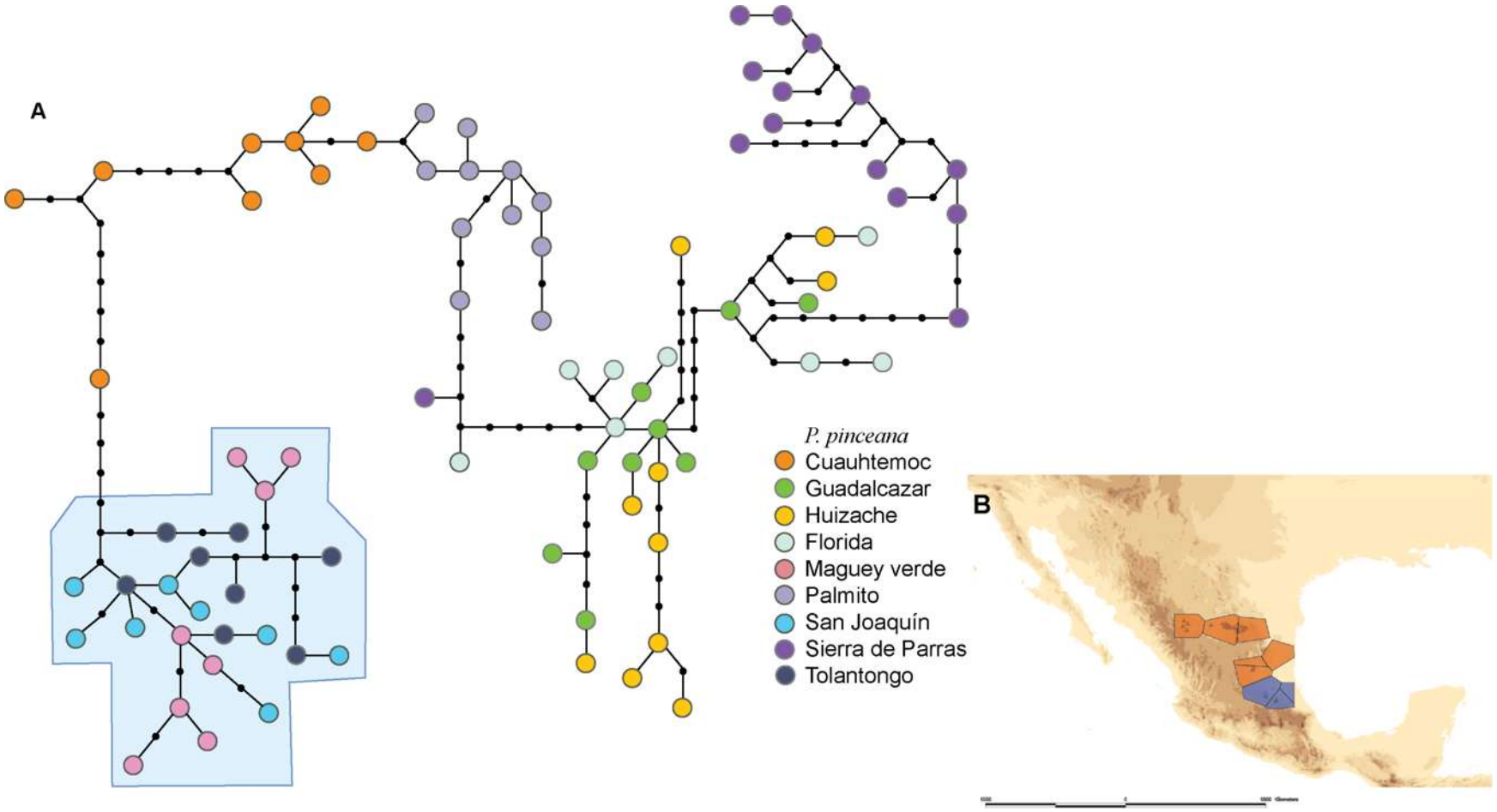


Figura 2. Red de haplotipos reconstruida a partir de 23 cpSSR, cada población es representada por un color y dentro del recuadro se integran la estructuración poblacional que corresponde a las poblaciones en el Sur de la distribución (tomado de Figueroa, *en preparación*

Estudios fenotípicos han reportado una amplia variación que sugiere una respuesta adaptativa al estrés hídrico y al tipo de suelo a lo largo de la distribución. Córdoba (*et al.,* 2008) han detectado cambios significativos en las cubiertas cerosas de las acículas y la proporción de follaje seco en función de su distribución. Martinez Martiñon *et al.*, (2010) encontraron una correlación significativa entre el potencial de crecimiento de la raíz y el estrés por temperatura, que pudo inferirse a partir de la reducción en la tasa de crecimiento de las raíces.

La distribución fragmentada, la alta diversidad genética y la estructuración poblacional que presenta *P. pinceana*, hacen de esta especie un buen modelo para identificar la variación involucrada en la adaptación a diferentes climas. Así, el reconocimiento y análisis de RNAseq del genoma de *P. pinceana* ayudará a comprender la dinámica entre el genoma y el cambio fenotípico modulado por el ambiente.

Este trabajo que constituye la primera caracterización de una conífera endémica mexicana comprende la descripción del transcriptoma de *Pinus pinceana* con el fin de analizar la diversidad de proteínas a partir de distintos tejidos y estadios, para reconocer las respuestas fisiológicas activas, y los marcadores que puedan ser utilizados para identificar la respuesta del genotipo en ambientes contrastantes. El trabajo está dividido en dos capítulos, en el primero se incluye todo el trabajo descriptivo bioinformático sobre los transcritos secuenciados por RNAseq e incorpora ejercicios comparativos entre caracterizaciones previas con otros organismos y el reconocimiento de las rutas metabólicas activas.

En el segundo capítulo se presentan estrategias para reconocer los caracteres genotípicos, fenotípicos y ambientales que se encuentran diferenciados a lo largo de la distribución de *P. pinceana*, con el objetivo de identificar marcadores y caracteres ligados a cambios adaptativos de la especie en distintas condiciones ambientales.

Capítulo I

Caracterización del transcriptoma de *Pinus pinceana*

Antecedentes

La genómica es la disciplina que se ha modulado con la interacción de herramientas tecnológicas y metodologías que permiten analizar el contenido, variación y estructura de todo el material genético de un organismo (Wiley-Liss *et al.*, 2002). El análisis genómico permite aumentar la resolución para la descripción de eventos de recombinación, mutación, detección de pseudogenes, elementos móviles, huellas de selección e identificación de marcadores moleculares **(**SSR's, SNP's, EST's).

La aplicación de estas herramientas, a partir de la reconstrucción y caracterización del genoma, permite el planteamiento de enfoques comparativos que buscan entender procesos evolutivos y ecológicos, con evidencias en procesos adaptativos, purgas genéticas y diferencias en la expresión de los genes para entender procesos de especiación (*e.g.,* Wachowiak *et al.*, 2015; Nadeau *et al.,* 2013; Niu *et al.,* 2013).

El avance de estas herramientas ha sido abrumador en los últimos cinco años se ha ampliando el campo, los costos se han reducido y el manejo de datos se ha vuelto más accesible. Actualmente se cuenta con 67,632 genomas secuenciados (GOLD; Genomes Online Database; 1,139 Archaeas, 50,409 Bacterias y 11,534 Eucariotas).

La secuenciación de RNA (RNAseq) se refiere a los procedimientos experimentales y análisis del RNA transcrito, en conjunto, todo el material transcrito de un organismo en un momento específico se denomina transcriptoma (Wiley-Liss *et al.*, 2002).

La caracterización de transcriptomas representan un esfuerzo descriptivo de la genómica funcional para identificar secuencias de RNA codificante, reconocer cambios en la estructura transcripcional, identificar nuevos marcadores genéticos, reconocer la expresión de genes y las redes de regulación de la expresión genética (Prunier *et al.,* 2015).

Estrategias de RNAseq representan una herramienta muy útil para entender las respuestas fisiológicas a partir de las reconstrucciones de los niveles de expresión de distintos tipos celulares en condiciones específicas. Provee mecanismos para medir y cuantificar la transcripción, localizar la expresión diferencial de transcritos entre distintos tejidos o tipos celulares (*e.g.,* Zhao *et al.*, 2005)

Así, los planteamiento en estudios comparativos entre la expresión en distintas condiciones conforman el primer paso para crear ensayos de secuenciación selectiva,

construir perfiles de expresión génica y el reconocimiento de eventos adaptativos a partir de la dinámica de las respuestas genéticas al ambiente (Savolainen *et al.,* 2007).

La reconstrucción de un transcriptoma implica la manipulación de una gran cantidad de datos y demanda grandes capacidades computacionales y retos bioinformáticos. Es importante hacer distinción en que el procesamiento y manejo de datos en un organismo no modelo depende en gran medida del tamaño del genoma del organismo, las referencias disponibles, la calidad y profundidad de la secuenciación.

Los genomas de las coníferas son muy grandes, de entre los 18 y 40Gb (Birol *et al.,* 2013; Wegrzyn *et al.,* 2014). Particularmente para el genoma de las pináceas se ha calculado 16 veces más grande que otras gimnospermas y 2.400 veces en comparación con angiospermas. Por ejemplo, se calcula que el genoma de *Pinus taeda* es 125 veces más grande que el genoma de *Arabidopsis thaliana* (Bennett y Leitch, 2005). En los pinos el genoma se organiza en 12 cromosomas y ha sido estimado en 30Gb, es un genoma grande y complejo con abundantes regiones intergénicas y de pseudogenes (Neale *et al*., 2014).

Se ha sugerido que el gran tamaño del genoma es resultado de la proliferación de elementos móviles y la retención de amplias regiones intergénicas y pseudogenes, ya que no existe evidencia de poliploidización.

Prunier *et al.,* (2015) señala que los genes de coníferas tienen un origen antiguo y muy conservado, esto a partir de que los eventos de duplicación de genes son anteriores a la división entre las angiospermas y las gimnospermas (hace 300 millones de años (Ma)).

Dado su tamaño, fue hasta la reciente reducción de costos y la optimización tecnológica que ha sido posible la reconstrucción de cuatro genomas completos, *Picea glauca*, *Picea sitchensis, Pinus taeda y Pinus lambertiana* (Birol *et al.,* 2013; Nystedt *et al.,* 2013; Neale *et al.,* 2014; Wegrzyn *et al.*, 2014). Se han caracterizado 28,354 y 50,172 genes en los genomas de *Picea sitchensis* y *Pinus taeda* respectivamente con regiones intergénicas muy grandes (≅2.7 Kb).

En lo que respecta a la estructura de genes, los genes de coníferas tienden a acumularse entre intrones muy largos, 60 Kb (*Picea abies*; Nystedt *et al.,* 2013) y 120 Kb (*Pinus lambertiana*; Wegrzyn *et al.,* 2014), además, del 30% a 40% de genes identificados no han podido ser caracterizados pues carecen de similitud con secuencias de genes conocidos (Eckert *et al.*, 2010; Kovach *et al.*, 2010; Nystedt *et al.,* 2013, Neale *et al.,* 2014; Wegrzyn *et al.,* 2008, 2014).

Gramzow *et al.,* y Guillet-Claude (2014; 2004) han detectado patrones de cambio lento en familias de genes KNOX-I y los genes MADS box en comparación con angiospermas. Mientras que Pavy y Rigault (*et al.*, 2012; *et al.,* 2012) han descrito que los factores de transcripción son menos abundantes en comparación con el número registrado en angiospermas.

Existen numerosos esfuerzos a nivel de análisis de transcriptomas para caracterizar patrones de expresión específicos en coníferas y se dispone de bases de datos para *Pinus spp., Picea spp., Pseudotsuga menziesii* y *Cryptomeria japónica* (Raherison *et al.*, 2012; Pavy *et al*., 2008; Canales *et al.*, 2014; Parchman *et al.*, 2010; Keeling *et al.,* 2011). Este grupo de plantas representan un buen modelo en estudios de genómica comparada, dada la estructuración genética de las poblaciones, su amplia distribución en diferentes ambientes, su sistema de apareamiento, sus grandes tamaños demográficos y el decaimiento en el desequilibrio de ligamiento debido al tamaño de su genoma (Keller *et al.,* 2011).

En éste capítulo se procura un esfuerzo descriptivo para reconstruir el transcriptoma de *Pinus pinceana* con la caracterización de mRNA a partir de RNAseq e identificar las proteínas presentes e inferir las rutas metabólicas y respuestas fisiológicas activas.

Metodología

Colecta del material biológico

Con la finalidad de tener una muestra más representativa de la diversidad genética de la especie, se colectaron dos muestras de tejido que incluyen acículas de individuos juveniles y adultos seleccionados al azar en las poblaciones de dos localidades Maguey Verde, Queretaro y Antena Núñez, San Luis Potosí (Tabla I.1; Figura I.1). El material colectado se mantuvo en una solución de RNAlater (SIGMA) o nitrógeno líquido y posteriormente se almacenó a -80ºC para la preservación adecuada del RNA.

Tabla I.1. Georreferencias de las localidades muestreadas.

Población		**Longitud**	**Latitud**	**Altitud** (msnm)
Antena Núñez	Guadalcázar, San Luis Potosí	-100.48	22.68	1900
Maguey Verde	Peñamiller, Querétaro	-99.66	21.11	2300

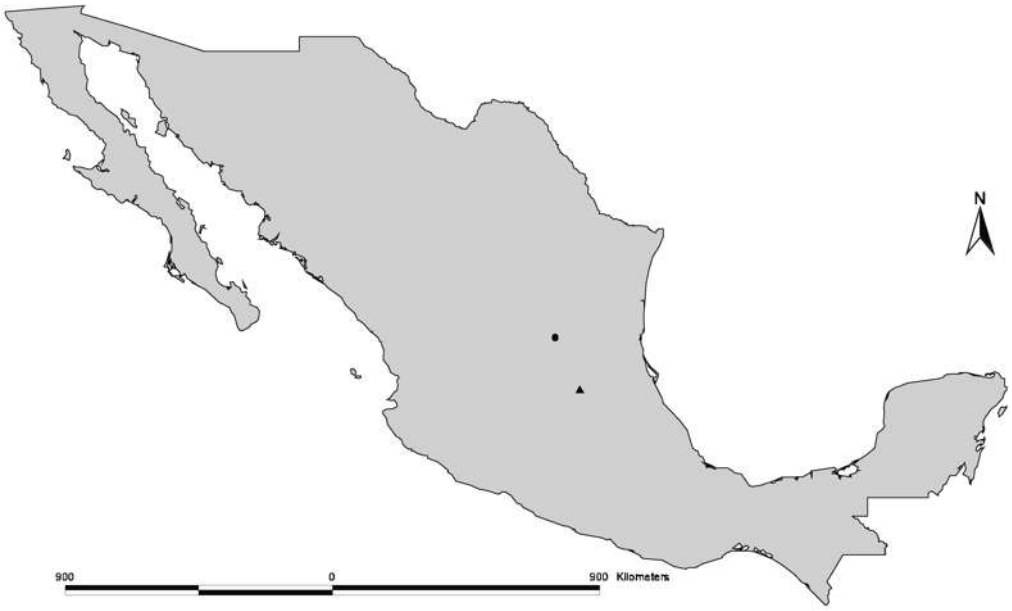


Figura I.1. Mapa con los sitios de colecta para el procesamiento genómico. Se representa a Maguey Verde con un triángulo oscuro y a Antena Nuñez con un círculo.

Extracción y purificación del mRNA

Para la extracción se utilizó la modificación del protocolo de Chang (1993). Debido a la cantidad de polisacáridos, fenoles y demás componentes secundarios fue necesaria la duplicación del procedimiento de extracción para recuperar concentraciones suficientes de mRNA.

La concentración de cada muestra se visualizó en un gel de agarosa al 1% con tinción de Bromuro de etidio y fue cuantificada y evaluada a partir de la proporción 260:280 con un espectrofotómetro (Thermo Nanodrop lite). La muestras en las que se apreció en el gel una mejor definición de las bandas de rRNA, menor ocurrencia de sales y cuantificadas con una mayor concentración a 40 ng/μl, se purificaron para eliminar DNA utilizando el kit RNA clean up (Qiagen) siguiendo el protocolo del fabricante. Se corroboró la eficiencia del procedimiento de purificación cuantificando la concentración con un espectrofotómetro (Thermo Nanodrop lite).

Finalmente las distintas extracciones provenientes de una misma localidad fueron mezcladas para su procesamiento en el *Genomics Sequencing Laboratory* del *Institute of Quantitative Biosciences* en *University of California*, Berkeley QB3 Vincent J. Coates, en donde se llevó acabo una revisión de una electroforesis de capilares en un bioanalyzer 2100 con chips RNA 6000 Nano Labchips (Agilent Technologies Ireland) para evaluar la

calidad y concentración a partir de la proporción 28S:18S (*28S:18S ratio*) en el RNA total y el Índice de Integridad del RNA (RIN).

Se realizó la retrotranscripción de cada muestra a cDNA, secuenciando ambos extremos (3'-5' y 5'-3') utilizando procedimientos clonales para la formación de las librerías genómicas, cada una fue etiquetada para su reconocimiento bioinformático.

Para favorecer la homogeneidad de fragmentos, facilitar la detección de transcritos de bajo nivel de expresión, y reducir la proporción de fragmentos de rRNA ambas muestras de cDNA fueron normalizadas.

Secuenciación masiva

Se llevó a cabo la secuenciación de fragmentos en la plataforma HiSeq2000 PE 150 de Illumina en el *Genomics Sequencing Laboratory* del *Institute of Quantitative Biosciences* en la Universidad de California, Berkeley QB3 Vincent J. Coates acorde al protocolo estándar que generaron lecturas de fragmentos de 100 pb.

