



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**Detección de selección natural en genes nucleares a través  
de una filogenia de *Abies***

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**BIÓLOGO**

**P R E S E N T A:**

**José Miguel Amaro Estrada**



**DIRECTOR DE TESIS:  
DR. JUAN PABLO JARAMILLO CORREA**

**CIUDAD UNIVERSITARIA, CD.MX. 2018**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

## **Hoja de datos del jurado**

### **1. Datos del alumno**

Amaro Estrada José Miguel

Universidad Nacional Autónoma de México

Facultad de Ciencias

Biología

### **2. Datos del tutor**

Dr. Juan Pablo Jaramillo Correa

### **3. Datos del sinodal**

Dr. Daniel Ignacio Piñero Dalmau

### **4. Datos del sinodal**

Dr. Lev Orlando Jardón Borbolla

### **5. Datos del sinodal**

Dra. Tania Garrido Garduño

### **6. Datos del sinodal**

Dr. Erick Alejandro García Trejo

### **7. Datos del trabajo escrito**

Detección de selección natural en genes nucleares a través de una filogenia de *Abies*.

pp. 69

2018

## **Agradecimientos**

A mi familia, en especial a mis padres, Sandra y Miguel, y a mi hermana, Galilea, por el amor, el apoyo y por darme todo lo que necesito y mucho más.

A mis amigos de la prepa, Helena, Cobo y Roy, gracias por diez maravillosos años de amistad, de risas, lágrimas y dramas, ojalá sean muchos más. Y a los que he conocido en el camino, Rini, César, Miriam, Karla, Thael, Vivianne...que son parte de la misma gran familia, gracias por todos los momentos y experiencias. A Laura, por todo lo anterior y por el inacabable apoyo emocional.

A mis amigos de la Facultad de Ciencias, Marco, Gustavo, Santiago, Gloria, Olivia, Zulo, Paul, Edgar. A Shary por tantas risas y consejos y a Beto, por las infinitas charlas. Todos ustedes hicieron de la carrera una mejor experiencia de lo que yo mismo esperaba.

A mis compañeros y amigos del laboratorio, Tania y Susette, por el apoyo técnico. A Eréndira y Alfredo por las primeras enseñanzas en el laboratorio y junto con Gustavo y Marco, por el aprendizaje y la amistad. Su apoyo fue profundamente valioso durante este proceso.

A mis profesores de la prepa 6, Sandra, Cosme, Marichuy y Juan, gracias a quienes descubrí y aprendí a valorar mi vocación. Y a los de la Facultad de Ciencias, quienes con su ejemplo y enseñanzas me inspiraron a seguir adelante.

A mis sinodales, los doctores Lev Jardón y Daniel Piñero por sus valiosos comentarios para mejorar este trabajo, y al Dr. Erick García Trejo y la Dra. Tania Garrido por ser además, mis profesores.

A mi tutor, el Dr. Juan Pablo Jaramillo, con quien tuve una experiencia de trabajo única, de mucho aprendizaje, no solo académico sino también personal.

A la UNAM, por toda la experiencia académica, a donde pertenezco ya por 10 años y gracias a la cual me asumo orgullosamente, puma.

# ÍNDICE

<b>Resumen</b>	6
<b>Introducción</b>	
1. Variación genética y la teoría neutral	8
2. Efectos de la selección natural a nivel molecular	10
2.1 Espectro de frecuencias alélicas, LD y hitchiking	10
2.2 Probabilidad de fijación y tasas de sustitución de nuevos alelos	11
2.3 Diferenciación y divergencia	12
3. Pruebas de neutralidad	13
3.1 Pruebas basadas en polimorfismos	13
3.2 Pruebas basadas en divergencia	15
3.3 Métodos combinados	15
4. Extensiones de las pruebas de neutralidad: métodos filogenéticos	17
4.1 Relative Rate Ratio test	17
4.2 Modelos evolutivos por codones	18
5. Aspectos evolutivos y adaptativos de las coníferas	20
5.1 Características que permiten estudiar la selección	20
5.2 Evolución molecular en coníferas y selección natural	21
5.3 El género <i>Abies</i>	23
<b>Objetivos</b>	26
General	
Objetivos particulares	
<b>Justificación</b>	27
<b>Hipótesis</b>	29
<b>Método</b>	30
Selección, amplificación de los genes y construcción de bases de datos	30
Análisis de diversidad y divergencia	31

Pruebas de neutralidad	32
<b>Resultados</b>	<b>37</b>
Obtención de las secuencias	37
Diversidad y divergencia	38
Pruebas de neutralidad	39
Pruebas filogenéticas: RRR test y Neutral Substitution Test	40
Pruebas filogenéticas: modelos evolutivos por codones	43
<b>Discusión</b>	<b>47</b>
Diversidad y divergencia	47
Patrones de selección a nivel filogenético	49
Genes candidatos bajo selección positiva	53
<b>Conclusiones</b>	<b>57</b>
<b>Anexos</b>	<b>59</b>
<b>Bibliografía</b>	<b>63</b>

## RESUMEN

Uno de los objetivos fundamentales de la biología evolutiva, es entender cómo las diferentes fuerzas evolutivas, tales como la mutación, deriva génica y selección natural afectan la diversidad genética de las poblaciones. En este campo, la teoría neutral de la evolución molecular ha jugado un papel central, ya que provee de bases teóricas para entender la variación genética. La acción de la selección natural altera los patrones de la diversidad genética con respecto a lo esperado bajo neutralidad, dando como resultado distribuciones sesgadas de las frecuencias alélicas, niveles de diversidad genética más altos o bajos de lo esperado y regiones genómicas que evolucionan a tasas contrastantes.

Partiendo de lo anterior se han desarrollado diversas pruebas estadísticas que utilizan las predicciones de la teoría neutral como modelo nulo para demostrar la ocurrencia de selección natural a lo largo de los genomas de las especies. Entre estas pruebas, son de particular interés aquellas que permiten estudiar la selección en un contexto filogenético, ya que con ellas es posible ubicar los eventos selectivos en ramas particulares dentro de una filogenia y por lo tanto, localizarlos en el tiempo y asociarlos con eventos evolutivos importantes, como la especiación.

Las coníferas representan buenos modelos para estudios de selección natural, ya que generalmente tienen grandes tamaños poblacionales efectivos, con poca diferenciación y alta diversidad genética, lo cual facilita la acción de la selección natural. El género *Abies* es el segundo más grande de la familia Pinaceae, y son componentes principales de los bosques templados de gran altitud y de las regiones boreales del hemisferio norte. Al igual que otras coníferas y especies de regiones templadas, las especies de *Abies* han experimentado grandes cambios ambientales y de distribución, tales como migraciones, expansiones y contracciones poblacionales asociados a sucesivos períodos de calentamiento y enfriamiento globales; por esto es de esperar que hayan estado sujetos a eventos selectivos que hayan dado lugar a adaptaciones a diferentes condiciones ambientales durante y después de sus procesos de especiación.

En este trabajo se evaluaron los patrones de variación y divergencia de 23 genes nucleares en diferentes especies del género *Abies* con el objetivo de detectar eventos de selección positiva en puntos específicos de la historia evolutiva del género. Para esto, se implementaron pruebas de neutralidad a nivel filogenético mediante la reconstrucción de una filogenia parcial.