Análisis y procesamiento bioinformático

Los análisis y el procesamiento bioinformático de los datos fue realizado en los servidores del centro de Bioinformática de la Universidad de California, Davis y en los servidores de la Universidad de Connecticut, UCONN Bioinformatics Facility en el Plant Computational Genomics Lab.

Evaluación de la calidad de los fragmentos

Buscando tener mayor eficiencia y cubrir la diversidad de todos los fragmentos transcritos en *Pinus pinceana* se conjuntaron todos los fragmentos respectivos a cada librería.

El análisis de control de calidad de los fragmentos comprendió la exploración a partir de la densidad, largo y calidad de cada fragmento con las paquetería FASTQ Quality Filter en FASTX Toolkit (H Lab, 2010) Solexa QA (LengthSort: Cox *et al.*, 2010) y Sickle v. 1.33 (Joshi y Fass, 2011). Debido a la gran cantidad de datos y para hacer más eficiente su procesamiento computacional, se realizaron filtros de calidad a un índice de calidad de 35 y un largo de 50 pb.

Ensamble

El proceso de empalme y alineamiento de lecturas cortas de secuencias (*reads*) se llevó a cabo para cada una de las librerías. Se utilizaron dos estrategias bioinformáticas, el ensamble *de novo* y el ensamble de referencia, que permitieran detectar el procedimiento que reconstruyera fragmentos más largos y que se describen a continuación.

i. Ensamble *de novo*

Se realizaron ensambles simples y para evitar la redundancia de *reads*, se utilizaron ensambles normalizados con el programa Trinity v. 2.06 (Grabherr *et al.*, 2011).

ii. Ensamble de referencia

Se realizaron ensambles guiados con los genomas de referencia de *P. lambertiana* y *P. taeda* (Neale *et al.,* 2014; Wegrzyn *et al.*, 2014). Este procedimiento se realizó a partir del mapeo de las lecturas en fragmentos de 1000 pb sobre los genomas de referencia, utilizando los algoritmos GMAP-GSNAPL (Wu, 2011) y Bowtie2-TopHat (Langmead, 2009; Kim-Salzberg, 2008). Cada mapeo fue transformado a componentes binarios con la paquetería de SAMtools (Li, 2009), posteriormente se realizó el ensamble utilizando el programa Trinity v.2.06 (Grabherr, 2011).

De manera paralela se hizo un ensamble de referencia con el programa BRANCH (Bao *et al.*, 2013) considerando como guía los genes de los genomas disponibles de *P. taeda* y *Picea glauca* (Birol *et al.,* 2014; Neale *et al.,* 2014; Wegrzyn *et al.*, 2014) mapeando sobre un ensamble de los transcritos producto del ensamble *de novo* normalizado (Trinity v. 2.06 Grabherr, 2011).

Caracterización de fragmentos

Para organizar los fragmentos para su procesamiento se hizo una agrupación a partir de su similitud, longitud y abundancia con el programa UCLUST (Edgar, 2010).

Se detectaron los marcos de lectura (ORF), y se identificaron plenamente los fragmentos que corresponden a genes íntegros mediante el programa Transdecoder (Hass, 2011).

Para buscar coincidencias con las secuencias ya identificadas en bases de datos y descartar secuencias de genes de insectos, hongos, y bacterias que pudieran haber contaminado el tejido, se caracterizó y etiquetó los fragmentos siguiendo el *pipeline* enTAP (Eukaryote Non-Model Transcriptome Annotation Pipeline; Ginzburg, 2014) utilizando uBLAST en la base de datos del NCBI con un corte de un E *value* $<10^{-9}$.

Identificación de dominios funcionales

Para clasificar las secuencias, predecir dominios funcionales y detectar motivos de identificación de familias genéticas se utilizaron tres aproximaciones. 1) con el programa InterProScan (Mitchell *et al.*, 2015), se detectaron las familias génicas, y motivos estructurales como dominios, sitios repetidos, conservados, sitios activos, y sitios de unión a partir de los fragmentos con un ORF identificado. 2) usando BLAST2GO (Conesa *et al.*, 2005) y la clasificación previa de enTAP se realizó una búsqueda exhaustiva a partir de un BLASTx con corte a un *E value* $<10^{-5}$ para anotar los dominios funcionales, la actividad

y las familias genéticas con los motivos clasificados en *Gene Onthology* (GO). 3) A partir de las secuencias traducidas de la clasificación previa de InterProScan se comparó con la base de datos de Kyoto Encyclopedia of Genes and Genomes (KEGG) usando el programa KEGG Automatic Anotation Server (KAAS; Moriya *et al.*, 2007) para detectar las rutas metabólicas y las características funcionales de los transcritos. Los motivos funcionales detectados fueron comparados con la base de GO registrada en TAIR (www.arabidopsis.org).

Los resultados de estas tres aproximaciones se conjuntaron para redondear todas las identificaciones, completar las caracterizaciones, compararlas y reducir las redundancias.

Resultados

Extracción

Las cuantificaciones de concentración iniciales y las evaluaciones en el *Genomics Sequencing Laboratory del Institute of Quantitative Biosciences*, se encuentran resumidas en la Tabla I.2.

Tabla I.2. Descripción de calidad y concentración de las muestras.

Localidad	Tejido	Concentración (ng/µL)	260/280	RIN
Maguey Verde	Acícula juvenil	107.1	2.1	8.1
	Acícula Adulto	117.7	2.0	
	Meristemo	173.1	1.9	
	Estróbilo masculino	388.9	2.2	
Antena Núñez	Acícula juvenil	186	2.4	4.5
	Acícula Adulto	104.3	2.5	
	Meristemo	70.5	2.1	

El RNA total se cuantificó con la proporción 260:280 en un rango entre 1.0 a 2.5, equivalente en el Índice de Integridad del RNA (RIN; del Ingles: *RNA Integrity Index Number*) de 8.10 y 4.50 lo que representa condiciones óptimas y degradaciones parciales respectivamente.

La amplificación clonal de material genético normalizado generó fragmentos con un tamaño promedio de 328 pb.

Secuenciación

En la secuenciación masiva se obtuvieron 115,248,758 *reads* y 275,116,776 *reads* para cada librería respectivamente, con un largo máximo de 102pb sin adaptadores, que se estima equivale a una cobertura aproximada a 32x.

Análisis y procesamiento bioinformático

Evaluación de la calidad de los fragmentos

En la revisión de calidad con la paquetería de FASTX Toolkit se cuantificaron las medias del índice de calidad (*Q value*) entre 33 y 39. Posteriormente con los programas Sickle v. 1.33 (Joshi, 2011) y Solexa QA LengthSort (Cox, 2010) se filtraron 252,807,146 *reads* para ambas librerías de fragmentos de longitud mayor a 50pb y un índice de calidad mayor a 30, con el propósito que fueron los mejores fragmentos en el procesamiento bioinformático.

Ensamble

Se generaron las estadísticas para evaluar la construcción de *clusters*, rearreglo y largo de los ensambles, y estimados de número de transcritos, así como los estimados de número de genes hipotéticos, en cada uno fueron descartados los fragmentos menores a 300pb. La eficiencia del ensamble fue cuantificada por la variable N50 (que se refiere a la longitud más corta dentro del 50% de los *scaffolds*). Obteniendo según el procesamiento valores desde 481 hasta 654 (ver Tabla I.3).

Para un manejo óptimo de los datos y hacer más eficiente la anotación del transcriptoma, se utilizó el ensamble con el consenso de los transcritos procesados con el ensamble normalizado que posteriormente fueron referenciados con los genes de *Picea spp*. y *Pinus spp*.

Tabla I.3. Comparativa con los estadísticos de los productos en procesos de ensamble.

	Ensamble	Procesamiento	N50	Longitud promedio [pb]	Transcritos	Genes**
Maguey Verde	*dn*	Trinity estándar	**562**	554.42	208253	105352
		Trinity normalizado*	**640**	608.47	301897	90389
	r	Trinity normalizado BRANCH	**654**	608.47	311531	122948
Antena Núñez	*dn*	Trinity estándar	**512**	513.6	98553	56603
		Trinity normalizado*	**534**	540.01	107339	51039
	r	Trinity normalizado BRANCH	**546**	548.89	112925	58616
Consenso	*r*	Trinity normalizado BRANCH	**481**	497.17		117824

***Ensamble considerado para formar el consenso; *dn* ensamble *de novo r* ensamble de referencia

****Fragmentos inferidos como genes hipotéticos

Caracterización de fragmentos

Para la clasificación y ordenamiento de fragmentos, la aproximación a partir de agrupación en UCLUST (Edgar, 2010) generó para el consenso de transcritos 117,824 contigs, lo que corresponde al 37% de genes predichos (ver Tabla I.4). En el análisis con Transdecoder (Hass, 2011) se restringió la identificación de ORF's en cada fragmento, y se encontraron 47,510 secuencias de genes únicos restringiendo la búsqueda en Pfam.

Tabla I.4. Identificación de regiones codificantes.

	UCLUST	**Transdecoder**
Reads	306806	117824
Clusters	179111	
Tamaño promedio	2.7	500
Grupos/ Genes únicos	117824	47510

La caracterización de fragmentos con enTAP se llevó a cabo a partir de dos bases de datos de los contigs codificados en aminoácidos y filtrados para posiciones no redundantes (*nr*). Se encontraron de los 47510 *reads*, 35966 transcritos correspondientes a genes descritos, 2589 sin identificación previa y 8915 transcritos desconocidos (ver Tabla I.5).

La identificación de secuencias previamente descritas corresponde en mayor proporción a descripciones en especies modelo que a la semejanzas filogenéticas en especies cercanas de coníferas, siendo *Vitis vinifera* la especie con la que se encontraron mayor proporción de coincidencias (ver Figura I.2). En proporción, de las 35,966 secuencias que fue posible describir como genes sólo el 5% (8632) corresponde a genes previamente descritos en coníferas (*Picea sitchensis*).

La proporción de fragmentos que no pudieron ser caracterizados equivale al 24% del total de transcritos ensamblados, de los cuales el 22% de estos corresponde a genes identificados pero sin caracterización previa, el resto corresponde a secuencias aún no registradas en la base de datos de NCBI (http://www.ncbi.nlm.nih.gov/).

Tabla I.5. Resultados de la caracterización de transcritos con enTAP (uBLAST).

Transcritos	47510
Longitud promedio	501.4 pb
Secuencias caracterizadas	35966
Secuencias sin caracterización	2589
Secuencias sin correspondencia	8915
Secuencias contaminantes	40

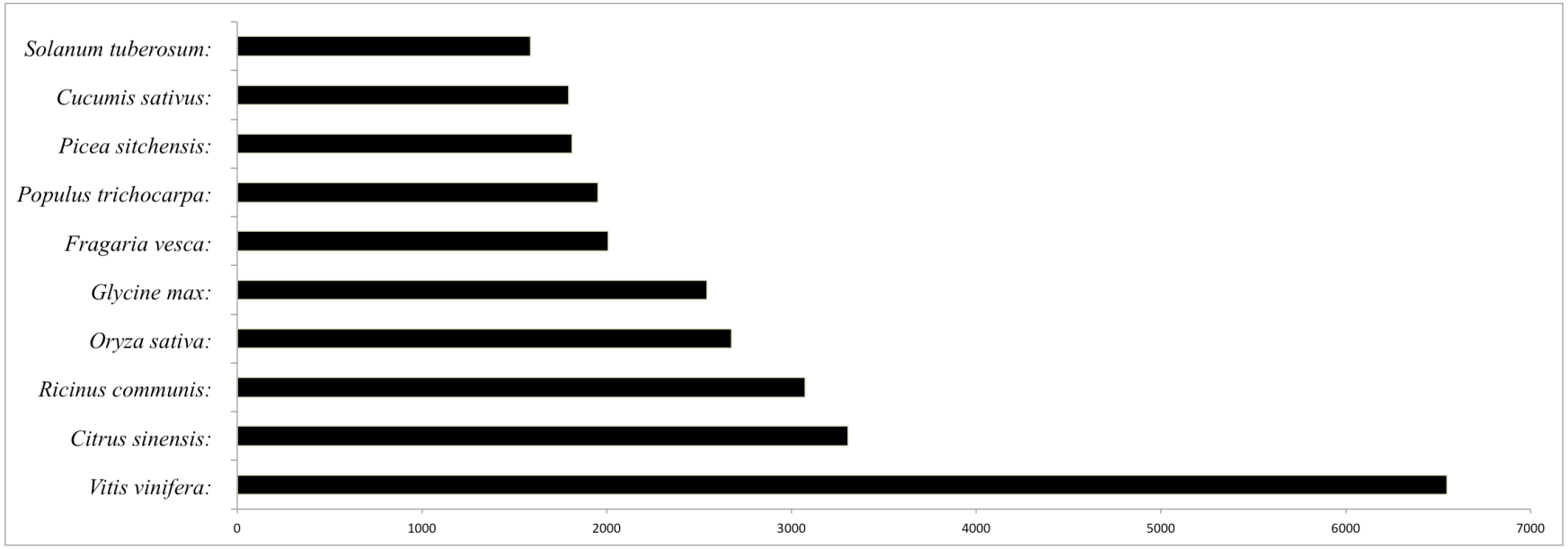


Figura I.2. Coincidencias por especie para la caracterización de fragmentos acorde al uBLAST.

En la búsqueda y filtrado de contaminantes se obtuvieron 40 descripciones correspondientes a secuencias características de los hongos patógenos *Rhizoctonia solani* (4), *Marssonina brunnea* (2), *Grosmannia clavigera* (2) y *Aspergillus kawachii* (2); de la bacteria *Gillisia* sp. (2) y de la hormiga *Acromyrmex echinatior* (2).

Identificación de dominios funcionales

En la identificación a partir de los identificadores del mapeo de GO se logró separar las secuencias implicadas en funciones moleculares, procesos biológicos y actividades ligadas y/o específicas a organelos.

Esta clasificación no es exclusiva para las categorías, es decir una proteína pudo ser identificada con varios motivos distintos codificados para varias funciones y procesos metabólicos (*e.g., Cystathionine beta-lyase*, participa en cinco rutas metabólicas: los metabolismos de metionina, cisteína, selenoaminoácidos, del nitrógeno y sulfuros).

Según el proceso biológico existen 5,208 genes involucrados en procesos metabólicos, y 2,048 genes relacionados a la regulación de los procesos biológicos. Las funciones moleculares con mayores coincidencias describen actividades catalíticas (52%) y de transporte (30%). Fue posible asociar la presencia de los genes identificados con las actividades ligadas a organelos, en acciones metabólicas específicas o como componentes estructurales (Figura I.3).

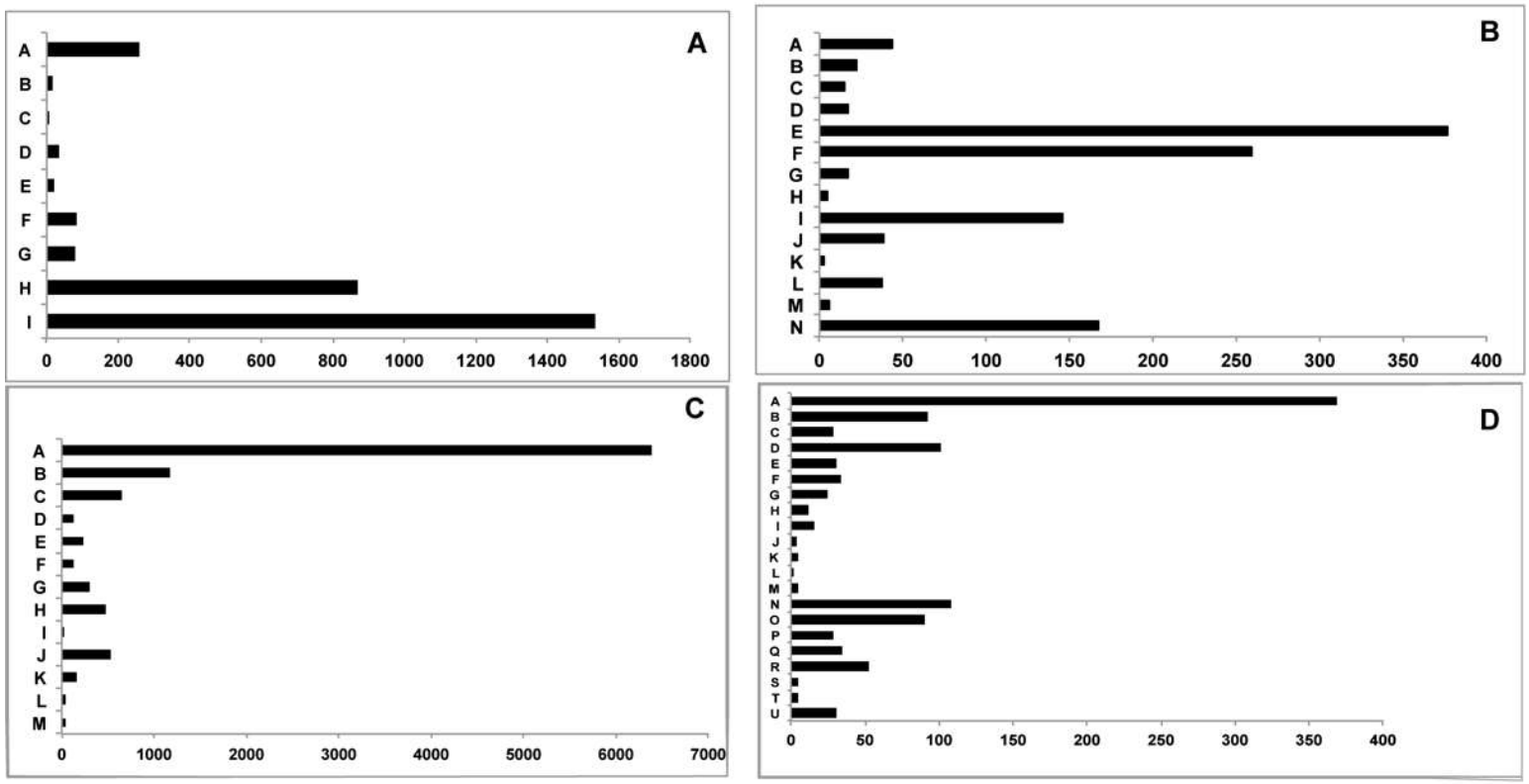
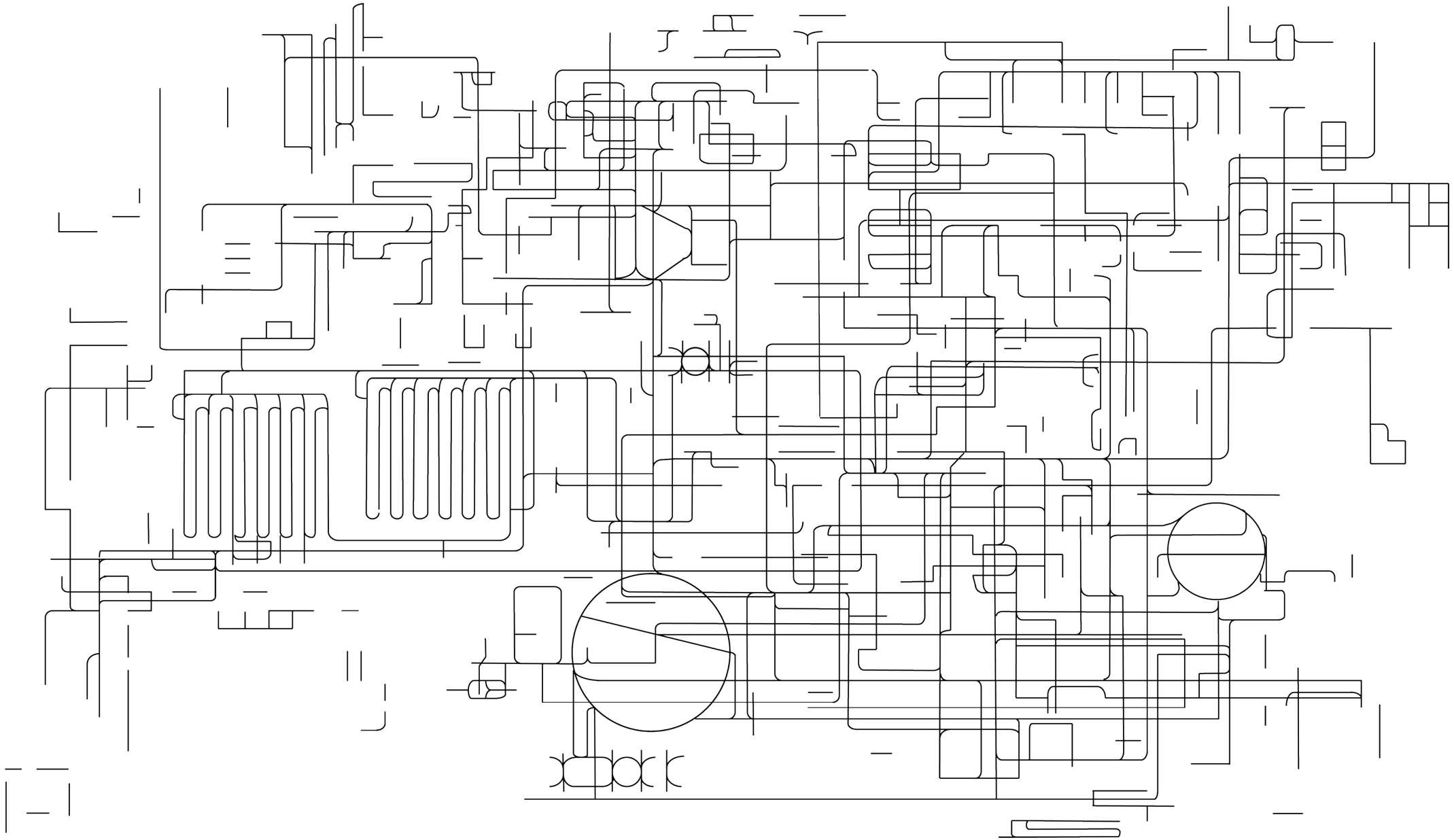


Figura I.3. Asociaciones de función de los transcritos. **A.** De acuerdo a función: (A) Catalítica (B) Transportador (C) Acarreador de electrones (D) Receptor (E) Regulador enzimático (F) Estructural (G) Antioxidante (H) Energética-Nutrición (I) Proteína de unión a factor de transcripción. **B.** De acuerdo al proceso biológico donde interactúan: (A) defensa (B) Respuesta al calor (C) Unión (D) Utilización de Nitrógeno (E) Proliferación celular (F)Crecimiento (G) Reproducción (H) Muerte celular (I) Estrés (J)Señalización (K)Localización (L)Regulación basal (M)Estrés hídrico (N) Resistencia a enfermedades. **C.** Organelo o sitio asociado a la función (A)Citoesqueleto (B)ribosomas (C) Cromatina (D)Ap. Golgi (E) Retículo Endoplasmico (F) Vacuola (G) Plastidos (H)Mitocondria (I)Cloroplasto (J)Núcleo (K)Membrana (L)Lisosoma (K) Peroxisoma. **D.** De acuerdo al proceso metabólico involucrado: (A) Biosíntesis de metabolitos secundarios (B) Metabolismo de Carbono (C) Metabolismo de ac. grasos (D) Interacción Planta-Patógeno (E) Fotosíntesis (F) Fijación del Carbono (G) Metabolismo del Nitrógeno (H) Biosíntesis de ac. grasos (I) Biosíntesis de hormonas esteroideas. (J) Síntesis de Lipopolisacaridos (K) Síntesis de péptidoglicanos (L) Citocromo P450 (M) Transporte de RNA (N) Replicación de DNA (O) Transducción hormonal (P) Endocitosis (Q) Absorción de minerales (R)Apoptosis (S) (T) (U)Glugolisis

Se logró reconstruir al menos parcialmente genes que involucran las respuestas a estímulos específicos por ejemplo: al estrés (259 genes) en donde se incluyen codificaciones de respuesta a la falta de agua (22 genes), la respuesta al calor (6), la resistencia a enfermedades (44 genes) y defensa contra patógenos (167 genes), entre otras.

Para la clasificación basada en motivos indexados en KAAS (Moriya, 2007) y el KEGG se utilizaron las secuencias traducidas obtenidas en InterProScan (Mitchell *et al.,* 2015) y se encontraron elementos de 751 rutas metabólicas (no excluyentes). Con esta información fue posible reconstruir parcialmente rutas metabólicas utilizando como esquema base la organización de *V. vinifera*, dado el mayor número de coincidencias que se encontró en el análisis de caracterización. Se logró reconstruir gran proporción de rutas de metabolismo del carbono (Figura I.4), se localizo vías de regulación en la respuesta a patógenos (Figura I.5).

No fue posible definir a partir de la base datos de *A. thaliana* las secuencias ortólogas con los motivos señalados por Blast2Go o KASS a partir de los motivos identificados por ontología.



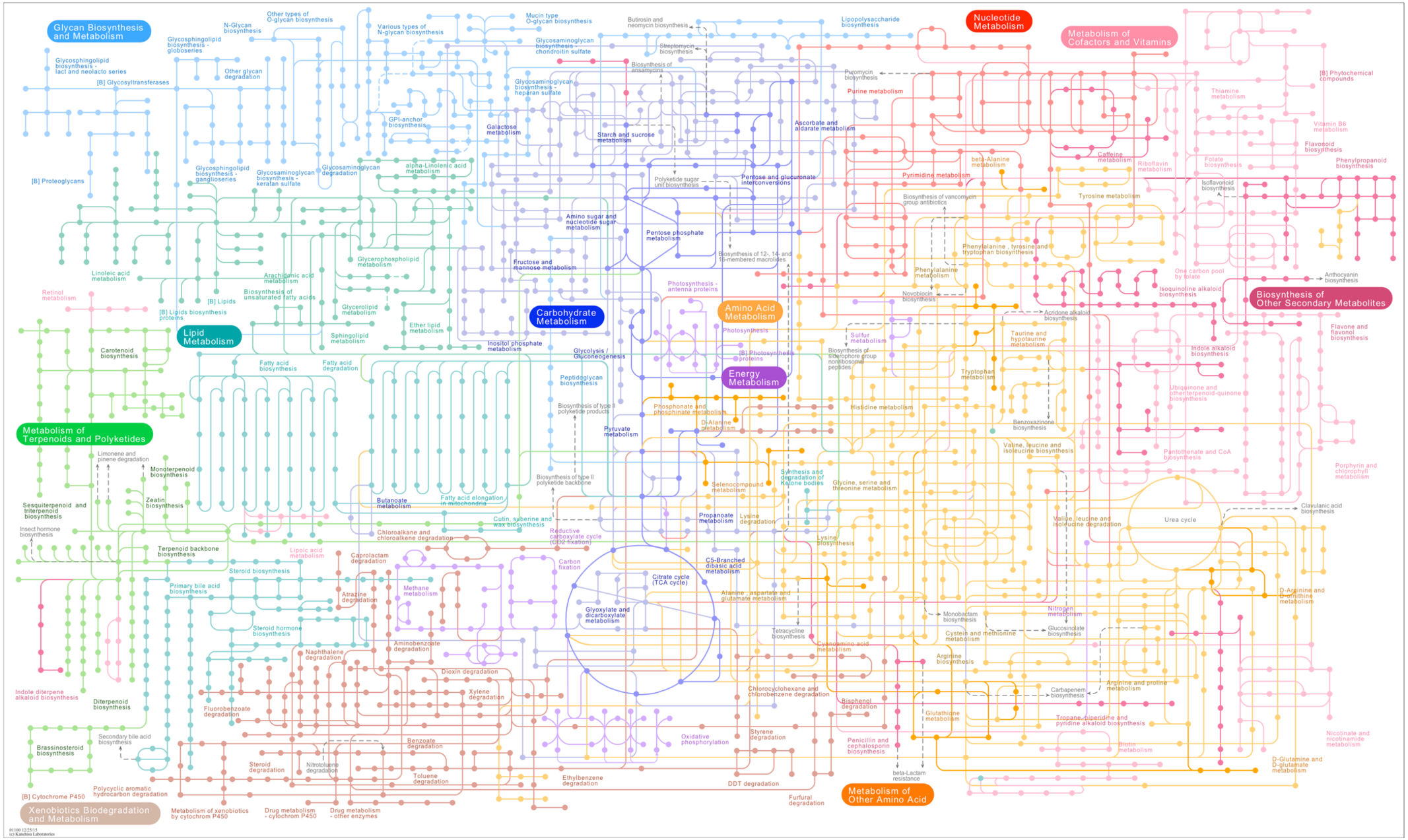


Figura I.4. Esquema de las rutas metabólicas, en líneas negras se encuentra señaladas las rutas metabólicas reconstruidas a partir de los transcritos anotados

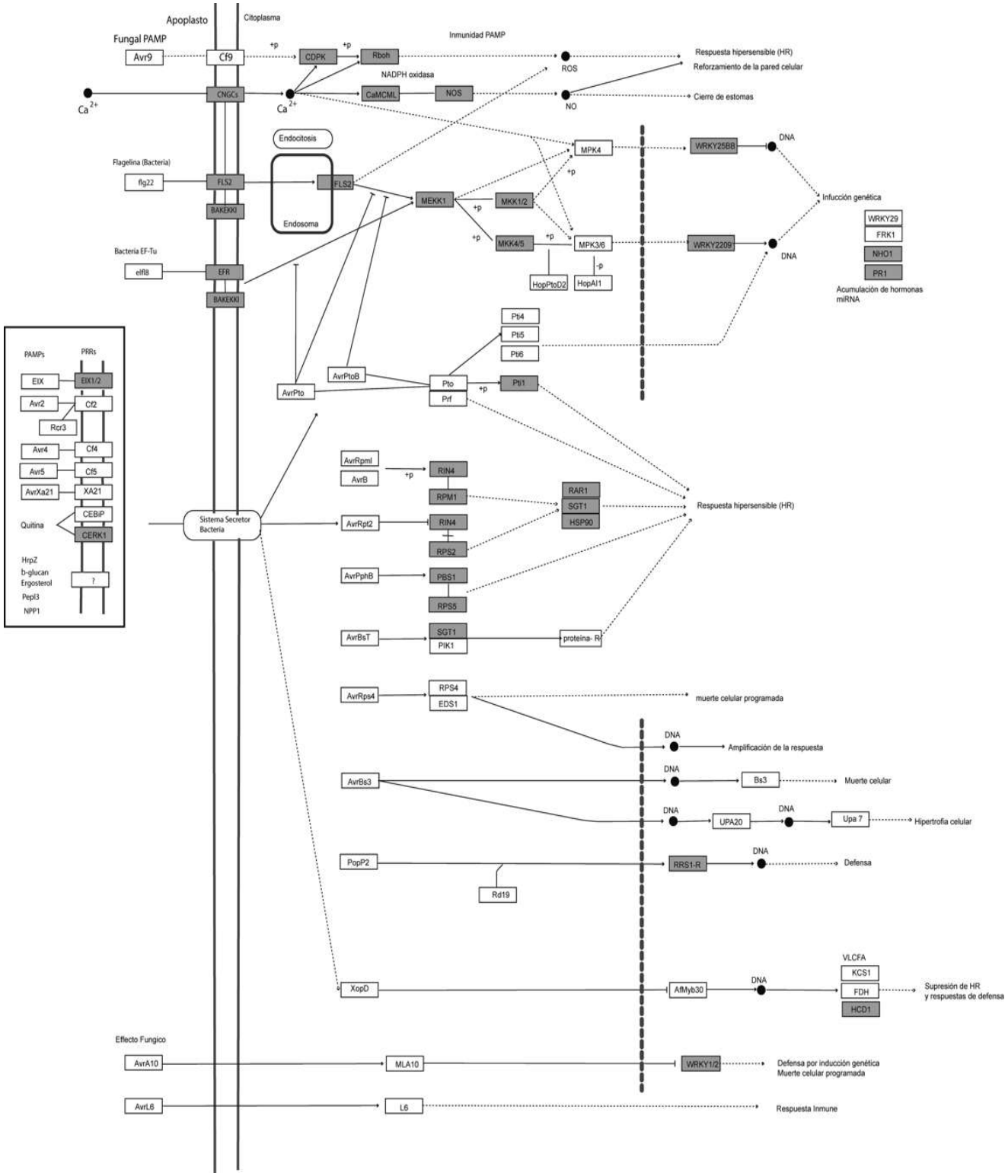


Figura I.5. Cascada de regulación genética en la interacción planta patógeno

Discusión

Si bien existen muchos ejemplos de genómica en microorganismos, son contados los trabajos que existen en genómica de organismos eucariontes no modelo desarrollados en México (Ibarra-Laclete *et al.,* 2013; Tsai *et al.*, 2013; Vielle-Calzada *et al.,* 2009; Qin *et al.,* 2014). Aunado a esto las herramientas bioinformáticas aplicadas en especies no modelo con genomas grandes, tienen una fuerte demanda de recursos informáticos y de capacidades computacionales que sobrepasa las capacidades disponibles en la mayoría de las instituciones dedicadas a la genómica en el país.

No obstante a ello, se obtuvo en este trabajo una gran cantidad de transcritos que describen de manera general las iteracciones y respuestas genéticas en *Pinus pinceana*.

El diseño experimental que se empleó pretendía describir la mayor proporción de respuestas genéticas y de transcritos posibles. Sin embargo con este mismo objetivo se pudieron abordar mejores estrategias, por ejemplo, como aumentar la cantidad de librerías, o bien las líneas de secuenciación no fueron óptimas para tal diversidad de fragmentos y tejidos. Dado que al combinar las muestras de mRNA de múltiples tejidos e individuos se redujo la resolución del trabajo para identificar mecanismos que controlan respuestas especificas en estadios y tejidos. Además de que se restó cobertura en el proceso de secuenciación.

La gran diversidad de transcritos generada en este trabajo fue tan abundante que la normalización de las librerías de cDNA que promueven la homogenización de secuencias no fue suficiente para la secuenciación de transcritos poco abundantes. Wegrzyn *et al.,* (2014) ha señalado que la limitante metodológica más importante para caracterizar fragmentos son las bajas abundancias y las longitudes cortas. Y si bien las cantidades y concentraciones de las muestras eran buenas según los protocolos previos a la secuenciación (Tabla 2), el proceso de extracción, el manejo en el procedimiento, o el traslado al Genome Center pudieron haber favorecido la fragmentación de las secuencias.

En el procesamiento bioinformático a pesar de contar con los recursos computacionales suficientes, surgieron problemáticas ligadas a la longitud y la diversidad de fragmentos, lo que generó la necesidad de explorar gran variedad de herramientas de análisis en cada uno de los procedimientos.