Se detectaron señales de selección negativa en al menos el 47% de los genes analizados, lo cual probablemente está relacionado con el sesgo hacia genes conservados que fueron analizados en este estudio. En uno de los genes se encontró evidencia de selección no direccional actuando en todo el género (DCL1), así como dos posibles candidatos más, con evidencias de selección direccional (CDPK6) y no direccional (0\_8531\_01) en ramas particulares de la filogenia. Estos resultados sugieren que a nivel filogenético, la selección natural ha estado actuando sobre genes asociados con la regulación de eventos del desarrollo, la fenología y con la respuesta a diferentes tipos de estrés biótico y abiótico, ya que los genes encontrados como sujetos de selección positiva están relacionados con estas funciones. Adicionalmente, se reforzó la necesidad de generar una mayor disponibilidad de recursos genéticos para especies no modelo como las coníferas del género *Abies*, lo cual permitirá no solo entender mejor sus procesos evolutivos, si no también mejorar las estrategias de conservación a implementar.

**Palabras clave:** selección natural, pruebas de neutralidad, adaptación, filogenia, *Abies*, genes nucleares.

# INTRODUCCIÓN

## 1. Variación genética y la teoría neutral

Uno de los objetivos fundamentales de la biología evolutiva es entender cómo las diferentes fuerzas evolutivas afectan la diversidad genética de las poblaciones. En este campo, la teoría neutral de la evolución ha jugado un papel central, ya que provee las bases teóricas para entender la variación genética (Kimura, 1991). Originalmente propuesta por Kimura, (1968), esta teoría se basó en diferentes propiedades de la evolución a nivel molecular, como la tasa aproximadamente constante con la que ocurren las mutaciones y la naturaleza conservadora de las mismas (regiones conservadas que evolucionan con restricción o más lentamente que otras), así como los descubrimientos sobre los niveles de diversidad molecular mucho más altos de lo esperado bajo los modelos de selección hasta entonces predominantes (Kimura, 1991; Nei, 2005; Nei, Suzuki, & Nozawa, 2010). De manera general, la teoría neutral postula que la gran mayoría de los cambios evolutivos a nivel molecular ocurren debido a la fijación aleatoria de variantes génicas neutrales, por efecto de la deriva génica, y a la entrada constante de nuevas variantes, producto de la mutación (Kimura, 1991). La teoría neutral también admite la existencia de una gran cantidad de mutaciones deletéreas, las cuales son eliminadas de la población por medio de selección natural, así como la ocurrencia de una pequeña porción de mutaciones ventajosas que eventualmente pueden llegar a fijarse por selección positiva (Nei *et al.*, 2010).

La neutralidad depende de si los cambios en las frecuencias alélicas están determinados principalmente por deriva génica. Entonces, si el coeficiente de selección,  $s$ , es la diferencia selectiva entre dos alelos y además  $s < 1/(2N_e)$ , siendo  $N_e$  el tamaño poblacional efectivo, podemos decir que los alelos son neutrales entre sí. Esto implica que el carácter neutral, ventajoso o deletéreo de un alelo, depende del tamaño poblacional efectivo (Hedrick, 2005).

Si los cambios en las frecuencias alélicas están determinados por la deriva y la mutación (neutralidad), entonces la tasa de sustitución nucleotídica ( $k$ ) estaría

determinada por la tasa de mutación  $\mu$  en una población de tamaño  $2Ne$ . En ésta, y bajo un modelo de mutación de sitios infinitos, el número esperado de mutantes por generación sería  $2Ne\mu$ . Como la probabilidad de fijación de una mutación nueva es igual a  $1/2Ne$ , la tasa de sustitución alélica  $k$  sería igual al producto del número de mutaciones nuevas por su probabilidad de fijación:  $k = 2Ne\mu \left(\frac{1}{2Ne}\right)$ , de lo cual deriva que  $k=\mu$  (Gillespie, 2010). Entonces, la teoría neutral predice que para mutaciones neutrales, la tasa de sustitución nucleotídica será igual a la tasa de mutación en ese locus (Kimura, 1991). Es importante resaltar la independencia entre la tasa de sustitución y el tamaño poblacional efectivo. Esto ocurre debido a que en poblaciones más pequeñas, hay menos mutantes, es decir, el valor de  $2Ne\mu$  es menor, pero la frecuencia inicial de los alelos mutantes es mayor, lo cual aumenta su probabilidad de fijación (Hedrick, 2005).

Cuando existe un equilibrio entre las variantes introducidas por mutación y las eliminadas por deriva, se alcanza un nivel de heterocigosis dado por  $H = \frac{\theta}{\theta+1}$ , donde  $\theta = 4Ne\mu$ . De acuerdo con esto, la teoría neutral predice que cualquier incremento en la tasa de mutación  $\mu$  o en el tamaño poblacional efectivo  $Ne$ , resultará en un aumento en la variación genética de las poblaciones. Esto ha sido demostrado en estudios con microsatélites y alozimas; los primeros tienen simultáneamente tasas de mutación y heterocigosis más altas que otros marcadores, mientras que para las segundas existe una correlación entre la heterocigosis y el tamaño poblacional efectivo (Hedrick, 2005), tal como se predice con esta fórmula.

Al comparar las predicciones sobre las tasas de sustitución y la variabilidad intraespecífica, puede derivarse una tercera predicción: que aquellas regiones que acumulan más variación a través del tiempo deberían presentar mayores niveles de polimorfismo intraespecífico (Kimura, 1991). Dicho de otra manera, debería existir una correlación positiva entre el nivel de divergencia de las secuencias y su nivel de variación intraespecífica. Esta predicción ha sido apoyada por el hecho de que los sitios sinónimos son más polimórficos y presentan mayores tasas evolutivas que los sitios no sinónimos. Igualmente, otros estudios han revelado una fuerte correlación

entre la distancia genética y la heterocigosis en proteínas de diversos vertebrados e invertebrados, apoyando esta predicción (Hahn, 2008; Kimura, 1991).

## **2. Selección natural a nivel molecular**

La selección natural puede ser definida como cualquier diferencia consistente en cuanto a supervivencia y éxito reproductivo entre entidades biológicas fenotípicamente distintas (Futuyma, 2013). De acuerdo con esta definición, cualquier carácter que aumente la probabilidad de supervivencia y reproducción de un organismo tiene mayores probabilidades de pasar a las siguientes generaciones y por lo tanto de aumentar su frecuencia en la población. A nivel molecular, esto se refiere a la propagación diferencial no aleatoria de un alelo como consecuencia de su efecto fenotípico (Vitti, Grossman, & Sabeti, 2013).

La selección natural puede actuar de diversas maneras. Por un lado, puede favorecer un alelo y promover su fijación, en cuyo caso se trata de selección direccional positiva. En caso contrario, puede actuar en contra de un alelo para eliminarlo de la población, lo cual se conoce como selección negativa. En otros casos, la selección puede mantener diversos alelos con frecuencias intermedias, lo cual se conoce como selección balanceadora (o de manera más general, selección no direccional); esta ocurre debido a fenómenos como la ventaja de los heterocigos (sobredominancia), por selección en ambientes heterogéneos (Gillespie, 2009), por selección dependiente de la frecuencia (Vitti *et al.*, 2013) o por fluctuaciones temporales de las condiciones ambientales.