En el ensamble se logró hacer una evaluación comparativa (Tabla I. 3) siendo el método de mayor eficacia la combinación de dos procedimientos, mientras que fue evidente que no es posible aplicar solamente algoritmos de los ensambles de referencia a genoma completo en sistemas con genomas tan grandes, como el de especies del género *Pinus*

aproximación que no había sido abordada previamente. Al comparar con otros transcriptomas de coníferas se encontraron similitudes en el tamaño y número de genes encontrados, secuenciados y descritos (Parchman, *et al.,* 2010; Shi-Niu *et al.*, 2013; Canales *et al.,* 2014; Wachowiak *et al.,* 2015) al igual que con los seis genomas secuenciados de gimnospermas (*Gynkgo biloba, Welwitschia mirabilis, Pinus taeda, Pinus lambertiana y Picea glauca* y *Picea abies*),

Se consideró a las referencias de N50 como el estimado de calidad del ensamble, donde observamos un cambio evidente entre métodos, donde valor más alto para el mejor procedimiento dobla el valor de N50 lo que evidencia la abundancia de fragmentos cortos.

Okeke (2014) señala que la abundancia de fragmentos cortos en un ensamble es debida a baja cobertura en la secuenciación, dado que entre más largos son los *scaffolds* de un ensamble el proceso de caracterización puede ser más especifico y completo. Además de este argumento también se asocian la abundancia de registros para el sistema y/o de homología con los fragmentos descritos para poder caracterizar adecuadamente a los transcritos (Wegrzyn *et al.,* 2014).

Las identificaciones a partir de las coincidencias en las bases de datos hacen evidente la desproporcionalidad que existe en los avances de la genómica en las especies no-modelo. Encontrar pocas referencias a genes descritos puede explicarse por diversos factoresmetodologicos: i) durante el procesamiento de la muestra el fragmento se haya degradado, o no haya sido lo suficientemente abundante para ser secuenciado dada la viabilidad bioquímica o la vida media del compuesto, ii) que el fragmento no haya sigo secuenciado con una buena calidad.

Además, puede estar sujeto a un efecto de la poca información de los genomas pues sólo existen seis genomas secuenciados de gimnospermas (*Gynkgo biloba, Welwitschia mirabilis, Pinus taeda, Pinus lambertiana y Picea glauca* y *Picea abies*), y es probable que correspondan a genes que son propios de grupos filogenéticos más cercanos de gimnospermas o del grupo de angiospermas basales (*e.g., Amborella*, *Austrobaileyales*, *Chloranthaceae*, y *Nymphaeales*).

El uso de distintas estrategias de caracterización hizo posible reconocer mayor proporción de transcritos, estimar los factores que participan en vías de señalización, el crecimiento celular y el desarrollo, mecanismos de defensa y otras vías metabólicas a partir de los dominios proteicos. La caracterización de estos 35,966 genes nos permitió generar mapas genéticos y reconstruir parcialmente rutas metabólicas muy importantes, que filogenéticamente son muy conservadas (Figuras I.4 y I.5).

Mapear las rutas metabólicas especificas nos permite reconocer patrones y respuestas locales, por ejemplo fue posible reconocer las interacciones de la respuesta a la interacción planta-patógeno, y si consideramos que la detección de secuencias contaminantes de hongos endófitos y parásitos en el tejido indican que más que una contaminación ocasionada por el mal manejo del tejido, son producto de contaminaciones en el medio natural .

Según Bonello *et al.,* (2006) el mecanismo de acción de las fitoalexinas en las coníferas se desconoce, y sólo es a partir de la acumulación de intermediarios como PR1 (*Pathogen resistance gene*), NH01 (*nonhost resistance gene*) que se ha inferido esta ruta de defensa a patógenos. Las fitoalexinas son compuestos de bajo peso molecular sintetizados *de novo* después de una infección microbiana en respuestas locales y sistémicas de resistencia, incluyen a los isoflavonoides, pterocarpanos, stilbenas y saponinas (Hammerschmidt, 1999).

Si bien la inferencia para la caracterización parte de la homología y de la ortología de secuencias, determinar y mapear procesos precisamente requiere de hacer análisis comparativos entre los transcritos de otras especies más cercanas filogenéticamente. Por ejemplo, Canales *et al.,* (2014) logra correlacionar los genes asociados al crecimiento con respuestas ambientales para poder estimar la diversidad y evolución de genes dentro de las coníferas.

De manera inequívoca se requiere tanto del avance tecnológico y de la disposición de tecnología y espacios computacionales para este tipo de procedimientos sin embargo, la principal limitante en el manejo de datos yace principalmente en el planteamiento y diseño experimental. El reto actual para la genómica en gimnospermas es reconocer las funciones en los genes que hasta ahora son desconocidas, para lo cual se requiere de la integración con la proteomica y metabolomica en estos sistemas que describan la rutas especificas, deduzcan los sesgos y puedan generar datos mucho mas específicos para la coníferas.

Conclusiones

La caracterización del transcriptoma de *P. pinceana* provee la descripción de la expresión general en distintos tejidos y etapas del desarrollo, de 47,510 genes transcritos, de los cuales se encontró la descripción y función de 35,916, con los cuales se pueden reconstruir importantes vías metabólicas para reconocer las interacción del genotipo ante procesos de crecimiento, respuesta a patógenos y el metabolismo basal.

La descripción del transcriptoma de *P. pinceana* permite una estimación global y muy general de las interacciones genéticas y fisiológicas de esta conífera en su ambiente, que contribuye a nuestra comprensión de la variación genética en sus poblaciones y el control genético de los rasgos.

La caracterización de este trabajo es una referencia para el futuro planteamiento de estudios de asociación, la búsqueda de nuevos marcadores moleculares (microarreglos, EST, SNP) para relacionar respuesta específica a un estímulo en un ensayo de expresión.

Este trabajo ejemplifica la gran utilidad y rendimiento de las técnicas de NGS y de análisis bioinformático como mecanismos eficientes y rentables para obtener información sobre la codificación de la variación genética.

La caracterización precisa de las proteínas es complicada y depende tanto de la calidad y la diversidad de transcritos amplificados como del avance y reconocimiento de la función de genes. Las caracterizaciones funcionales reflejan el avance sesgado de estas tecnologías en especies modelo y cultivadas (*e.g.*, *Arabidopsis thaliana, Phaseolus vulgaris, Zea mays*)

Las descripciones e inferencias de los análisis en GO y KEGG clasificaron categorías funcionales a fin de comprender sus funciones y formas de regulación de rutas metabólicas.

Este estudio es el primer esfuerzo NGS y de análisis de la función de genes en el transcriptoma de *P. pinceana* y representa el estudio más amplio de caracterización en una conífera mexicana hasta la fecha.

Capítulo II

Una aproximación para estudiar la adaptación local en *Pinus pinceana*

Antecedentes

Una población ante condiciones ecológicas distintas, ya sea por cambios en el ambiente original, por la migración, y/o relocalización de individuos, puede: 1) permanecer, 2) extinguirse o 3) diferenciarse (Schlichting, 1986; Pfennig, 2010). Este último escenario puede deberse a dos vías, la adaptación (Pespeni *et al.,* 2013) o la plasticidad fenotípica (Evans y Hofmann, 2012).

El potencial adaptativo de un organismo a un ambiente depende de la variación genética (Frankham *et al.,* 2002). Los mecanismos evolutivos que fijan estas variantes surgen con alelos neutrales o poco deletéreos que insertan variaciones que podrían ser benéficas en las nuevas condiciones ambientes, y como resultado pueden modificar el fenotipo y/o los niveles de expresión aumentando la adecuación de la población (King y Wilson, 1975; Romero *et al.*, 2012).

La dinámica de los procesos evolutivos, como el efecto fundador, los cuellos de botella, los patrones de mutación, la recombinación y el desequilibrio de ligamiento pueden mimetizar los procesos de selección. Sin embargo las pruebas de neutralidad ayudan a reconocer el efecto de la Selección Natural (Tajima, 1989; Fu Li, 1997; Wang, 2015). Esencialmente se han tratado de identificar procesos de selección positiva y de selección balanceadora (Smith y Haigh, 1974; Darden y Marks, 1988) a partir de cambios en la diversidad nucleotídica, en un panorama comparativo y multilocus. Para identificar la selección positiva, se parte de la detección de mutaciones favorables fijas y por tanto, se presenta como una reducción en los niveles de diversidad nucleotídica. Mientras que en la selección balanceadora, se mantienen distintos alelos durante largos intervalos de tiempo, y así eleva o mantiene la diversidad de los polimorfismos.

La adaptación local supone que existen disyuntivas en la adecuación entre los distintos hábitats que pueden reconocerse en cambios y frecuencias diferenciales en genes que moldean el fenotipo, y que repercuten en las tasas de crecimiento, la capacidad reproductiva y la sobrevivencia (Kuparinen *et al.,* 2010).

Por su parte, la plasticidad fenotípica es reconocida como una respuesta rápida, fisiológica o morfológica de un organismo ante un cambio o condición ambiental por lo que no promueve necesariamente un valor adaptativo local (Bradshaw, 1965). Potencialmente, para identificar las variantes adaptativas, se requiere detectar los

polimorfismos involucrados y la frecuencia de su distribución, para establecer correlaciones entre el ambiente, la estructura poblacional y la diversidad (Endler 1977, 1986).

Schlichting (1986) reconoce que existe una dinámica entre el genotipo, el fenotipo y el ambiente, conformando un paisaje adaptativo, donde el ambiente tiene un efecto sobre el fenotipo que consta de componentes genéticos, entonces, considera a la plasticidad fenotípica como un instrumento de adaptación. Así la distinción entre la plasticidad y la adaptación se ha convertido en una discusión amplia y abierta, pues son procesos que no surgen independientemente, donde uno explica consecuentemente al otro.

Para Lewontin (1957) la variabilidad genética implicada en la adaptación está presente en un grupo, mientras que la plasticidad es un caracter individual, de manera que debe aumentar a medida que la cantidad de heterocigosidad disminuye, por el efecto deletéreo sobre homocigotos recesivos de los componentes genéticos que modulan el rasgo (Schlichting, 1986; Pfennig *et al.,* 2010). Sin embargo, las respuestas plásticas pueden estar correlacionadas entre varios caracteres, tener un efecto aditivo o bien podrían ser diferentes respuestas al mismo factor ambiental.

Detectar los cambios adaptativos depende de la expresión, la fuerza de la selección, el tiempo transcurrido desde la fijación de la mutación benéfica, y la cantidad de la recombinación entre los sitios seleccionados y los neutros (Savolainen *et al.,* 2007). Teóricamente factores determinantes para la regulación de la homeostasis de un organismo en su ambiente son más susceptibles al efecto de la selección (Tiffin y Ross Ibarra, 2014). Por ejemplo, los mecanismos que modulan el estrés a la temperatura y el estrés hídrico serían más fácilmente detectados (Condit *et al.*, 1995; Storey, 2004).

Avances en la genómica evolutiva y la transcriptómica comparada han generado un campo emergente en la formulación de estudios de asociación con el objetivo de entender la dinámica del genotipo, el fenotipo y el ambiente, para reconocer el efecto de los factores ambientales que modulan la variación genética (Rellstab *et al.*, 2015). Los estudios de asociación plantean comparaciones entre poblaciones que han evolucionado en distintas condiciones ambientales, bajo el supuesto de que la Selección Natural ha actuado diferencialmente aumentando la frecuencia de las mutaciones ventajosas según el ambiente.

Desde la perspectiva genómica se consideran dos distintas aproximaciones para establecer correlaciones entre el genotípico y el fenotipo: la primera parte del reconocimiento de cambios y del análisis de la variación fenotípica que estén asociadas a

las variaciones genotípicas, y la segunda, a la inversa procura a partir de la información genómica detectar huellas adaptativas en la variación genética que se reflejen en cambios fenotípicos (Tiffin y Ross-Ibarra, 2014).

En especies forestales se han detectado correlaciones entre la variación genética y los gradientes ambientales. Particularmente se ha tenido especial interés en reconocer rasgos relacionados con el estrés por déficit hídrico, rasgos fenológicos, la aclimatación al frío, la resistencia a la aridez y la sequía (*e.g.,* Zhang y Marshall 1994; Aitken *et al.,* 1995; Skrøppa y Johnsen *et al.,* 1999; Olivas-García *et al.,* 2000; Brendel *et al.,* 2002; González-Martínez *et al.,* 2006; Baltunis *et al.,* 2008; Newton *et al.,* 1991; Ingram y Bartels, 1996).

De manera prospectiva entender el impacto del cambio climático a través de los mecanismos adaptativos que han surgido en especies arbóreas es central para la genómica evolutiva comparada, en particular los mecanismos evolutivos que se han generado en las coníferas que se distribuyen en las regiones áridas.

Las zonas áridas son aquellas regiones cuya provisión de agua es deficiente con precipitaciones medias anuales menores a 400 mm (Le Houerou y Dregne, 1970), conforman el 6% de la superficie forestal mundial (FAO, 2002). En México las zonas áridas y semiáridas ocupan más de la mitad del territorio, en donde se ubican centros de origen y/o diversificación de cactáceas, agaváceas, crasuláceas, y algunas especies forestales como *Pinus y Juniperus* que son los únicos recursos forestales.

Los primeros pinos piñoneros (Sección *Nelsoniae*, Gernandt, 2008) según el registro fósil datan de hace 22 Ma (Mioceno), en zonas áridas del Altiplano Mexicano, donde se establecieron y formaron grandes extensiones boscosas durante las glaciaciones del Cuaternario (Little, 1969; Farjon y Styles, 1997). Sin embargo, actualmente los factores climáticos de precipitación y temperatura y el tipo de suelo limitan la distribución de los árboles a lugares con acumulación de agua de escurrentía o en lugares accesibles al agua subterránea (Granados, 2015).

Bajo esta premisa, si las condiciones climáticas son distintas, y se han ejercido presiones de selección diferentes, se podrán identificar marcadores fenotípicos y genotípicos candidatos asociados a diferencias ambientales en las dos regiones geográficas de *Pinus pinceana*.

El objetivo general de este capítulo es detectar diferencias en caracteres morfológicos y genéticos a nivel de expresión o del polimorfismo detectado en los transcriptomas reportados para *P. pinceana* (Capitulo I), que sean de utilidad en en estudios de adaptación local que permitan inferir cambios adaptativos en relación a la variación

ambiental. Los objetivos particulares son, i) determinar si las condiciones climáticas son diferentes a lo largo de la distribución de *P. pinceana*, si el perfil climático es distinto y si ésta diferenciación afecta el área potencial de distribución, ii) identificar si existen cambios fenotípicos en la morfología foliar y en características dasometricas de los individuos (*e.g.,* grosor de corteza, diámetro y altura de los individuos) entre las dos zonas de estudio de la especie, iii) detectar marcadores genéticos potenciales relacionados a respuestas adaptativas (EST's, SSR's y SNP's) a partir de polimorfismos en las secuencias y iv) reconocer en los perfiles de expresión cambios asociados al gradiente ambiental.

Metodología

Caracterización ambiental de las regiones de la distribución

Para caracterizar el ambiente en el que se distribuye *P. pinceana* se obtuvieron los perfiles climáticos de las dos localidades muestreadas, a partir de los datos de precipitación media y temperatura registrados en las estaciones meteorológicas automáticas (EMAS) de la Comisión Nacional del Agua (CONAGUA) y los registros de Climate-Data (http://es.climate-data.org/).

Con el objetivo de reconocer el efecto y cambio de las variables climáticas se elaboró el modelo de distribución potencial con el programa MAXENT ver. 3.2.19 (Phillips *et al*., 2006) con el uso de las georreferencias de los 57 localidades de *P. pinceana* del Herbario Nacional del Instituto de Biología, (UNAM). Para la reconstrucción se consideraron las capas ambientales de WorldClim con la información de precipitación y temperatura de diferentes temporadas del año (BIO #1, #8, #9, #11, #14 Y #15; Hijmans *et al*., 2005) una capa con el mapeo de un índice de estrés hídrico del periodo de sequía marzo-junio (Trabucco, 2010) y la información de los tipos de suelo (CONABIO, 2008).

A partir de los valores de las características ambientales (precipitación, temperatura y estrés hídrico) en los sitios de presencia, se elaboró una base de datos con el programa de información geográfica ArcView 3.2 y el paquete de herramientas *ArcView Projection Utility* y GARP (ver. 1.0, 1999). Debido a que la caracterización de suelos es una categorización cualitativa se decidió eliminar esta variable para el procesamiento estadístico.

Con los datos cuantitativos correspondientes a las localidades de distribución se elaboró un análisis de componentes principales con el paquete estadístico JMP 8 (SAS Institute, 2014) para reconocer la varianza y el efecto de cada una de las características ambientales y así evaluar si existe agrupación y diferenciación entre las variables

Detección de cambios fenotípicos

Se consideraron dos aproximaciones para identificar cambios morfológicos que potencialmente indiquen la respuesta a cambios ambientales. La primera consistió en ubicar y caracterizar la morfología foliar y los factores que posiblemente pudieran estar relacionados a la regulación fisiológica de la evapotranspiración en condiciones áridas en las hojas; y la segunda fue un análisis estadístico de caracteres dasometricos que podrían ser indicadores del crecimiento y desarrollo de individuos en su ambiente.

La caracterización de la organización de la anatomía foliar se visualizó la distribución de estomas y la proporción de ceras en las acículas con técnicas de Microscopía Electrónica de Barrido (MEB). Se seleccionaron fragmentos de un centímetro de la región media de la hoja en cuatro individuos adultos de cuatro poblaciones: Sierra de Parras, Tamaulipas, El Palmito, Durango, Maguey Verde, Querétaro y Antena Núñez, San Luis Potosí, que en conjunto representan los extremos de toda la distribución de *P. pinceana*, (Figura II.1; Tabla II.1). En total para cada tratamiento se analizaron 32 muestras de las que se hicieron registros fotográficos a un aumento de 500nm.

Figura II.1. Mapa de las localidades colectadas para el análisis cambios en la morfología foliar

Tabla II.1. Georeferencias de las localidades que fueron muestreadas para el análisis micrográfico

Localidad	Latitud	Longitud	Altitud
Sierra de Parras Gral. Cepeda, Coahuila	25.433	-102.02	2450
El Palmito Hidalgo, Durango	25.741	-104.881	2000
Maguey Verde Pinal de amoles, Querétaro	21.116	-99.666	2300
Antena Núñez Guadalcázar, San Luis Potosí	22.683	-100.483	1900

El procesamiento del tejido pasó por deshidratación en cambios sucesivos de etanoles [70, 96 y 100%]. Para mejorar la visualización del arreglo de los estomas se utilizaron replicados de todas las muestras con un tratamiento para la eliminación de ceras, que consistió en la incubación en Xileno, seguida de una fijación en etanol absoluto (Reséndiz-Arias, 2014).

Posteriormente, utilizando el servicio e instalaciones del Laboratorio de Microscopía del Instituto de Biología, UNAM, las 32 muestras se secaron hasta el punto crítico, se montaron sobre cinta de carbono, utilizando tres vistas; abaxial, adaxial y transversal por cada muestra, seguida de una cobertura por oro. De cada muestra analizada se recopilaron fotografías, se cuantificó el número de filas que conforman los estomas en la superficie de la epidermis en la proporción abaxial (inferior) y adaxial (superior), y de manera cualitativa se apreció el grosor y abundancia del recubrimiento ceroso.

El análisis estadístico de componentes principales (PCA) de los 93 registros dasometricos del Inventario Forestal **(**CONAFOR, 2012) de los caracteres morfológicos; diámetro a la altura del pecho (DAP), fuste limpio y diámetro de copa, con el objetivo de identificar posibles efectos de la posición geográfica sobre la varianza y agrupación de las características fenotípicas. El PCA fue realizado con el paquete estadístico JMP 8 (SAS Institute, 2014).

Detección de cambios a nivel genético

Identificación de polimorfismos genéticos

La asociación de los cambios y el reconocimiento de la función de los polimorfismos a partir del análisis de los 35,966 genes transcritos que fueron identificados en la caracterización del transcriptoma de *P. pinceana,* se realizó con la detección de cambios

de un solo nucleótido (por sus siglas en ingles *single nucleotide polymorfisim*; SNP), a partir de la detección de inserciones, sustituciones y/o deleciones .

Este procedimiento se realizó con un mapeo de los *reads* procedentes de cada librería y un ensamble consenso de referencia (la descripción y procesamiento de éste se describe en el capítulo 1), construido con los programas Bowtie2-TopHat (Langmead *et al.*, 2009; Kim-Salzberg, 2008).

Las posiciones de cada polimorfismo encontrado se detectaron con el programa Freebayes (Garrison, 2012) y se definieron de acuerdo a la caracterización de la anotación producto del análisis de los transcritos en enTAP (Ver capítulo 1). Se detectó el tipo de cambio (sinónimo o no sinónimo) a partir de la comparación codón por codón de las secuencias traducidas a aminoácidos. La cobertura mínima para la detección de los polimorfismos fue fijada en dos copias distintas y dada la composición de tejido en la muestra no se estableció la ploidia.

Análisis de expresión de transcritos

Para estimar el nivel de expresión de los transcritos que fueron amplificados se consideró la abundancia de los transcritos en cada librería sobre un mapa de las lecturas en fragmentos, reconstruido con los programas Bowtie2 y TopHat (Langmead, 2009; Kim-Salzberg, 2008).

La evaluación de las diferencias de expresión se llevó a cabo a partir del análisis comparativo pareado entre la desviación estándar de la abundancia del transcrito de las dos muestras y el valor medio de expresión de la isoforma en cada muestra, con el objetivo de identificar y mapear la expresión y reconocer los valores contrastantes entre las dos muestras. El valor de abundancia total de cada uno de los fragmentos se obtuvo con el programa eXpress (Roberts *et al.*, 2011) y las comparaciones se efectuaron con la paqueterías de R, EDGE-pro (Magoc *et al.*, 2013) y edgeR (Robinson *et al.*, 2010)

Resultados

Caracterización ambiental de las regiones de distribución

Perfil climático

La obtención de perfiles climáticos regionales correspondió a los datos disponibles en las estaciones meteorológicas más cercanas (http://es.climate-data.org); Peñamiller ubicada a 14 km de Maguey Verde, Queretaro y la estación en Huizache a 15 km de Guadalcazar, San Luis Potosí.

En particular Peñamiller está clasificada con clima Csb, templado seco, la temperatura media anual es de 15.3°C con una precipitación promedio de 1044 mm al año. La diferencia en la precipitación entre el mes más seco y el mes más lluvioso es de 198 mm con temperaturas medias que varían durante el año en un 6.9°C (Figura II.2a). Mientras que para la estación en Huizache está clasificada con un clima BSh, semiárido cálido, con una temperatura media anual de 20.4°C y precipitación de 384mm al año. En donde la diferencia en la precipitación entre el mes más seco y el mes más lluvioso es de 58mm, con un cambio en temperaturas medias durante el año en un 9.2 °C (Figura II.2b).

Los perfiles climáticos de las dos zonas que fueron comparadas coinciden en tener dos periodos en el año con abundante precipitación. La comparación entre las temperaturas de los sitios de muestreo no resulta tan diferente, sin embargo al incluir en la comparación a la localidad El Palmito, ubicada en el extremo Noroeste de la distribución se observan diferencias en las temperaturas (Figura II.2c).

Con pruebas de kolmogorov-smirnov se evaluaó que no existía una diferenciación estadística significativa (p=0.8) entre la precipitación o temperatura entre las estaciones meteorológicas de Peñamiller y Huizache.

Modelo de distribución potencial

El área de distribución modelada abarcó hasta el valor probabilístico más bajo que coincidiera con una localidad conocida (0.5). Las variables con mayor efecto sobre la reconstrucción fueron; las características del suelo, la temperatura media anual y la precipitación por estación.

El modelo abarca en buena proporción el área de distribución conocida e incorpora a todos los sitios registrados en el Herbario Nacional (MEXU), reflejando la fragmentación del hábitat y la escasez de condiciones favorables para la distribución de la especie (Figura II.3).

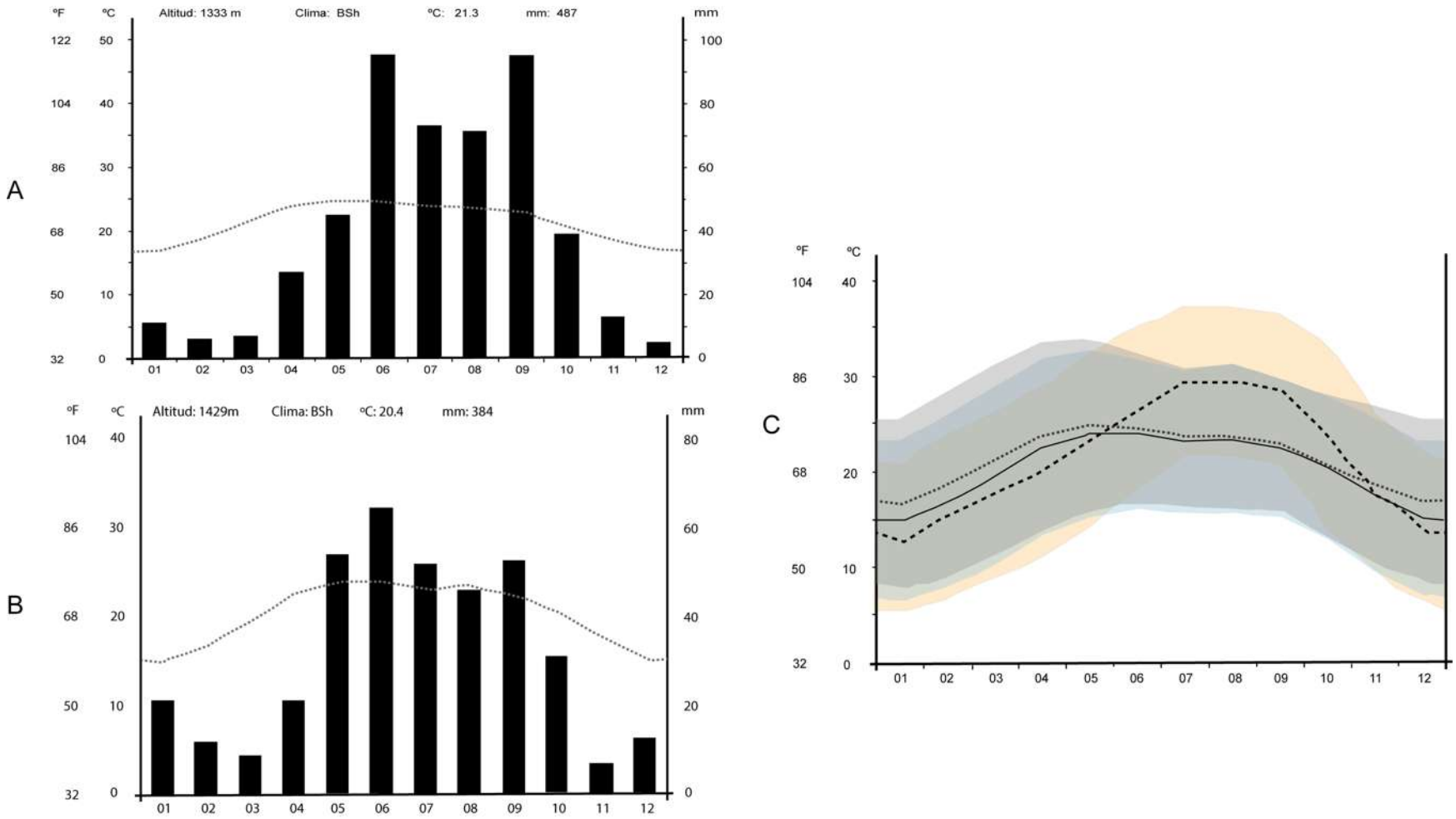


Figura II.2. Climogramas de las regiones cercanas a las localidades muestreadas. A. Corresponde a Maguey Verde B. A Guadalcazar. C. Gradiente de precipitación comparando las diferencias de precipitación entre las localidades muestreadas, en gris, y en naranja, corresponde a General Cepeda en el extremo Norte de la distribución.

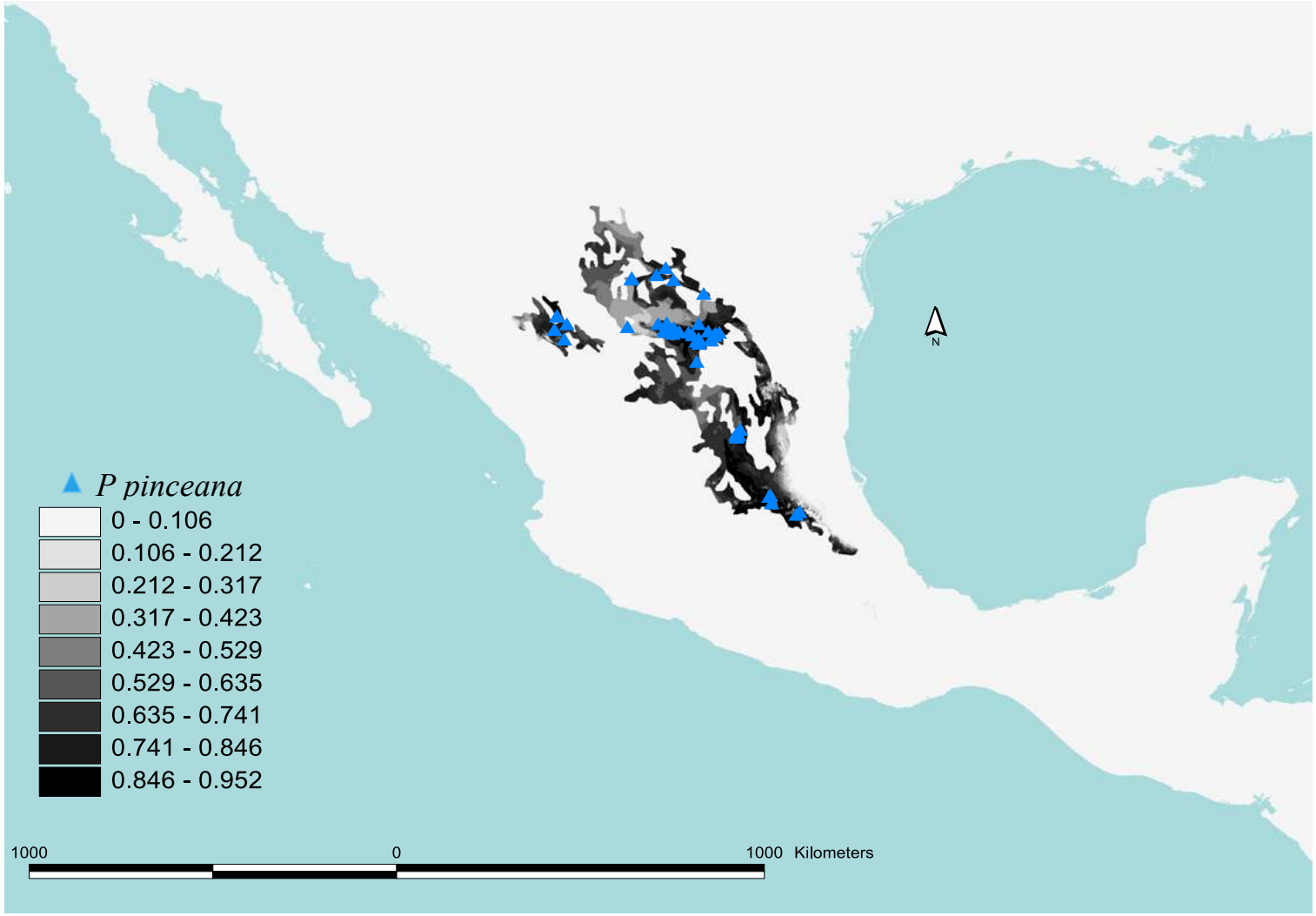


Figura II.3. Modelo de distribución potencial, en gradiente de grises se ilustra las probabilidades arrojadas por el modelo, y en triángulos azules los sitios de presencia de *P. pinceana*.

Con los valores correspondientes a las variables utilizadas en el modelo en cada una de las localidades, se realizó un análisis de componentes principales en donde la aridez, la temperatura del semestre más húmedo y la precipitación anual, explican la mayor proporción (58%) de la varianza de los datos, siendo la aridez la que aporta más a la explicación de la varianza (Figura II.4).

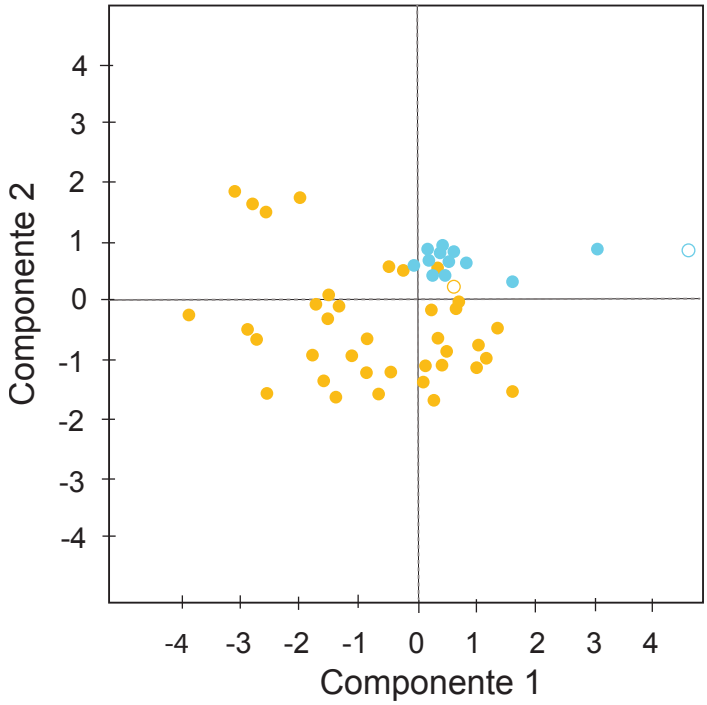


Figura II.4. Gráfico de componentes principales con la dispersión de las variables climáticas en las localidades conocidas. En azul las localidades correspondientes a la grupo genético del Sur y en amarillo las correspondientes al grupo genético del Norte. Los simbolos en donde sobresale el contorno son las localidades que fueron muestreadas para el procesamiento transcriptómico.

El componente principal dos es explicado principalmente por la temperatura media anual, la temperatura en el semestre más seco y la precipitación en el mes más seco, explicando el 16% de la varianza total.

Detección de cambios fenotípicos

Caracterización de la anatomía foliar

Con el tratamiento para cualificar la cantidad de ceras se encontró que en los individuos de la localidad del Sur (Maguey Verde) hay una mayor proporción y grosor de ceras en comparación con el resto de las localidades. Mientras que la localidad Noroeste (El Palmito) se encontró una cubierta cerosa más delgada.