### 2.1 Efectos de la selección natural a nivel molecular

#### *Espectro de las frecuencias alélicas, desequilibrio de ligamiento y hitchhiking*

El espectro de frecuencias es una forma de computar el número de mutaciones que existen en una determinada frecuencia a lo largo de una secuencia de DNA en una población; es decir, es un resumen de las frecuencias alélicas de las diferentes mutaciones dentro de una población (Nielsen, 2005). Bajo neutralidad, se espera que

la mayoría de las mutaciones se encuentren con frecuencias bajas, mientras que las mutaciones de alta frecuencia deberían ser raras (Nielsen & Slatkin, 2013). Bajo los efectos de la selección negativa, las mutaciones tienden a segregarse con bajas frecuencias y por lo tanto, el espectro estará sesgado hacia un exceso de este tipo de variantes. Por otro lado, la selección positiva provoca que las mutaciones aumenten su frecuencia, lo cual causa que el espectro de frecuencias se sesgue hacia mutaciones de frecuencias altas. En este caso, es importante considerar los efectos del desequilibrio de ligamiento (LD) y el hitchhiking genético. El primero es una asociación no aleatoria entre alelos de diferentes *loci*, el cual aumenta durante el proceso selectivo en las regiones que están sujetas a selección (Nielsen, 2005). Una vez terminado el evento de selección, el LD se observa en los sitios aledaños a la región seleccionada. El *hitchhiking* genético se define por su parte como el cambio en las frecuencias de los alelos neutrales debido a una asociación con alelos que están sujetos a selección (Hedrick, 2005). Estos fenómenos generan un patrón de barrido selectivo, en el cual los sitios ligados a los sitios seleccionados positivamente aumentarán su frecuencia junto a estos últimos, mientras que los sitios no ligados, tenderán a disminuir su frecuencia. Esto causará que el espectro de frecuencias esté sesgado hacia un exceso de variantes de alta y baja frecuencia, mientras que las variantes con frecuencias intermedias serán muy pocas. Finalmente, la selección balanceadora tendría el efecto opuesto, promoviendo un exceso de variantes de frecuencia intermedia, ya que los sitios ligados se mantendrían junto con los sitios seleccionados (Nielsen & Slatkin, 2013).

#### *Probabilidad de fijación y tasas de sustitución de los nuevos alelos*

Cuando la selección natural actúa favoreciendo a un alelo, aumentan sus probabilidades de fijación. Para un alelo  $A$  sujeto a selección positiva, la probabilidad de fijación  $u$ , con un coeficiente de selección  $s$  es:

$$u(s, N) = \frac{1 - e^{-2s}}{1 - e^{-4Ns}}$$

Para una  $N$  determinada, podemos encontrar tres rangos de intensidad de la selección. Mientras  $s$  aumenta,  $u$  se vuelve aproximadamente igual a  $2s$ , y cuando  $u \approx 2s$ , se trata

de una mutación altamente ventajosa, en cuyo caso,  $2Ns > 1$ , y el destino de esa mutación está determinado principalmente por la selección. Si la selección es más débil,  $-1 < 2Ns < 1$ , el comportamiento del alelo será muy parecido al de una variante neutral. Hablamos entonces de mutaciones casi neutrales y la selección será un factor importante en el comportamiento del alelo, pero también estará involucrado el tamaño poblacional y por lo tanto la deriva génica. Finalmente, cuando  $2Ns < -1$ , la mutación es altamente deletérea y su probabilidad de fijación disminuye rápidamente mientras  $s$  se vuelve más negativa (Nielsen & Slatkin, 2013).

Al estar involucrado un coeficiente de selección, la tasa de sustitución para los alelos nuevos también cambia. En este caso, al igual que para las mutaciones neutrales, se considera el número de mutantes que aparecen en cada generación,  $2N\mu$ , y la probabilidad de fijación de estas nuevas variantes considerando ahora su coeficiente de selección asociado,  $u(s, N)$ , entonces:  $k(s, N) = 2N\mu u(s, N)$  (Gillespie, 1998). De acuerdo con esta ecuación,  $u(s, N) > 1/2N$  si  $s > 0$  y  $u(s, N) < 1/2N$  si  $s < 0$ ; por lo tanto, los alelos ventajosos serán sustituidos a una tasa más alta que los alelos neutrales en una misma población, mientras que los alelos desventajosos serán sustituidos a tasas más bajas, siendo la tasa de sustitución para los alelos altamente deletéreos esencialmente igual a 0 (Nielsen & Slatkin, 2013).

## 2.2 Diferenciación y divergencia

Cuando las especies están divididas en poblaciones, o cuando el proceso de especiación es incipiente, los niveles de variación genética en las poblaciones pueden verse afectados de manera diferencial si ocurre selección natural.

Por un lado, la selección puede actuar de manera contrastante en dos poblaciones diferentes, promoviendo diferencias a largo plazo en las frecuencias alélicas de cada una y por lo tanto, afectando también los niveles de divergencia entre ellas en los *loci* seleccionados. Por otro lado, la selección positiva puede aumentar las diferencias en las frecuencias alélicas entre poblaciones durante un breve periodo de tiempo, en el cual la frecuencia del alelo seleccionado es mayor en la población en la que se originó.

Sin embargo, es importante considerar que la selección positiva también puede homogeneizar las frecuencias entre poblaciones, por ejemplo, cuando el mismo tipo de selección balanceadora actúa sobre ellas (Nielsen & Slatkin, 2013).

La selección también puede modificar los patrones de variación y los niveles de divergencia en diferentes regiones genómicas, debido a que esta actúa de manera diferente dependiendo de si se trata, por ejemplo, de regiones codificantes o no codificantes, regiones constreñidas (proteínas cuya función se puede ver comprometida si cambian los aminoácidos que las componen) o no constreñidas funcionalmente, etc. Así, pueden observarse diferentes regiones genómicas que evolucionan con tasas contrastantes (Biswas & Akey, 2006).

### **3. Pruebas de neutralidad**

Uno de los principales aportes de la teoría neutral es que sus predicciones sobre la distribución de las frecuencias alélicas en las poblaciones y sobre la variación genética dentro y entre las especies son lo suficientemente simples para ser derivadas matemáticamente y ser probadas mediante experimentos y observaciones (Kimura, 1991; Nei, 2005). Se han desarrollado diversas pruebas estadísticas que buscan demostrar la existencia de selección natural a lo largo de los genomas de las especies. Estas utilizan las predicciones de la teoría neutral como modelo nulo y sugieren selección cuando estas no se cumplen. Estas pruebas han sido utilizadas ampliamente para reportar la ocurrencia de selección en los genomas de diferentes especies (Nei *et al.*, 2010). De manera general, podemos clasificar estas pruebas estadísticas en aquellas que usan datos de polimorfismos, las que usan datos de divergencia entre secuencias y las que combinan información de ambos tipos (Jensen, Wong, & Aquadro, 2007).

#### *3.1 Pruebas basadas en polimorfismos*

La mayoría de estas pruebas, asumen modelos en los que la variación observada es selectivamente neutral y que la población está en equilibrio, mientras que otras

pruebas de este tipo consideran los modelos de *hitchhiking*, basados en los patrones de variación molecular previamente señalados para este fenómeno (Jensen *et al.*, 2007).