Los individuos de la localidad Sur presentan arreglos estomáticos adaxiales en cuatro líneas, mientras que en la cara abaxial los estomas están ausentes o presentes arreglados en dos líneas. Este patrón abaxial también se encontró en las muestras de la población Noroeste. En general, en la superficie adaxial de las muestras de las localidades Norte se apreció mayor abundancia de estomas arreglados en cuatro y hasta siete líneas (Tabla II.2; Figuras II.5 y II.6).

Se obtuvieron correlación no significativas entre el número de filas en ambas superficies de la acícula, (r^2= 0.10 para la supercie adaxial, y r^2=0.16 para la superficie abaxial). Por tanto, con la muestra analizada la dispersión de los datos no permite establecer una relación estadística significativa entre las variables.

Tabla II.2. Síntesis de la descripción de los caracteres morfológicos de las acículas analizadas.

		Adaxial		Abaxial		
		# filas	# estomas*	# filas	# estomas*	Ceras
Norte	Sierra de Parras	4	6	2	6.5	++
		5	5.8	2	5.5	
		4	6.2	2	5.5	
	Palmito	4	5.2	2	4.5	+
		6	5.1	-	-	
		6	4.3	-	-	
		6	4	-	-	
	Guadalcazar	7	5.1	1	5.5	++
		6	5.5	2	6	
		4	5	-	-	
		4	5.7	2	6	
		4	5.5	1	6	
Sur	Maguey Verde	4	6.1	-	-	++++
		4	5.7	-	-	
		4	6	-	-	
		4	6.2	2	5.5	

*Promedio de estomas por fila en 500nm

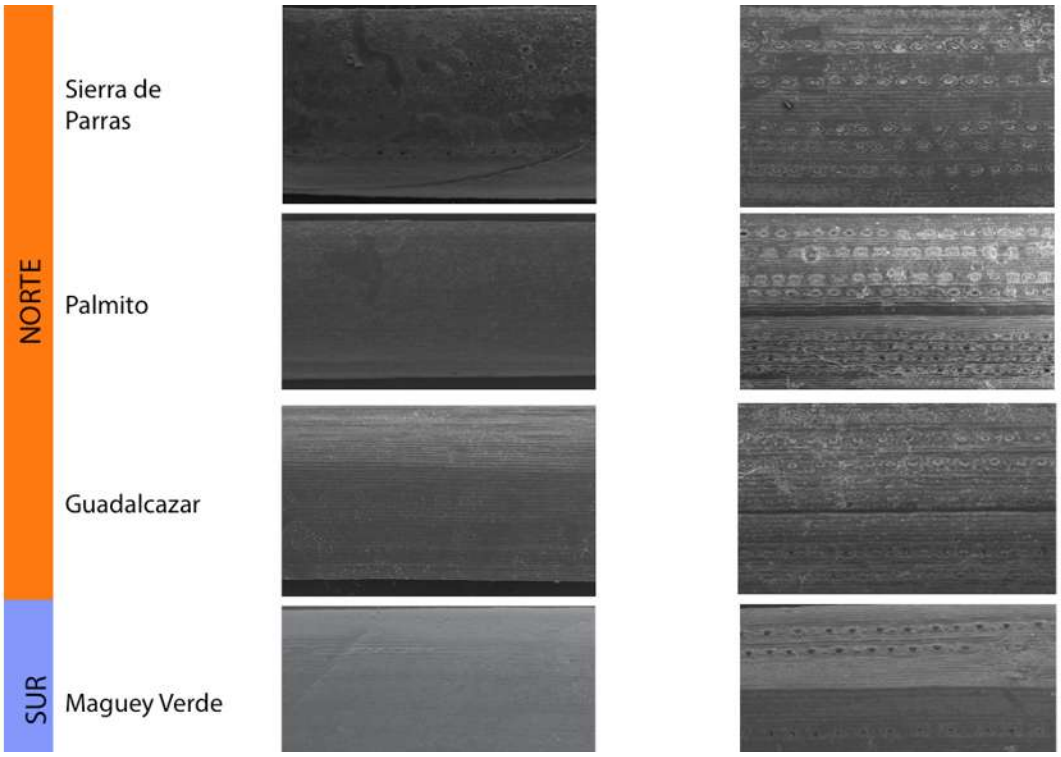


Figura II.5. Micrografías de las acículas analizadas para la visualización de la cubierta cerosa.

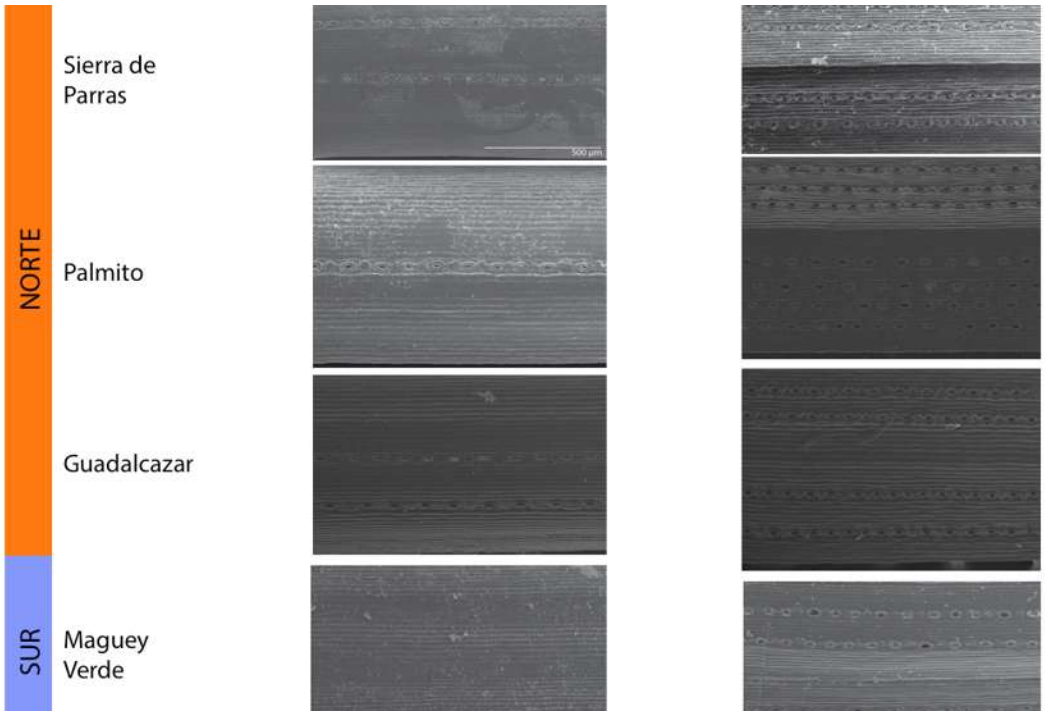


Figura II.6. Micrografías de las acículas analizadas para la visualización del arreglo de los estomas en la acícula

Caracterización morfométrica

Con los datos dasometricos disponibles en el Inventario Forestal **(**CONAFOR, 2012) se encontraron 93 registros, donde 91 corresponden a las localidades de la región Norte y sólo dos a las localidades ubicadas en la parte Sur de la distribución de la especie. Para el análisis estadístico sólo se incluyeron los datos de diámetro a la altura del pecho, la altura total y el diámetro de copa.

En el análisis de componentes principales se encuentran correlacionadas en el componente principal 1, el diámetro a la altura del pecho y la altura total explican el 67.6% de la varianza muestral, lo que sugiere que el individuo al alcanzar mayor altura también tendrá un grosor (DAP) mayor (Figura II.7). En el componente principal dos el diámetro de copa explica el 23.7% de la varianza.

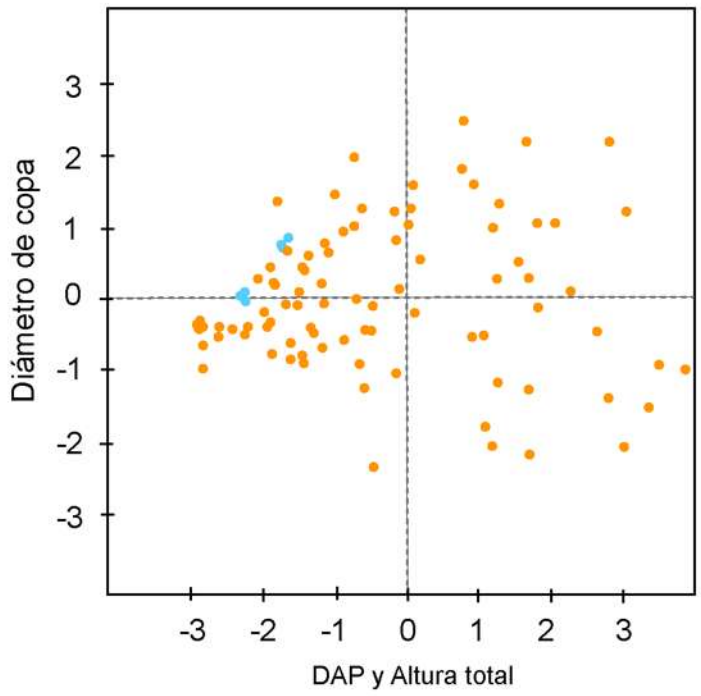


Figura II.7. Gráfica de dispersión con el análisis de componentes principales de los datos anatómicos recabados en el inventario forestal. En amarillo se muestran los individuos muestreados correspondientes al grupo genético del Note, en Azul los individuos de la región Sur.

Diferenciación genética entre las regiones geográficas de distribución

Detección de SNP's

Se identificaron 34 polimorfismos, de los cuales 28 (82.3%) corresponden a transcritos no caracterizados previamente en la base de datos del GenBank (NCBI) y 17 a transcritos descritos; donde se incluyen 3 genes predichos, 3 genes identificados que están involucran los procesos biológicos de acciones perirribosomales, el crecimiento inicial del meristemo y actividades de fotorespiración (Tabla II.3).

A partir de cambio aminoácidico, se encontraron cuatro cambios sinónimos y dos cambios no sinónimos.

En la identificación de polimorfismos se hace importante cotejar, y revalidar los cambios a nivel fisiológico, así como distinguir la frecuencia poblacional de este cambio entre los ambientes en los que se distribuye.

Tabla II.3. Descripción de polimorfismos encontrados

Gen/ Función	#	Posición	Cambio	
Función desconocida	28			
Periribosomal	1	381	C/G	Grl/Trh
Fotorespiración	1	261	A/T	Sinónimo
Succinato deshidrogenasa	1	159	G/T	Sinónimo
Inc. Meristemo	1	423	A/T	Grl/ Csy
Ubiquina	1	500	C/G	Sinónimo
S-adenosilmetionina sintasa	1	72	T/A	Sinónimo

Análisis de expresión

Con los 126 millones de *reads* obtenidos del filtrado y control de calidad (ver Capítulo 1) se encontraron 421 genes con DE (*differentially expressed*) a partir de los cuales fue posible reconstruir la agrupación jerárquica entre los módulos (Dendrograma) y el mapa de calor (*heatmap*) para visualizar los bloques de cambios entre las dos muestras en función de su abundancia de transcripción (Figura II.8).

Entre los 421 genes con expresión diferencial, 282 (66.9%) corresponden a fragmentos sin caracterización, 48 (11.40%) son genes predichos de los cuales cinco están involucrados en el crecimiento y la germinación, esta diferenciación se atribuye a las variación en la composición de tejido en la muestra, 70 (16.6%) factores de regulación de la transcripción y del metabolismo basal, y sólo 10 (2.37%) están involucrados en

actividades están relacionadas a la termorregulación, estrés hídrico y defensa de patógenos.

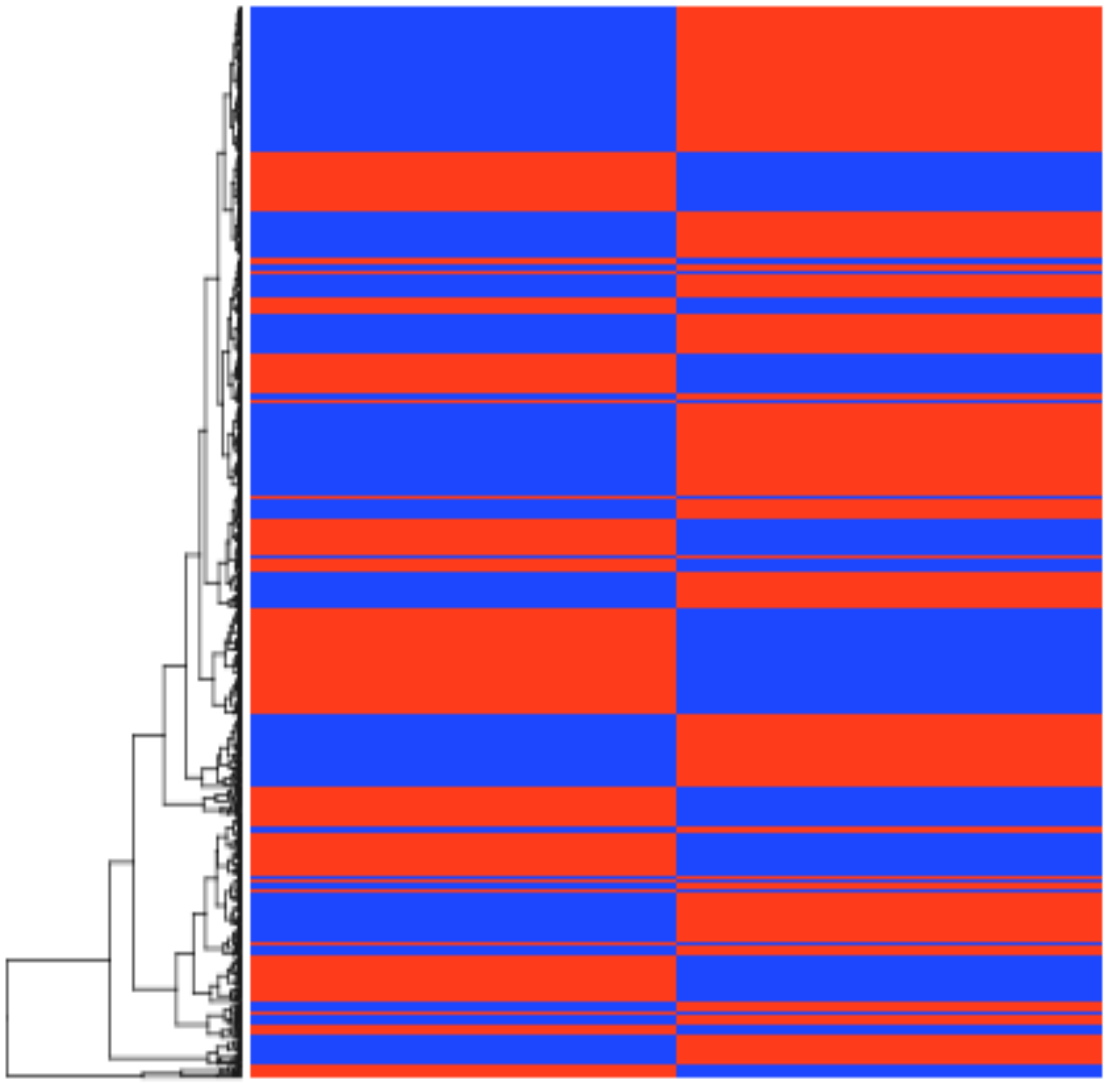


Figura II.8. *Heatmap* comparando los contigs expresados diferencialmente, el color indica la expresión.

Discusión

En este trabajo se logró identificar marcadores candidatos potenciales, genéticos y morfológicos, que podrían estar involucrados en respuestas adaptativas a la aridez y en la defensa a patógenos. La integración e identificación de cambios explora preliminarmente la dinámica genotipo-fenotipo-ambiente.

Para el análisis de las variables climáticas en las zonas de colecta (Maguey Verde y Antena Núñez) los perfiles climáticos de las zonas muestran sitios áridos con dos periodos anuales de lluvia abundante. Estas localidades no muestran cambios contrastantes, en precipitación o temperatura, esto puede explicarse porque las condiciones de los sitios de muestreo y las estaciones meteorológicas (altitud, temperatura y presión) no son los mismos que en los sitios de muestreo. Pero al comparar localidades del extremo Norte de la distribución se aprecia que el intervalo termodinámico donde ésta conífera se distribuye va de condiciones semiáridas a más secas.

En el modelo de distribución potencial se logró distinguir que las condiciones propicias para la especie no son abundantes en el país y que *P. pinceana* se encuentra restringida debido a las características del suelo, la temperatura y la precipitación, lo que sugiriere que en la distribución actual se conjuntan efectos climáticos e históricos-demográficos. El análisis de componentes principales permite distinguir en los extremos de la dispersión de los datos, que las condiciones opuestas de temperatura y precipitación corresponden a los extremos geográficos de la distribución.

La variación morfológica responde al mantenimiento homeostático y al efecto de los factores bióticos y abióticos del ambiente (Hart *et al.,* 2000, Smith, 2011, St. Clair *et al.,* 2009). El variación del arreglo en los estomas que observamos en los resultados, puede ser explicado por tres escenarios plausibles; respuestas plásticas, adaptativas y variaciones ontogénicas (cambios morfológicos a lo largo de los estadios del organismo; Coleman, 1994).

Smith (2011) señala que los rasgos que varían significativamente entre poblaciones y que se encuentras asociados con el genotipo son el crecimiento (Oksanen *et al.,* 2001), la morfología foliar (Barnes, 1975), y rasgos fenológicos (Yu *et al.,* 2001). Así, incluir indicadores directos o indirectos de estos caracteres podrían puntualizar con mayor precisión cambios entre las regiones de distribución.