Para datos de variación nucleotídica, la prueba más común es la  $D$  de Tajima. Esta prueba compara dos estimadores del parámetro  $\theta$ . El primero, es el estimador de Watterson ( $\theta_W$ ), que se basa en el número de sitios segregantes. El segundo es el estimador de Tajima ( $\theta_T$ ), que utiliza el número promedio de diferencias nucleotídicas entre pares de secuencias, de tal forma que:

$$\theta_W = S / [\sum_{i=1}^{n-1} 1/i] \text{ y } \theta_T = \frac{\sum_{(i,j):i<j} d_{ij}}{n(n-1)/2}$$

Donde  $n$  es el número de cromosoma muestreados,  $S$  el número de sitios segregantes y  $d_{ij}$  es el número de diferencias nucleotídicas entre las secuencias  $i$  y  $j$ . La prueba de Tajima, compara los valores de estos parámetros de la siguiente manera:

$$D = \frac{\theta_T - \theta_W}{\sqrt{V(\theta_T - \theta_W)}}$$

Donde  $V$  representa la varianza. Bajo neutralidad, se espera que estos dos estimadores sean iguales y por lo tanto que  $D$  sea aproximadamente igual a cero. Cuando sus valores son significativamente diferentes entre sí, el modelo neutral es rechazado. De esta forma, un barrido selectivo deja un exceso de variantes de alta y baja frecuencia, lo que se traduce en un mayor efecto sobre  $\theta_W$  que sobre  $\theta_T$ , dando como resultado un valor negativo de  $D$ . Por otro lado, cuando existe selección balanceadora, los valores de  $D$  son significativamente positivos, ya que hay un exceso de variantes de frecuencia intermedia que aumenta los valores de  $\theta_T$  (Nielsen & Slatkin, 2013). Adicionalmente, se han desarrollado otras pruebas basadas en estadísticos ligeramente diferentes, tales como la  $F$  de Fu y Li (Fu & Li, 1993) y la  $H$  de Fay y Wu (Fay & Wu, 2000). Esta última prueba, compara el número de diferencias entre pares de individuos con el número de individuos que son homocigos para el alelo derivado; cuando se obtienen valores pequeños de  $H$  se infiere un exceso de alelos de alta frecuencia, y por lo tanto evidencia de selección positiva (Vitti *et al.*, 2013).

Estas pruebas se han aplicado para probar desviaciones de la neutralidad, sin embargo algunos estudios han demostrado que estos estadísticos también son susceptibles a eventos demográficos como los cuellos de botella y la estructura poblacional, los cuales producen valores similares a los de la selección. Por lo tanto, es posible que estas pruebas tengan menos poder cuando se prueban modelos comunes de selección que modelos que consideran eventos demográficos (Nielsen, 2001). En pruebas como a  $D$  de Tajima, los eventos demográficos pueden distinguirse de los de selección ya que los patrones dejados por los primeros son homogéneos a nivel genómico, mientras que los segundos ocurren en *loci* específicos.

### 3.2 Pruebas basadas en divergencia

Estos métodos permiten utilizar datos de diferentes especies y pueden utilizarse para detectar selección en genes específicos o a nivel genómico. Generalmente suponen que existen regiones que evolucionan con poca o ninguna restricción e infieren selección positiva cuando una región evoluciona más rápido que las regiones no constreñidas. Para regiones codificantes, las pruebas consideran que una comparación entre la tasa de sustituciones no sinónimas ( $dn$ ) y sinónimas ( $ds$ ) proveen de una buena medida de la fuerza y el tipo de selección que actúa sobre las secuencias (Jensen *et al.*, 2007). Bajo neutralidad, se espera que el coeficiente  $dn/ds$  sea igual a uno. Cuando existe selección negativa este debería ser menor que uno y bajo selección positiva, mayor que uno.

Sin embargo, para la mayoría de los genes, el coeficiente  $dn/ds$  promedio entre todos los sitios es menor que uno, ya que la mayoría de las mutaciones no sinónimas son deletéreas. Sin embargo, es posible que si un número suficiente de mutaciones es afectada por selección positiva, el coeficiente  $dn/ds$  de en esa región sea igual a 1. Por otro lado, si la selección actúa solo sobre una misma mutación o si afecta a regiones no codificantes, estos métodos no tienen poder de detectar selección (Nielsen & Slatkin, 2013).

### 3.3 Métodos combinados

Este tipo de pruebas utilizan información tanto de variación intraespecífica como de divergencia entre especies. Una de las primeras pruebas de este tipo es la de HKA (Hudson, Kreitman, & Aguadé, 1987). Ésta utiliza la correlación predicha por la teoría neutral entre la variación intrapoblacional y la divergencia interespecífica. Para realizarla es necesario contar con información de dos regiones genómicas diferentes de dos o más especies, de tal forma que se posea información tanto de variación como de divergencia (Hedrick, 2005). Si la tasa de sustitución entre especies es igual a la tasa de mutación  $\mu$ , el número de diferencias fijas entre especies es igual a  $2T\mu$ ; donde  $T$  es el tiempo de divergencia de las especies medido en número de generaciones. Si consideramos la variación poblacional ancestral, tenemos  $2T\mu+4N\mu$ , de tal forma que el número esperado de mutaciones dentro y entre especies es proporcional a la tasa de mutación; así, el coeficiente entre mutaciones esperadas dentro y entre especies depende del tamaño poblacional efectivo, del tiempo de divergencia y del valor de  $\theta$  para cada gen (Nielsen & Slatkin, 2013; Hedrick, 2005). Para datos de varios *loci*, con un mismo tamaño de muestra, el número de diferencias fijas entre especies ( $F$ ) debería ser igual al número de diferencias segregantes dentro de cada especie ( $S$ ); de tal forma que  $E(S_1)/E(F_1) = E(S_2)/E(F_2)$ . Cuando estos coeficientes son significativamente diferentes, puede indicar selección natural. Por ejemplo, un coeficiente  $S/F$  pequeño, indicaría un barrido selectivo que ha eliminado la variación intraespecífica alrededor de un sitio seleccionado con respecto al número de mutaciones fijas entre especies, mientras que un coeficiente  $S/F$  grande, podría suponer que se han manteniendo niveles de mutaciones segregantes en mayor número que las mutaciones fijas entre especies a través de selección balanceadora (Nielsen & Slatkin, 2013; Vitti *et al*, 2013).

Otra prueba que considera la variación intraespecífica y la divergencia entre especies es la prueba MK (McDonald & Kreitman, 1991). Esta se basa en los mismos principios que la prueba HKA, pero considera dos clases de mutaciones: las sinónimas y las no sinónimas. Bajo neutralidad, se espera que el coeficiente entre variantes no sinónimas y sinónimas fijas entre especies ( $F_N$  y  $F_S$ ) sea igual al coeficiente de variantes no

sinónimas y sinónimas que segregan ( $S_N$  y  $S_S$ ) dentro de las especies. En ambos casos, el cociente estará determinado por la diferencia en la tasa de mutación entre los sitios sinónimos y no sinónimos. De acuerdo con esto, tendríamos que  $E(S_N)/E(S_S) = E(F_N)/E(F_S)$ ; y de manera similar  $E(S_N)/E(F_N) = E(S_S)/E(F_S)$  (Hedrick, 2005; Nielsen & Slatkin, 2013). Cuando el coeficiente entre especies es mayor que el coeficiente dentro de especies, la hipótesis de neutralidad puede ser rechazada e indicar selección positiva, mientras que si el coeficiente entre especies es menor que el intraespecífico, puede tratarse de selección balanceadora o de un exceso de variantes ligeramente deletéreas que se encuentran bajo selección negativa débil dentro de alguna de las especies (Vitti *et al.*, 2013). Debido a que esta prueba considera los patrones de variación y divergencia dentro de la misma región genómica, se considera que es estadísticamente más poderosa que la prueba HKA (Hedrick, 2010).

#### **4. Extensiones de las pruebas de neutralidad: métodos filogenéticos**

Para algunas de las pruebas anteriores se han desarrollado variantes que permiten estudiar la selección en un contexto filogenético. Estas son de particular interés, ya que con ellas es posible ubicar los eventos selectivos en ramas particulares dentro de una filogenia y por lo tanto, localizarlos en el tiempo y asociarlos con eventos evolutivos importantes, como la especiación (Palmé, Pyhäjärvi, Wachowiak, & Savolainen, 2009).

##### *4.1 “Relative Rate Ratio” test*

Este método fue desarrollado por Creevey y McInerney (Creevey & McInerney, 2002). A partir de una base de datos multiespecífica se construye una filogenia enraizada con un grupo externo para polarizar los datos. Posteriormente se reconstruyen las secuencias ancestrales hipotéticas para cada rama interna mediante un método de parsimonia. Luego, se toman en cuenta las sustituciones que han ocurrido a lo largo de la filogenia, considerando la secuencia en cada rama interna y las secuencias de los grupos derivados de ese ancestro y se determina si se trata de sustituciones

sinónimas o no sinónimas. Con esto se producen cuatro valores correspondientes al tipo de mutación: **RI** y **RV** son sustituciones de reemplazo (no sinónimas) invariables y variables, respectivamente; mientras que **SI** y **SV** corresponden a sustituciones silenciosas (sinónimas) invariables y variables. Por sustitución invariable se entiende a aquellos cambios que una vez ocurridos, se mantienen en todos los linajes descendientes, mientras que las sustituciones variables son aquellas que cambian en solo algunos de los linajes descendientes. Posteriormente se comparan los coeficientes **SI/SV** y **RI/RV**, mediante un *G*-test o una prueba exacta de Fisher. Bajo neutralidad se espera que ambos coeficientes sean iguales, de lo contrario puede inferirse selección. Bajo selección direccional se espera que las mutaciones ventajosas ocurran y se mantengan invariables a una tasa mayor que la esperada bajo neutralidad, en cuyo caso el valor de **RI** será mayor. Por otro lado, bajo selección no direccional, las mutaciones ventajosas ocurrirán y se mantendrán variables a una tasa más alta, en cuyo caso los valores de **RV** serán más altos (Creevey & McInerney, 2002). Como se mencionó previamente, los valores para los cuatro tipos de mutaciones, así como sus coeficientes, son calculados para cada secuencia ancestral y sus secuencias derivadas inmediatas; por lo cual, ésta prueba permite detectar los eventos de selección en ramas particulares de la filogenia y asociar los eventos selectivos y evolutivos.

#### *4.2 Modelos evolutivos por codones*

Son métodos estadísticos para hacer inferencias sobre coeficientes  $dn/ds$  ( $\omega$ ) mediante modelos específicos (Goldman & Yangf, 1994). Estos se caracterizan por que consideran como unidad evolutiva al codón y no al nucleótido. El modelo generalmente usado no considera las diferencias químicas entre los nucleótidos, por lo cual utiliza el mismo coeficiente,  $\omega$ , para todas las sustituciones no sinónimas. Como se mencionó anteriormente, el promedio de  $\omega$  de todos los sitios difícilmente es superior a 1, debido a la acción de la selección negativa sobre la mayoría de las sustituciones no sinónimas; por lo cual estos modelos han buscado resolver este

problema mediante la detección de selección natural sobre linajes específicos de una filogenia o sobre sitios (codones) específicos dentro de una secuencia (Yang, 2007).

Los modelos por rama buscan utilizar diferentes coeficientes  $\omega$  para diferentes ramas de una filogenia; esto permite detectar selección sobre ellas sin necesidad de promediar los valores de  $\omega$  a lo largo de la misma (Yang, 1998). El modelo por rama más sencillo supone un solo valor de  $\omega$  para toda la filogenia, mientras que el más general o “modelo libre” puede llegar a tener tantos valores de  $\omega$  como ramas tenga la filogenia. Este último es rico en parámetros y por lo tanto se recomienda su uso en filogenias con pocas ramas. También existe una gama de modelos intermedios que han sido implementados en el programa PAML (Yang, 2007) para facilitar su cálculo (Yang, 1998). Por ejemplo, el modelo de dos coeficientes supone que ciertas ramas de interés (previamente identificadas) tienen un parámetro  $\omega_1$  distinto del parámetro  $\omega_0$  o de fondo; mientras que el modelo de tres coeficientes permite que dos ramas de interés tengan un coeficiente  $\omega$  diferente del de fondo y diferentes entre sí (Yang, 1998). Posteriormente, estos modelos pueden compararse mediante pruebas de verosimilitud para probar hipótesis. Por ejemplo, al comparar el modelo de un solo coeficiente contra el modelo libre, puede detectarse heterogeneidad en los valores de  $\omega$  entre ramas, lo cual podría ser causado por selección positiva. Al comparar los valores de  $\omega$  entre el modelo de un coeficiente y el de dos, puede detectarse si las ramas de interés tiene un valor significativamente diferente del resto, lo cual examina directamente la posibilidad de que la selección haya actuado sobre linajes específicos (Yang, 1998).

Por otro lado, existen los modelos por sitios. Estos consideran el valor de  $\omega$  para cada codón como una variable aleatoria dentro de una distribución estadística. Por lo tanto, permite diferentes valores del parámetro entre los codones de una secuencia y se detecta un sitio sujeto a selección positiva cuando  $\omega > 1$ . Al igual que los modelos por ramas, diferentes modelos pueden probarse mediante pruebas de verosimilitud. Se ha sugerido que dos pares de modelos son particularmente efectivos: un modelo neutro (M1a) contra un modelo de selección (M2a) y un modelo M7 (beta) contra un modelo M8 (beta+w) (Yang, 2007).

El modelo neutral supone dos categorías de sitios, los neutrales, cuyas sustituciones no sinónimas tienen un coeficiente  $\omega$  igual a 1, y los conservados, para los cuales las sustituciones no sinónimas son deletéreas y eliminadas por selección y solo las sustituciones sinónimas son permitidas, por lo tanto,  $\omega = 0$ . Los parámetros de este modelo incluyen la longitud de las ramas de la filogenia ( $t$ ), el coeficiente de transiciones /transversiones ( $k$ ) y las proporciones de las dos categorías de sitios ( $p_1 + p_2$ ). El modelo de selección, agrega una categoría extra de sitios seleccionados positivamente, los cuales tiene un valor de  $\omega$  mayor que 1 (Nielsen & Yang, 1998).

Los modelos M7 y M8, consideran los mismos tres parámetros generales que el modelo M0, pero implementan de diferentes formas la distribución de los valores de  $\omega$  que pueden ocurrir en datos reales. El modelo M7 utiliza una distribución beta, la cual puede tomar distintas formas para considerar la heterogeneidad de los valores de  $\omega$  a través de los diferentes sitios, pero sin que existan sitios seleccionados positivamente ( $\omega > 1$ ). El modelo M8 considera una proporción de sitios muestreados a partir de la distribución beta y una proporción de sitios con el mismo valor de  $\omega$ , lo cual puede ser útil para detectar sitios seleccionados positivamente (Yang, Nielsen & Goldman, 2000).

Estas pruebas basadas en determinar si  $\omega$  es igual o mayor que cero proveen evidencia directa de la ocurrencia de selección natural. Sin embargo, poseen limitaciones, como no considerar la ocurrencia de recombinación. Por esto, pueden no ser adecuadas en algunos casos para datos intraespecíficos (Nielsen, 2001).