El análisis estadístico con los datos dasometricos del Inventario Forestal no sugieren una correlación entre las variables fenotípicas y su procedencia geográfica, se observa en

cambio, que existe una gran variabilidad y dispersión de los caracteres morfológicos,lo que puede deberse al manejo, crecimiento y condición de los individuos muestreados. Haciendo que los cambios morfológicos no sean observable en este nivel, con estas características en esta muestra. Los datos del Inventario Forestal si bien aportan generalidades sobre las características morfológicas de la especie, la muestra disponible no abarca toda la distribución de la especie, y no se dispone de todos los datos para todas las poblaciones.

Para este estudio se encontraron cambios en la distribución y abundancia de los estomas que sugieren, según la muestra analizada, que el arreglo de los estomas cambia y podría relacionarse al cambio en las condiciones climáticas. En la literatura (Salisbury, 1927; Tichá, 1982; Woodward, 1987; Hetherington, 2003; Chaerle *et al.,* 2005; Israelsson *et al.*, 2006) se han documentado que los cambios de la morfología foliar en la frecuencia de estomas y la formación de ceras se ven afectados por los cambios en la temperatura atmosférica, la regulación metabólica del consumo de CO_2 y la pérdida de agua por efecto de factores ambientales. Mientras que el diferencial entre las superficies foliares, con mayor abundancia de estomas en la parte abaxial (inferior) que en la superficie adaxial (superior) ha sido explicado por Limm *et al.,* (2009) como una respuesta para evitar la saturación de agua cuando llueve.

Sin embargo, existen investigaciones en angiospermas donde se ha corroborado la plasticidad de este carácter (Clifford *et al.,* 1995; Salisbury, 1927; Quarrie y Jones, 1977). Por ejemplo en condiciones diversas de humedad y estrés hídrico se ha reportado que existe reducción del número de estomas en *Caltha palustris* (Ranunculaceae), *Triticum spp.* (Poaceae; trigo) y *Arachis hypogaea* (Fabaceae; cacahuate), mientras que con el aumento de la humedad no generó una reducción de estomas en *Scilla nutans* (Asparagaceae). Por su parte, MacDonald, (2002) y Kouwenberg *et al*., (2003) han documentado que pueden haber reducciones en la densidad de los estomas cuando se cultiva en niveles elevados de CO_2.

Las diferencias en el arreglo de los estomas no sugiere una correlación significativa con las condiciones de aridez. Sin embargo, se requiere para corroborar la relación de estos caracteres con un análisis con una muestra más grande que permita validar las diferencias que integre la la variablidad de toda la distribución de la especie e incluya la variación en el número de estomas y la cantidad de ceras.

Por su parte, en la detección de SNP's refleja muy pocas diferencias entre los dos sitios muestreados, esto puede deberse a dos factores: 1) que los polimorfismos se deban a

las anotaciones dada la calidad de secuenciación y cobertura, *ie.*, sea un efecto del método y diseño experimental que se siguió, la composición de la muestra y diversidad de fragmentos o, 2) que la escases de anotaciones producidas y disponibles reduzcan naturalmente la proporción de información disponible.

Si bien los seis cambios identificados reflejan una pauta de cambios ecológicos y fisiológicos importantes por lo cual se hace necesario corroborar bioquímicamente del efecto de las isoformas que están involucradas en la fotorespiración y la germinación, evaluar la frecuencia de cada isoforma por las regiones de distribución.

La expresión de transcritos en este trabajo fue interpretada como el efecto del genotipo a los estímulos y condiciones ambientales; pero es importante reconocer aspectos de la biología de la especie, por ejemplo, factores fisiológicos y/o reproductivos, características demográficas y de la historia evolutiva (Rajkumar *et al.*, 2015).

Los cambios en la expresión detectados marcan una pauta general sobre las diferencias en respuestas genéticas para regular y responder fisiológica y metabólicamente ante distintas condiciones. Savolainen *et al.,* (2007) señala que los ejercicios comparativos entre poblaciones o especies promueven mejor comprensión de la base genética para entender los mecanismos evolutivos de la adaptación, al explorar con mayor resolución los efectos fenotípicos múltiples o de múltiples loci.

Los DE arrojados por el análisis de cambios de expresión son marcadores potenciales para enfocar un estudio a nivel de tejido específico, comparando cuantitativamente el efecto de estas diferencias, y los blancos fisiológicos donde repercuten estos cambios. En este trabajo el análisis de expresión no considera la estructura y diversidad de tejido en cada muestra, además de que carece de replicas por lo que es necesario dar continuidad al estudio para corroborar la detección de polimorfismos y revisión de cambios en los niveles de expresión, con una muestra significativa con replicas, distintos tipos de tejido y tratamientos control.

Los análisis hasta estos resultados especulan y dan una pauta para estudios futuros donde puedan ser considerados como posibles genes candidatos asociados a cambios adaptativos. Y permitan explicar el componente genético de la variación fenotípica (heredabilidad, *h*) para enfocar los estudios de asociación hacia el entendimiento cuantitativo de los mecanismos de Selección Natural.

Conclusiones

Se asocio la variación genética y fenotípica en dos localidades donde se distribuye *Pinus pinceana,* logrando establecer relaciones entre estos cambios y la variación climática que existe a lo largo de la distribución. Ésta aproximación parte de la integración de datos morfológicos y genéticos para la predicción de marcadores potenciales adaptativos.

La distribución de *P. pinceana* está fuertemente restringida por las características del suelo, la temperatura y precipitación. Si bien en los puntos de muestreo no se aprecia el contraste climático, en el análisis estadístico se observa un gradiente ambiental que aumenta las condiciones áridas en la proporción Norte.

La varianza morfológica en caracteres dasométricos no refleja un patrón asociado a las diferencias climáticas a partir de la muestra disponible. Cuantificar caracteres de crecimiento neto y desarrollo podrían acotar e identificar niveles de variación.

Respecto a la variación en la organización de la morfología foliar se distinguen diferencias en las cubiertas cerosas y en la organización de los estomas en las caras de la hoja que en la literatura se atribuyen a los cambios climáticos en el ambiente.

La variación encontrada en los loci expresados diferencialmente sugieren un potencial adaptativo, además de que en ambos sitios se emplean estrategias distintas de resistencia a patógenos. Sin embargo, es necesario corroborar su participación en la adaptación a partir de un nuevo esquema de diseño experimental con una evaluación con un enfoque poblacional.

Se identificaron cambios genéticos que están involucrados a procesos metabólicos y fisiológicos. Sin embargo se hace importante revalidar y cotejar los cambos a nivel fisiológico así como distinguir la frecuencia poblacional de estos polimorfismos entre los ambientes en los que se distribuye la especie.

Conclusiones generales

Se reconocieron en el transcriptoma de *P. pinceana* 47,510 genes transcritos, de los cuales se encontró la descripción y función de 35,916.

Este trabajo ejemplifica el alcance y el rendimiento de las técnicas de NGS, para entender las respuestas genéticas activas, reconocer la diversidad genética en genomas grandes y esbosar las relaciones entre el genotipo, el fenotipo y el ambiente.

La caracterización del transcriptoma de una conífera depende tanto de la calidad y la diversidad de la muestra, así como del avance tecnológico y del reconocimiento de genes de especies de grupos filogenéticos cercanos.

El mapeo de rutas de metabólicas facilita la comprensión de funciones, formas de regulación y respuestas ante estímulos y/o condiciones especificas. Se encontraron así, promotores e intermediarios de rutas de fitoalexinas ruta que no ha sido bien caracterizada en coníferas.

La distribución de *P. pinceana* está restringida por las características de la temperatura y precipitación, aspectos que permiten plantear estudios que evidencien las repercusiones de los cambios abientales dado que existe un gradiente ambiental que aumenta las condiciones aridas en el extremo Norte de la distribución.

En la organización de la morfología foliar hay diferencias en las cubiertas cerosas y en la organización de los estomas. Siendo que la abundancia en ceras y estomas en la cara adaxial aumenta al mismo tiempo que las condiciones aridas. Sin embargo en este estudio no se corroboro estadísticamente esta relación, pero para otras especies en la literatura se atribuyen como una respuesta a cambios climáticos en el ambiente.

La variación genética, a partir de polimorfismos y expresión diferenciales, sugieren que se expresan distintos mecanismos de resistencia a patógenos, control hídrico y termorregulación que reflejan una pauta de cambios ecológicos y fisiológicos. No obstante se hace importante revalidar y cotejar los cambos a nivel fisiológico además distinguir su frecuencia poblacional.

Perspectivas

Las bases de este trabajo permiten dar seguimiento al planteamiento de un estudio de asociación, que procure una comparación poblacional de los cambios morfológicos y genéticos que se destacan en este trabajo, que permita explorar alternativas para reconocer si *P. pinceana* ha modulado un proceso de adaptación a la diversidad de su ambiente, que pueda descartar los efectos de la plasticidad de los caracteres y los efectos epigenéticos.

Para los cambios genéticos se sugieren las siguientes consideraciones: (1) Determinar el efecto estructural y fisiológico del cambio de los genes candidatos, (2) Validar la frecuencia de los cambios genéticos en tejido haploide (megagametofítico), para descartar falsos positivos debidos a cambios heterocigos en tejido diploide, (3) Evaluar si hay una efecto sinténico (multilocus y/o poligénicos) dentro de los polimorfismos, (4) Analizar los cambios de expresión en muestras en tejidos específicos, estadios de crecimiento para reconocer dinámicas y los patrones específicos de expresión, (5) Correlacionar los loci, mutaciones y cambios de expresión a los fenotipos específicos.

Procurar consolidar un enfoque cuantitativo que permita analizar la heredabilidad (*h*) de los rasgos fenotípicos, considerando: (1) Una evaluación morfométrica de caracteres foliares con una muestra más grande, que procure analizar si existe una variación individual y diferencias poblacionales, (2) La elaboración de trasplantes recíprocos con el objetivo de evaluar la plasticidad fenotípica, (3) Cuantificar las diferencias en los factores abióticos y cambios altitudinales en las zonas de distribución, que se han señalado en la literatura como determinantes para la organización de la anatomía foliar.

Glosario

BLAST [*Basic Local Alignment Search Tool*]: Algoritmo computacional que realiza comparaciones entre secuencias permite distinguir a partir de la similitudes, relaciones funcionales y filogenéticas.

Cobertura: parámetro para la cuantificar el número promedio de lecturas secuenciadas por locus, también se reconoce como profundidad de secuenciación. Puede ser calculada por la longitud acumulada de *reads* o *pair reads* como múltiplo del tamaño del genoma.

Contig: Construido a partir de las superposiciones fragmentos de secuencias (*reads*).

Ensamble: Algoritmo de superposición por consenso estructural o de gráficos de Bruijn de contigs para generar secuencias consenso.

Ensamble *De novo*: Tipo de ensamble que no utiliza un mapping o secuencias de referencia.

Etiquetado: Caracterización de las secuencias en formas genéticas, genes o isoformas con información funcional y descriptiva.

EST [*Expressed Sequence Tag*]: secuencia de transcritos específicos.

***Gene Onthology* (GO):** Base de datos clasificada a partir de las características funcionales, entre las especies y la funcionalidad de los genes.

Inserto: fragmentos cortados al azar (desde el genoma o transcriptoma) secuenciados.

Librería: Conjunto de fragmentos de DNA o RNA.

mRNA: RNA mensajero

Mapping: alineamiento de reads con un genoma de referencia.

N50: Estadística de un conjunto de contigs (scaffolds) reconocida como la longitud para la cual la colección de todos los contigs de longitud o más largo que contiene al menos la mitad del total de las longitudes de los contigs

ORF [*Open Reading Frame*] marco de lectura abierto

Pair-ends: reads pareados de secuenciación que derivan de l secuenciación por ambos extremos de la misma molécula de cDNA.

Reads: secuencia producida por la secuenciación una molécula de cDNA.

RIN: índice de Integridad del RNA

rRNA: RNA ribosomal

Script: Archivo que enlista los comandos que se requieren para ejecutar un programa escritas en un determinado lenguaje de programación.

SNP: Single Nucleotide Polymorphism. Una única base de diferencia se encontró al comparar la misma secuencia de ADN de dos individuos diferentes.

Scaffold: unión de contigs que no se superponen ordenados a partir de los extremos.

Referencias

Altschul *et al.,* 1990 Basic local alignment search tool. *J Mol Bio* 215:403-10.

Aitken, S.N. *et al.,* 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, 1(1), pp.95–111.

Aitken, S.N., Kavanagh, K.L. & Yoder, B.J., 1995. Genetic variation in seedling water-use efficiency as estimated by carbon isotope ratios and its relationship to sapling growth in Douglas-fir. *For. Genet*, 2, pp.199–206.

Baltunis, B.S. *et al.,* 2008. Inheritance of foliar stable carbon isotope discrimination and third-year height in Pinus taeda clones on contrasting sites in Florida and Georgia. *Tree Genetics & Genomes*, 4(4), pp.797–807.

Bao, E., Jiang, T. & Girke, T., 2013. BRANCH: boosting RNA-Seq assemblies with partial or related genomic sequences. *Bioinformatics*, 29(10), pp.1250–1259.

Barnes, B. V, 1975. Phenotypic variation of trembling aspen in western North America. *Forest Science*, 21(3), pp.319–328.

Bennett, M.D. & Linch, 2005. Plant Genome Size Research: A Field In Focus. *Annals of Botany*, 95(1), pp.1–6.

Birol, I. *et al.,* 2013. Assembling the 20 Gb white spruce (Picea glauca) genome from whole-genome shotgun sequencing data. *Bioinformatics (Oxford, England)*, 29(12), pp.1492–7.

Bonello, P. *et al.,* 2006. Nature and ecological implications of pathogen-induced systemic resistance in conifers: A novel hypothesis. *Physiological and Molecular Plant Pathology*, *68*(4–6), 95–104.

Bradshaw, A.D., 1965. Evolutionary significance of phenotypic plasticity in plants. *Advances in genetics*, 13(1), pp.115–155.

Brendel, O. *et al.,* 2002. Genetic parameters and QTL analysis of δ13C and ring width in maritime pine. *Plant, Cell & Environment*, 25(8), pp.945–953.

Canales, J. *et al.*, 2014. *De novo* assembly of maritime pine transcriptome: implications for forest breeding and biotechnology. *Plant Biotechnology Journal*, 12(3), pp.286–299.

Cañas, R. a. *et al.,* 2015. Understanding developmental and adaptive cues in pine through metabolite profiling and co-expression network analysis. *Journal of Experimental Botany*, 66(11), pp.3113–3127.

Chaerle, L., Saibo, N. & Van Der Straeten, D., 2005. Tuning the pores: towards engineering plants for improved water use efficiency. *Trends in biotechnology*, 23(6), pp.308–15.

Chang, S., Puryear, J. & Cairney, J., 1993. A simple and efficient method for isolating RNA from pine trees. *Plant molecular biology reporter*, 11(2), pp.113–116.

Chen, J. *et al.*, 2012. Sequencing of the needle transcriptome from Norway spruce (Picea abies Karst L.) reveals lower substitution rates, but similar selective constraints in gymnosperms and angiosperms. *BMC Genomics*, 13, p.589.

Clifford, S.C. *et al.,* 1995. The effect of elevated atmospheric CO2 and drought onstomatal frequency in groundnut (Arachis hypogaea (L.)). *Journal of Experimental Botany*, 46(7), pp.847–852.

Coleman, J.S., McConnaughay, K.D. & Ackerly, D.D., 1994. Interpreting phenotypic variation in plants. *Trends in ecology & evolution*, 9(5), pp.187–91.

Comisión Nacional para el conocimiento y uso de la Biodiversidad (CONABIO) *Tipos de suelo* modificado por CONABIO. 1999.

Comisión Nacional Forestal (CONAFOR) *Inventario Forestal* 2012.

Condit, R., Hubbell, S.P. & Foster, R.B., 1995. Mortality Rates of 205 Neotropical Tree and Shrub Species and the Impact of a Severe Drought. *Ecological Monographs*, 65(4), pp.419–439

Conesa, A. *et al.,* 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* , 21 (18), pp.3674–3676.

Corbett-Detig, R.B., Hartl, D.L. & Sackton, T.B., 2015. Natural Selection Constrains Neutral Diversity across A Wide Range of Species. *PLOS Biology*, 13(4), p.e1002112.

Cox, M.P., Peterson, D.A. & Biggs, P.J., 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC bioinformatics*, 11(1), p.485.

Cuenca, A., Escalante, A.E. & Piñero, D., 2003. Long-distance colonization, isolation by distance, and historical demography in a relictual Mexican pinyon pine (Pinus nelsonii Shaw) as revealed by paternally inherited genetic markers (cpSSRs). *Molecular Ecology*, 12(8), pp.2087–2097.

Darden, J.R. & Marks, H.L., 1988. Divergent selection for growth in Japanese quail under split and complete nutritional environments. 2. Water and feed intake patterns and abdominal fat and carcass lipid characteristics. *Poultry science*, 67(8), pp.1111–1122.

De La Torre, A.R. *et al.*, 2015. Genome-wide analysis reveals diverged patterns of codon bias, gene expression, and rates of sequence evolution in picea gene families. *Genome biology and evolution*, 7(4), pp.1002–15.

Delgado, P. *et al.,* 1999. High population differentiation and genetic variation in the endangered Mexican pine Pinus rzedowskii (Pinaceae). *American Journal of Botany*, 86(5), pp.669–676.

Eckert, A.J. *et al.,* 2010. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (Pinus taeda L., Pinaceae). *Genetics*, 185(3), pp.969–982.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)*, 26(19), pp.2460–1.

Ekblom, R. & Wolf, J.B.W., 2014. A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, 7(9), p.n/a–n/a.