## **5. Aspectos evolutivos y adaptativos de las coníferas**

### *5.1 Características que permiten estudiar la selección*

Las coníferas son un buen modelo para los estudios de selección, ya que poseen características particulares que determinan su evolución y adaptabilidad. Son organismos longevos que requieren de un buen nivel de plasticidad fenotípica para acoplarse a las condiciones ambientales cambiantes (Parent *et al.*, 2015). Estas aún forman poblaciones naturales con altos niveles de variación genética, en las cuales la

selección natural es más efectiva. Sus poblaciones generalmente son grandes y están conectadas por un flujo génico extenso. Todo esto propicia la acción de la selección natural y facilita la distinción de los eventos selectivos y los procesos estocásticos que han actuado a lo largo de su historia evolutiva (Buschiazzo, Ritland, Bohlmann, & Ritland, 2012; Neale & Kremer, 2011). Las tasas de diversificación de las coníferas son generalmente bajas, lo cual podría reflejar que han ocurrido menos eventos de cuello de botella asociados a especiación y por lo tanto menores niveles de variación interpoblacional (Buschiazzo *et al.*, 2012). Esto a su vez puede promover que ocurran eventos de adaptación local en las poblaciones a lo largo de su rango de distribución.

Las coníferas habitan una gran variedad de ambientes. Generalmente se les asocia con regiones templadas, como los bosques boreales del hemisferio norte y locaciones de alta montaña; sin embargo, también se les puede encontrar en bosques subtropicales siempre verdes, así como en poblaciones a nivel del mar. En general su distribución actual presenta un patrón espacial de clina, a través de los diferentes tipos de ambientes en los que habitan (Parent *et al.*, 2015; Aravanopoulos *et al.*, 2015). Los caracteres fenotípicos en coníferas, como el tiempo de crecimiento y la tolerancia al frío, se ajustan también a este patrón de variación clinal, lo cual facilita los estudios y la inferencia de la selección (Savolainen & Pyhäjärvi, 2007). Además, las coníferas han presentado a lo largo de su historia una serie de contracciones y expansiones poblacionales, así como grandes eventos migratorios asociados a los diferentes ciclos glaciales; esta historia evolutiva es fácilmente rastreable gracias al rico registro fósil (Aravanopoulos *et al.*, 2015; Neale & Kremer, 2011). Los genomas de organismos longevos, como las coníferas, retienen los efectos de estos eventos demográficos por más tiempo, pero en contraste, la selección natural tiene mayores efectos en sus poblaciones grandes. Con sus bajos niveles de diferenciación poblacional y bajos niveles de desequilibrio de ligamiento, es más fácil detectar los efectos de la selección natural que en organismos con historia y estructura poblacional más complejas (Savolainen & Pyhäjärvi, 2007).

## 5.2 Evolución molecular en coníferas y selección natural

En coníferas y otros árboles, la tasa de evolución neutral es generalmente baja. Comparada con los estimados de sustituciones sinónimas de  $1.5 \times 10^{-8}$  sustituciones sinónimas por año en *Arabidopsis*, la tasa neutral en *Populus* es de solo una sexta parte de la misma, mientras que en *Pinus* se encontraron tasas relativamente más altas (Savolainen & Pyhäjärvi, 2007; Willyard, Syring, Gernandt, Liston, & Cronn, 2007). Por otro lado, otro estudio revela que en las coníferas, los niveles de desequilibrio de ligamiento (LD) decaen rápidamente con la distancia, lo cual puede deberse a los altos niveles de entrecruzamiento y recombinación a nivel genómico que presentan estos organismos (Neale & Kremer, 2011).

Debido a las características poblacionales de las coníferas, esperaríamos altos niveles de diversidad en poblaciones en equilibrio. Sin embargo, los estimados de diversidad para estas especies se encuentran en el mismo nivel que las de otras especies de árboles, incluyendo angiospermas y en general son más bajos que los de plantas no arbóreas. Esto sugiere dos cosas: que las poblaciones de coníferas no se encuentran actualmente en equilibrio (no han recuperado sus niveles de diversidad genética después de los eventos migratorios y demográficos que han experimentado) o que existen otras fuerzas afectando su diversidad genética, como la selección natural (Savolainen & Pyhäjärvi, 2007). Se ha observado que las grandes poblaciones de coníferas y otros árboles poseen distribuciones alélicas bastante uniformes. Por ejemplo, en especies europeas de *Pinus* y *Populus*, la mayoría de la diversidad nucleotídica se encuentra dentro de las poblaciones, mientras que menos del 5% se encuentra entre poblaciones (Savolainen & Pyhäjärvi, 2007).

Diferentes estudios han detectado la existencia de casos de selección en estos organismos. De 151 *loci* analizados en diferentes estudios, aproximadamente el 15% ha sido reportado con alguna huella de selección natural (Savolainen & Pyhäjärvi, 2007). En *Pseudotsuga menziesii* se ha encontrado evidencia de barridos selectivos en cuatro genes candidatos mediante pruebas basadas en diversidad, mientras que con pruebas combinadas se han logrado detectar diferencias significativas entre los niveles de divergencia y polimorfismo en al menos tres genes (Pm\_CL61Contig1, Pm\_CL908Contig1, CN638556.1) así como cuatro genes con evidencia de barridos

selectivos detectados mediante una prueba compuesta DHEW, basada en la  $D$  de Tajima, la  $H$  de Fay y Wu y la prueba de Ewens-Waterson (Eckert *et al.*, 2009). Por otro lado, en *P. pinaster* se encontró evidencia de selección balanceadora y diversificadora en genes “outliers” relacionados con la tolerancia a la desecación (Eveno *et al.*, 2008). Otros estudios han buscado detectar selección natural por medio de la asociación de la variación alélica con clinas y variantes geográficas y ambientales. En *P. taeda* se han encontrado cinco genes asociados a una variación clinal de aridez (0-16169-01-216, 0-2179-01-216, 0-10240-01-410, 0-18281-01-345, UMN-6195-01-397) así como 24 genes con valores extremos de  $F_{ST}$ , varios de los cuales también presentan patrones de variación a lo largo de su distribución geográfica (Eckert *et al.*, 2010); de manera similar, en *Picea abies* se encontró un enriquecimiento de SNP's en genes candidatos asociados a variación latitudinal a lo largo de su distribución, lo cual sugiere que la selección local en las diferentes poblaciones actuales o selección en la población ancestral han sido las fuerzas que moldearon esta clina (Chen *et al.*, 2012). A nivel interespecífico, se han encontrado genes sujetos a selección de manera específica en una o varias especies de pinos, así como un gen (4-coumarate: CoA ligase) en el que la huella de selección era común a ambas especies (Grivet *et al.*, 2011); en este mismo estudio, se encontró que había una asociación entre la variación genética y la temperatura a lo largo del rango de distribución de estas especies. Finalmente, a nivel filogenético los resultados han sido complementarios y reflejan distintos escenarios de selección. Por una lado, no se ha encontrado un número significativo de sustituciones fijas por selección en genes nucleares a lo largo de una filogenia de *Pinus*, mientras que la mayor parte de las sustituciones no sinónimas resultan ser deletéreas o moderadamente deletéreas, sugiriendo la acción de selección negativa o purificadora (Eckert *et al.*, 2013). Por otro lado, se ha encontrado principalmente evidencia de selección negativa en 18 genes nucleares a lo largo de otra filogenia de *Pinus*, sin embargo, en este mismo estudio encontraron diferencias significativas en el coeficiente  $dn/ds$  entre la diferentes ramas para el gen *dhn1*, presumiblemente en la rama que lleva a *P. ponderosa*, lo cual, entre otros escenarios, puede ser evidencia de que la selección positiva ha estado afectando a ese gen en esa rama en particular (Palmé *et al.*, 2009).

### 5.3 El género *Abies*

El género *Abies* es el segundo más grande dentro de la familia Pinaceae. Comprende 48 especies, distribuidas en Asia, Norte y Centro América y en las regiones mediterránea y alpina de Europa (Xiang, Xiang, Guo, & Zhang, 2009; Xiang *et al.*, 2015). Los árboles de este género habitan ambientes diversos, incluyendo bosques boreales y templados, hasta aquellos de las regiones montañosas intertropicales (Jaramillo-Correa *et al.*, 2008). Diferentes estudios han confirmado la monofilia del género, lo cual, aunado a la apariencia uniforme de las especies, complica la clasificación al interior del grupo (Eckenwalder, 2009); esto ha generado dos principales esquemas de clasificación: uno que considera dos subgéneros, *Abies* y *Pseudotsuga* (Liu, 1971) y otro que divide las 48 especies en 10 secciones (Farjon & Rushforth, 1989)

Otros procesos que complican la taxonomía del género incluyen la hibridación y la introgresión que resultan en una evolución reticulada, y en una historia biogeográfica compleja. En este sentido, se ha encontrado evidencia para un escenario en el que el origen del grupo es en alguna región en el norte del Pacífico a partir de donde ocurrieron diferentes eventos migratorios hacia el Sur y a través del estrecho de Bering, lo cual explica la actual distribución disyunta entre Norte América y Asia de la Sección *Amabilis*; también se ha sugerido un origen híbrido para la Sección *Balsamea*, distribuida en Norte América y el Noreste de Asia así como un efecto de barrera para el mar de Turgait durante las migraciones hacia el sur en el Eoceno tardío, lo cual explicaría la distribución disyunta entre los *Abies* asiáticos y los de la Europa mediterránea (Xiang *et al.*, 2015). Para las especies Mesoamericanas, se ha encontrado que la diferenciación genética en marcadores mitocondriales no se relaciona con la taxonomía de las especies, mientras que los marcadores de cloroplasto reflejan mejor las relaciones evolutivas de este grupo (Jaramillo-Correa *et al.*, 2008); también se encontró, que de las cuatro especies analizadas en este estudio, solo una tenía niveles significativos de divergencia, mientras que el resto forman un grupo relativamente homogéneo. Además, para este grupo, se ha sugerido que durante los periodos glaciales han presentado distribuciones amplias y continuas, mientras que durante los

períodos interglaciales (como el actual Holoceno) sus poblaciones se han fragmentado y quedado aisladas en las regiones montañosas altas.

El papel de la selección natural en este género ha sido poco estudiado. A partir de datos de transcriptoma se ha logrado detectar 88 genes candidatos en *A. alba*, relacionados con funciones como la síntesis de compuestos secundarios, la regulación del crecimiento, la respuesta al estrés y el transporte de agua entre otros (Roschanski, Fady, Ziegenhagen, & Liepelt, 2013). De manera similar y también en *A. alba*, un estudio de expresión diferencial de genes logró detectar 296 candidatos relacionados con la respuesta al estrés hídrico. La mayoría de estos codifican proteínas aún no caracterizadas y de los cuales, 247 estaba regulados positivamente y 49, negativamente (Behringer, Zimmermann, Ziegenhagen, & Liepelt, 2015). Finalmente, otro estudio en esta misma especie en localidades de los Alpes Franceses logró detectar al menos ocho genes candidatos que presentaban patrones de selección a lo largo de un gradiente de humedad durante el invierno y la primavera. Este patrón también se reflejó en caracteres fenotípicos, ya que los individuos de estas mismas localidades presentaron diferencias significativas en el tiempo de aparición de brotes y en la duración del período de crecimiento (Roschanski *et al.*, 2016).

Las características antes citadas de las coníferas y en particular del género *Abies*, resultan atractivas para continuar esta clase de estudios. Por un lado, la amplia distribución, variedad de ambientes que habitan, así como los diferentes eventos migratorios y demográficos que han experimentado, pueden promover la adaptación local a las diferentes condiciones ambientales. Por otro lado, los grandes tamaños poblacionales que presentan estos organismos propician la acción de la selección natural, mientras que el intenso flujo génico puede, en algunos casos, retrasar la divergencia adaptativa de las poblaciones, pero en otros, promover la adaptación al introducir alelos ventajosos en otras poblaciones (Nosil, Funk, & Ortiz-Barrientos, 2009).

## Objetivos

### General

- Detectar eventos de selección natural en genes nucleares en un contexto filogenético en el género *Abies*.

### Particulares

- Recopilar, identificar y evaluar la utilidad para estudios de evolución molecular de las secuencias genéticas disponibles para el género *Abies* en la base de datos de GenBank.
- Identificar genes nucleares que se encuentren bajo diferentes tipos de selección natural (direccional, no direccional, negativa).
- Ubicar los eventos de selección en puntos específicos de la historia evolutiva del género.

## **Justificación**

La influencia humana sobre el clima de la tierra es cada vez más evidente y observaciones climáticas han probado la existencia de una tendencia hacia el calentamiento: la temperatura global promedio ha incrementado 0.8°C desde 1900 (Hansen *et al.*, 2006). En un escenario de cambio climático acelerado, investigar la respuesta y los efectos del calentamiento sobre la biodiversidad se ha convertido en uno de los campos más activos de investigación. En el nivel más básico de la biodiversidad, el efecto principal del cambio climático es el de disminuir la diversidad genética de las poblaciones, debido a la acción de la selección direccional y a las migraciones rápidas que comprometen el funcionamiento y resiliencia de los ecosistemas (Bellard, Bertelsmeier, Leadley, Thuiller, & Courchamp, 2012). En este sentido, los árboles forestales son particularmente sensibles, ya que sus largos tiempos generacionales no permiten una rápida adaptación ante los cambios climáticos tan drásticos que se están presentando (Lindner *et al.*, 2010).

La respuesta migratoria de las especies ante el cambio climático puede ser la de seguir a través del espacio las condiciones apropiadas para su supervivencia. La forma más común de hacer esto es por medio de la dispersión, sin embargo, casos más extremos pueden llevar a la necesidad de ocupar otros hábitats. Cambios latitudinales y altitudinales ya han sido observados para más de 1000 especies, y en general estos no solo han implicado un desplazamiento, si no también una reducción de su área de distribución, particularmente en especies polares y de alta montaña. En eventos migratorios, la evolución por selección natural también llega a jugar un papel importante, ya que si bien las especies se desplazan para buscar condiciones climáticas óptimas, estas no necesariamente estarán bien adaptadas a otras variables, tales como el fotoperiodo o nuevas interacciones bióticas (Bellard *et al.*, 2012).

La otra manera en que las poblaciones pueden subsistir al cambio climático es mediante la adaptación a las nuevas condiciones ambientales dentro de su rango local, la cual solo puede darse si existen niveles suficientes de variación genética o de plasticidad fenotípica. Las diferentes especies de coníferas han estado expuestas a lo largo de su historia evolutiva a cambios climáticos a largo plazo y han mostrado

capacidad para adaptarse a ellos. En términos genéticos, podemos entender el cambio climático como un cambio ambiental gradual y direccional que es esporádicamente sacudido por eventos extremos, y las especies deben adaptarse a estos para sobrevivir. En este aspecto, la selección natural subyace a la adaptación local, la cual lleva a las poblaciones hacia niveles más altos de adecuación para acoplarse a los cambios ambientales. Diversos casos de cambios continuos en las frecuencias alélicas o en los valores de ciertos caracteres fenotípicos han sido reportados como resultado de estudios en esta área. Por otro lado, la adaptación local a nivel de especies puede verse promovida por el flujo génico vía el polen proveniente de poblaciones mejor adaptadas a una condición ambiental particular; particularmente en el caso de un cambio hacia temperaturas más altas, es probable que las poblaciones de latitudes menores sean una fuente valiosa de polen para las poblaciones de latitudes mayores (Lindner *et al.*, 2010).