Endler, J.A., 1977. *Geographic variation, speciation, and clines*, Princeton University Press.

Endler, J.A., 1986. *Natural selection in the wild*, Princeton University Press.

Escalante A. E. *Estructura genética de poblaciones de Pinus pinceana usando como marcadores moleculares microsatelites de cloroplasto (cpssr's*) Tesis de Licenciatura. Universidad Nacional Autónoma de México, Facultad de Ciencias; 2001.

Escaramís, G., Docampo, E. & Rabionet, R., 2015. A decade of structural variants: description, history and methods to detect structural variation. *Briefings in functional genomics*, p.elv014.

Evans, T.G. & Hofmann, G.E., 2012. Defining the limits of physiological plasticity: how gene expression can assess and predict the consequences of ocean change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1596), pp.1733–1745.

Food and Agruculture Organization of the United Nations (FAO), 2002. The arid environments en *Arid zone forestry: A guide for field technicians*.ISBN 92-5-102809-5

Farjon, A., de la Rosa, J.A.P. & Styles, B.T., 1997. *A field guide to the pines of Mexico and Central America.*, Royal Botanic Gardens.

Feder, M.E. & M.-O.T., 2003. Evolutionary and ecological functional genomics. *Nature reviews. Genetics*, 4(August), pp.651–657.

Frichot, E. *et al.,* 2015. Detecting adaptive evolution based on association with ecological gradients: Orientation matters&excl. *Heredity*.

Fu, Y.X. & Li, W.H., 1997. Estimating the age of the common ancestor of a sample of DNA sequences. *Molecular Biology and Evolution*, 14(2), pp.195–199.

enTAP (Eukaryote Non-Model Transcriptome Annotation Pipeline) 2014. *Ver. 2.01*

González-Martínez, S.C. et al., 2006. DNA sequence variation and selection of tag SNPs at candidate genes for drought-stress response in Pinus taeda L. *Genetics*, 172, pp.1915–1926.

González-Medrano F. 2012. Las zonas áridas y semiáridas de México y su vegetación SEMARNAT.

Götz, S., 2015. Blast2GO Revisited: State of the Art in Functional Annotation and Analysis of Non-Model Organisms. In *Plant and Animal Genome XXIII Conference*. Plant and Animal Genome.

Grabherr, M.G. *et al.*, 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, 29(7), pp.644–652.

Gramzow, L., Weilandt, L. & Theißen, G., 2014. MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Annals of botany*, p.mcu066.

Granados V., R. Linx, *et al*., 2015. Caracterización y ordenación de los bosques de pino piñonero (Pinus cembroides subsp. orizabensis) de la Cuenca Oriental (Puebla, Tlaxcala y Veracruz). *Madera y bosques*, *21*(2), 23-43.

Guillet-Claude, C. *et al.,* 2004. The evolutionary implications of knox-I gene duplications in conifers: correlated evidence from phylogeny, gene mapping, and analysis of functional divergence. *Molecular biology and evolution*, 21(12), pp.2232–45.

Guo, B. *et al.,* 2015. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biology*, 13(1), p.19.

Haas, B.J. *et al.,* 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, 8(8), pp.1494–1512.

Hart, M., Hogg, E.H. & Lieffers, V.J., 2000. Enhanced water relations of residual foliage following defoliation in Populus tremuloides. *Canadian Journal of Botany*, 78(5), pp.583–590.

Hammerschmidt, R.L. Nicholson 1999. A survey of plant defense responses to pathogens A.A. Agrawal, S. Tuzun, E. Bent (Eds.), *Induced plant defenses against pathogens and herbivores*, APS Press, St. Paul, MN, pp. 55–71

Han, S. *et al.,* 2015. RNA-Seq analysis for transcriptome assembly, gene identification, and SSR mining in ginkgo (Ginkgo biloba L.). *Tree Genetics & Genomes*, 11(3).

Hetherington, A.M. & Woodward, F.I., 2003. The role of stomata in sensing and driving environmental change. *Nature*, 424(6951), pp.901–908.

Hijmans R. J. C, J. L. Parra, P. G Jones y A. Jarvis. Very high resolution interpolated climates surfaces for global land areas. *International Journal of Climatogy* 2005, 25.

Ibarra-Laclette, E. *et al.,* 2013. Architecture and evolution of a minute plant genome. *Nature*, 498(7452), pp.94–98.

Ingram, J. & Bartels, D., 1996. The molecular basis of dehydration tolerance in plants. *Annual review of plant biology*, 47(1), pp.377–403.

Israelsson, M. *et al.,* 2006. Guard cell ABA and CO 2 signaling network updates and Ca 2+ sensor priming hypothesis. *Current opinion in plant biology*, 9(6), pp.654–663.

Joshi, N.A. & Fass, J.N., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software].

Karhu, A. *et al.,* 2006. Analysis of microsatellite variation in Pinus radiata reveals effects of genetic drift but no recent bottlenecks. *Journal of Evolutionary Biology*, 19(1), pp.167–175.

Karhu, A. et al., 1996. Do molecular markers reflect patterns of differentiation in adaptive traits of conifers? *Theoretical and Applied Genetics*, 93(1-2), pp.215–221.

Karhu, J.A. & Holland, H.D., 1996. Carbon isotopes and the rise of atmospheric oxygen. *Geology*, 24(10), pp.867–870.

Keller, S.R. *et al.,* 2011. Climate-driven local adaptation of ecophysiology and phenology in balsam poplar, Populus balsamifera L.(Salicaceae). *American Journal of Botany*, 98(1), pp.99–108.

Keeling, C.I. *et al.*, 2011. Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (Picea spp.). *BMC plant biology*, 11(1), p.43.

King, M.-C. & Wilson, A.C., 1975. *Evolution at two levels in humans and chimpanzees*, na.

Kouwenberg, L.L.R., Kürschner, W.M. & McElwain, J.C., 2007. Stomatal frequency change over altitudinal gradients: prospects for paleoaltimetry. *Reviews in Mineralogy and Geochemistry*, 66(1), pp.215–241.

Kovach, A. *et al.,* 2010. The Pinus taeda genome is characterized by diverse and highly diverged repetitive sequences. *BMC genomics*, 11(1), p.420.

Kuparinen, A., Savolainen, O. & Schurr, F.M., 2010. Increased mortality can promote evolutionary adaptation of forest trees to climate change. *Forest Ecology and Management*, 259(5), pp.1003–1008.

Labrou, N.E. *et al.,* 2015. Plant GSTome: structure and functional role in xenome network and plant stress response. *Current opinion in biotechnology*, 32, pp.186–194.

Le Houerou, H.N. & DREGNE, H.E., 1970. North Africa: past, present, future. *Publ. Amer. Assoc. Advanc. Sci., No. 90*, pp.227–278.

Langmead, B. *et al.,* 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biol*, 10(3), p.R25.

Ledig, F.T. *et al.,* 2001. Genetic diversity and the mating system of a rare Mexican pinon, Pinus pinceana, and a comparison with Pinus maximartinezii (Pinaceae). *Am. J. Botany*, 88(11), pp.1977–1987.

Life, Earth and Environmental Sciences (LESC) 2012. Conservation Genomics: amalgamation of conservation genetics and ecological and evolutionary genomics (ConGenOmics) Research Networking Programme

Limm, E.B. *et al.,* 2009. Foliar water uptake: a common water acquisition strategy for plants of the redwood forest. *Oecologia*, 161(3), pp.449–459.

Little 1969 Little EL, Critchfield WB (1969) Subdivisions of the genus Pinus(pines). US Dep Agric For Serv Misc Pub11144 Washington D.C.

Lewontin, R.C., 1957. The adaptations of populations to varying environments. In *Cold Spring Harbor Symposia on Quantitative Biology*. Cold Spring Harbor Laboratory Press, pp. 395–408.

McDonald, E.P., Erickson, J.E. & Kruger, E.L., 2002. Research note: Can decreased transpiration limit plant nitrogen acquisition in elevated CO2? *Functional Plant Biology*, 29(9), pp.1115–1120.

Magoc, T., Wood, D. & Salzberg, S.L., 2013. EDGE-pro: estimated degree of gene expression in prokaryotic genomes. *Evolutionary bioinformatics online*, 9, p.127.

Marshall, H.D., Newton, C. & Ritland, K., 2002. Chloroplast phylogeography and evolution of highly polymorphic microsatellites in lodgepole pine (Pinus contorta). *Theoretical and Applied Genetics*, 104, pp.367–378.

Marshall, J.D. & Zhang, J., 1994. Carbon isotope discrimination and water-use efficiency in native plants of the north-central Rockies. *Ecology*, pp.1887–1895.

Martiñón-Martínez, R. J., Vargas-hernández, J. J., López-Upton, J., Gómez-guerrero, A., Vaquera-huerta, (2010)..RESPUESTA DE Pinus pinceana Gordon a estrés por sequía y altas temperaturas. *Fitogenética, S. M.*

Molina-Freaner, F. *et al.,* 2001. Do rare pines need different conservation strategies? Evidence from three Mexican species. *Canadian Journal of Botany*, 79(2), pp.131–138.

Morgante, M. & De Paoli, E., 2011. Toward the conifer genome sequence. *Genetics, Genomics and Breeding of Conifers Trees*, pp.389–403.

Moriya, Y. *et al.,* 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic acids research*, 35(suppl 2), pp.W182–W185.

Nadeau, N.J. *et al.,* 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*, 22(3), pp.814–826.

Nair, P., 2014. Conservation genomics. *Proceedings of the National Academy of Sciences*, 111(2), p.569.

Neale, D.B. *et al.,* 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome biology*, 15(3), p.R59.

Neale, D.B. & Savolainen, O., 2004. Association genetics of complex traits in conifers. *Trends in Plant Science*, 9(7), pp.325–330.

Newton, R.J. *et al.,* 1991. Molecular and physiological genetics of drought tolerance in forest species. *Forest Ecology and Management*, 43(3), pp.225–250.

Niu, S.-H. *et al.,* 2013. Transcriptome characterisation of Pinus tabuliformis and evolution of genes in the Pinus phylogeny. *BMC genomics*, 14(1), p.263.

Nosil, P., Funk, D.J. & ORTIZ-BARRIENTOS, D., 2009. Divergent selection and heterogeneous genomic divergence. *Molecular ecology*, 18(3), pp.375–402.

Nystedt, B. *et al.,* 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497(7451), pp.579–84.

Oksanen, E., Sober, J. & Karnosky, D.F., 2001. Impacts of elevated CO 2 and/or O 3 on leaf ultrastructure of aspen (Populus tremuloides) and birch (Betula papyrifera) in the Aspen FACE experiment. *Environmental Pollution*, 115(3), pp.437–446.

Oleksiak, M.F., Churchill, G.A. & Crawford, D.L., 2002. Variation in gene expression within and among natural populations. *Nature Genetics*, 32(2), pp.261–266.

Olivas-García, J.M., Cregg, B.M. & Hennessey, T.C., 2000. Genotypic variation in carbon isotope discrimination and gas exchange of ponderosa pine seedlings under two levels of water stress. *Canadian journal of forest research*, 30(10), pp.1581–1590.

Parchman, T.L. *et al.,* 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC genomics*, 11(1), p.180.

Paudel, Y. *et al.,* 2015. Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. *BMC genomics*, 16(1), p.330.

Pavy, N. *et al.*, 2008. Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC genomics*, 9(1), p.21.

Pavy, N. *et al.,* 2012. A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. *BMC Biology*, 10(1), p.84.

Pespeni, M.H. *et al.,* 2013. Evolutionary change during experimental ocean acidification. *Proceedings of the National Academy of Sciences*, 110(17), pp.6937–6942.

Pfennig, D.W. *et al.,* 2010. Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution*, 25(8), pp.459–467.

Phillips, S.J. & Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31(2), pp.161–175.

Prunier, J. *et al.,* 2015. From genotypes to phenotypes: expression levels of genes encompassing adaptive SNPs in black spruce. *Plant Cell Reports*, 34(12), pp.2111–2125.

Pyhajarvi, T. *et al.,* 2013. Complex Patterns of Local Adaptation in Teosinte. *Genome Biology and Evolution*, 5(9), pp.1594–1609.

Qi, Q., Li, J. & Cheng, J., 2014. Reconstruction of metabolic pathways by combining probabilistic graphical model-based and knowledge-based methods. *BMC Proceedings*, 8(Suppl 6), p.S5.

Quarrie, S.A. & Jones, H.G., 1977. Effects of abscisic acid and water stress on development and morphology of wheat. *Journal of Experimental Botany*, 28(1), pp.192–203.

Raherison, E. *et al.*, 2012. Transcriptome profiling in conifers and the PiceaGenExpress database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. *BMC genomics*, 13(1), p.434.

Rajkumar, H. *et al.,* 2015. De Novo Transcriptome Analysis of Allium cepa L.(Onion) Bulb to Identify Allergens and Epitopes. *PloS one*, 10(8), p.e0135387.

Reséndiz Arias, Cecelic, sustentante Caracteres morfoanatómicos para la inferencia filogenética de la familia pinaceae / 2014

Rellstab, C. *et al.,* 2015. A practical guide to environmental association analysis in landscape genomics. *Molecular ecology*, 24(17), pp.4348–4370.

Rigault, P. *et al*., 2011. A White Spruce Gene Catalog for Conifer Genome Analyses. *Plant Physiology*, 157(1), pp.14–28. .

Ritland, K., 2012. Genomics of a phylum distant from flowering plants: conifers. *Tree Genetics & Genomes*, 8(3), pp.573–582.

Roberts, A. *et al.* 2011 Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 12.3 R22.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139–140.

Romero, I.G., Ruvinsky, I. & Gilad, Y., 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, 13(7), pp.505–516.

Salisbury, E.J., 1928. On the causes and ecological significance of stomatal frequency, with special reference to the woodland flora. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, pp.1–65.

Savolainen, O., Pyhäjärvi, T. & Knürr, T., 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics*, pp.595–619.

Schlichting, C.D., 1986. The evolution of phenotypic plasticity in plants. *Ann. Rev. Ecol. Syst.*, 17(1986), pp.667–693.

Smith, E.A. *et al.,* 2011. Developmental contributions to phenotypic variation in functional leaf traits within quaking aspen clones. *Tree physiology*, 31(1), pp.68–77.

Smith, J.M. & Haigh, J., 1974. The hitch-hiking effect of a favourable gene. *Genetical research*, 23(01), pp.23–35.

Skrøppa, T. & Johnsen, Ø., 1999. Patterns of adaptive genetic variation in forest tree species; the reproductive enviroment as an evolutionary force in Picea abies. In *Forest genetics and Sustainability*. Springer, pp. 49–58.

St Clair, S.B. *et al.,* 2009. Soil drying and nitrogen availability modulate carbon and water exchange over a range of annual precipitation totals and grassland vegetation types. *Global Change Biology*, 15(12), pp.3018–3030.

Stillman, J.H. & Armstrong, E., 2015. Genomics Are Transforming Our Understanding of Responses to Climate Change. *BioScience*, XX(X), pp.1–10.

Storey, K.B., 2004. Strategies for exploration of freeze responsive gene expression: advances in vertebrate freeze tolerance. *Cryobiology*, 48(2), pp.134–145.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), pp.585–595.

Tichá, I., 1982. Photosynthetic characteristics during ontogenesis of leaves. 7. Stomata density and sizes. *Photosynthetica*.

Trabucco, A. & Zomer, R.J., 2009. Global aridity index (global-aridity) and global potential evapo-transpiration (global-PET) geospatial database. *CGIAR Consortium for Spatial Information. Published online, available from the CGIAR-CSI GeoPortal at: http://www. csi. cgiar. org/(2009). Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database. In.*

Tsai, I.J. *et al.,* 2013. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*, 496(7443), pp.57–63.

Vielle-Calzada, J.-P. *et al.,* 2009. The Palomero genome suggests metal effects on domestication. *Science*, 326(5956), p.1078.

Wachowiak, W. *et al.,* 2015. Comparative transcriptomics of a complex of four European pine species. *BMC Genomics*, 16(1).

Wang, F., Polydore, S. & Axtell, M.J., 2015. More than meets the eye? Factors that affect target selection by plant miRNAs and heterochromatic siRNAs. *Current Opinion in Plant Biology*, 27, pp.118–124.

Warren, R.L. *et al.,* 2015. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *The Plant Journal*, 83(2), pp.189–212.

Wegrzyn, J.L. *et al*., 2008. TreeGenes: a forest tree genome database. *International journal of plant genomics*, 2008.

Wegrzyn, J.L. *et al.,* 2014. Unique features of the loblolly pine (Pinus taeda L.) megagenome revealed through sequence annotation. *Genetics*, 196(3), pp.891–909.

Woodward, F.I., 1987. Stomatal numbers are sensitive to increases in CO_2 from pre-industrial levels. *Nature*, 327(6123), pp.617–618.

Wu, T.D. & Watanabe, C.K., 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, 21(9), pp.1859–1875.

Yu, Q., Tigerstedt, P.M.A. & Haapanen, M., 2001. Growth and phenology of hybrid aspen clones (Populus tremula L. x Populus tremuloides Michx.). *Silva fennica*, 35(1), pp.15–25.

Zhao, C. *et al.,* 2005. The xylem and phloem transcriptomes from secondary tissues of the Arabidopsis root-hypocotyl. *Plant Physiology*, 138(2), pp.803–818.

Zhou, D. *et al.,* 2004. DNA microarray analysis of genome dynamics in Yersinia pestis: insights into bacterial genome microevolution and niche adaptation. *Journal of bacteriology*, 186(15), pp.5138–5146.