El éxito de las poblaciones actuales para adaptarse a las condiciones climáticas cambiantes, dependerá entonces de los niveles de variación genética que presenten, así como de la tasa con la que surjan nuevas variantes, se mantengan y eventualmente se fijen en las poblaciones. Por lo tanto, nuestra capacidad para entender la dinámica evolutiva de las especies y de detectar las variantes génicas adaptativas presentes en las poblaciones resulta de vital importancia al momento de diseñar modelos predictivos y estrategias de conservación ( Jaramillo-Correa *et al.*, 2015). Este trabajo representa la primera búsqueda de selección a nivel filogenético en el género *Abies*. Los resultados aquí obtenidos, combinados con los de otros trabajos realizados a escala poblacional en este mismo género, no solo ayudarán a entender mejor la historia evolutiva de estos organismos, si no que busca mejorar las medidas implementadas para la conservación de estas especies y otras relacionadas. Esto es una tarea importante, particularmente en un país como México, el cual posee altos niveles de endemismos y una historia biogeográfica compleja.

## **Hipótesis**

- La amplificación de los genes nucleares entre especies solo será exitosa en aquellos que estén más conservados, por lo cual se observará además una tasa de sustituciones no sinónimas generalmente baja.
- Se detectarán eventos de selección positiva en los genes cuya función está asociada a características adaptativas específicas de cada grupo.
- Los distintos tipos de prueba detectarán huellas de selección natural en diferentes genes, dependiendo de la escala a la que haya ocurrido el evento selectivo.

## Método

### *Selección, amplificación de los genes y construcción de bases de datos*

Se seleccionaron y re-secuenciaron 788 genes candidatos identificados en *P. taeda* para ser probados en otras especies de la familia Pinaceae (Eckert *et al*, 2010); la re-secuenciación se realizó en dos especies de *Abies*: *A. pinsapo* y *A. concolor*. Para cada especie se seleccionaron cinco individuos de cinco poblaciones diferentes repartidas a lo largo de su distribución. Se colectaron acículas de estos individuos y a partir de ellas se extrajo el DNA con el kit DNeasy plant de Qiagen. Los detalles de los genes y las condiciones de amplificación fueron los mismos que aquellos reportados en Mosca *et al.* (2012). Posteriormente, se secuenciaron con el método Sanger aquellos genes cuya amplificación fue exitosa en el Centro Nacional de Análisis Genómico, en Barcelona, España. Los cromatogramas resultantes se evaluaron visualmente en 4Peaks (Griekspoor & Groothuis) y aquellos que mostraron una única secuencia limpia (aunque se toleraron hasta cinco SNPs por gen) fueron retenidos y alineados con MUSCLE (Edgar, 2004). Las secuencias de estos genes fueron luego utilizadas para realizar búsquedas en las bases de datos con los transcriptomas disponibles para especies del género *Abies* en GenBank con el algoritmo BLAST (ver Tabla 1).

Se buscó tener las secuencias de al menos una especie por cada sección reconocida dentro del género *Abies* y se realizaron búsquedas de esos mismos genes para las especies *Pseudolarix amabilis* y *Cedrus atlantica* los cuales se utilizaron como grupos externos y para enraizar las filogenias. Estos grupos externos fueron seleccionados debido a la disponibilidad de datos en GenBank y porque se encuentran dentro de la subfamilia Abietoideae, junto con *Abies* (Wang, Tank, & Sang, 2000; Wang & Ran, 2014). En cada caso se seleccionaron las secuencias con mayor porcentaje de cobertura con respecto a la referencia (*A. pinsapo*). Por cuestiones de disponibilidad de datos, la cobertura del grupo externo, el *score* y *E-value* del alineamiento fueron utilizados como referencia; es decir, se retuvieron las secuencias con una cobertura mínima de 20% y un *E-value* no mayor a 1E-10 (ver Anexo 1).

Las secuencias de cada gen se alinearon de manera individual mediante el algoritmo MUSCLE (Edgar, 2004) implementado en MEGA 7 (Kumar, Stecher, & Tamura, 2016). Se utilizaron los parámetros de alineamiento establecidos por defecto, con 8 iteraciones en el proceso y utilizando el método de clustering por UPGMA. Para cada gen se determinó el marco de lectura así como las regiones codificantes a partir de la información disponible en las secuencias anotadas de *Abies alba* en GenBank. Los *indels* que causaron desplazamientos en el marco de lectura (tomando como referencia la secuencia original de *A. pinsapo*) fueron eliminados. Finalmente, se clasificaron los genes en dos grupos, basados en sus anotaciones reportadas previamente por Mosca *et al.* (2012). Por un lado, se agruparon los genes candidatos, identificados por su función biológica a partir de anotaciones hechas en *Arabidopsis thaliana* y en estudios previos de expresión génica en coníferas, en los casos en los que esta información estuvo disponible. También se consideró si estudios poblacionales intra-específicos (por ej. Mosca *et al.*, 2012) ya habían identificado señales de selección en algún gen. El resto de los genes se mantuvieron como grupo control y se consideraron como posiblemente neutrales.

#### *Análisis de diversidad y divergencia*

Para cada gen se analizaron los niveles de polimorfismo dentro del género *Abies*, utilizando dos estimadores,  $\theta_w$  (Watterson, 1975), basado en el número de sitios segregantes, y  $\theta_T$  (Tajima, 1989), basado en el número promedio de diferencias nucleotídicas entre pares de secuencias. Adicionalmente se computó el número de haplotipos ( $h$ ) y se evaluó la divergencia nucleotídica promedio ( $D_{xy}$ ) entre pares de secuencias con respecto al grupo externo (Nei, 1987). Todos estos análisis se realizaron en DNAsp (Librado & Rozas, 2009).

Tabla 1. Número de acceso para los transcriptomas explorados en éste trabajo y número de secuencias retenidas para los análisis posteriores.

Especie	# Secuencias	Acc. number
<i>A. balsamea</i>	60	SRX039631
<i>A. alba</i>	40	JQ440374- JQ445205
<i>A. sachalinensis</i>	45	DRX001582
<i>A. sibirica</i>	15	ERX242658
<i>P. amabilis</i>	65	SRX096935
<i>C. atlantica</i>	65	SRX655858

### *Pruebas de neutralidad*

A partir de las anotaciones de los genes retenidos para los análisis, así como de la división realizada entre candidatos y controles, se buscó encontrar huellas de selección. En estudios previos realizados en coníferas a nivel poblacional, se han encontrado huellas de selección en genes asociados a dos tipos principales de funciones. Por un lado, genes reguladores de procesos de desarrollo y fenológicos, y por otro, genes involucrados en respuestas a diferentes tipos de estrés como el oxidativo o por congelamiento (Mosca *et al.*, 2012; Namroud *et al.*, 2010). Considerando esto, se esperó encontrar huellas de selección a nivel filogenético en genes con funciones relacionadas a las anteriores.

Inicialmente se utilizó la  $D$  de Tajima, implementada en DNAsp (Librado & Rozas, 2009), de manera individual para cada gen; se consideraron todas las mutaciones y se excluyeron los sitios con indels. Al tratarse de genes codificantes, se calculó un valor total de  $D$ , así como los valores para los sitios sinónimos y no sinónimos por separado

(únicamente para aquellos genes en los que se logró determinar las regiones codificantes). Los valores críticos de  $D$  para determinar su significancia estadística fueron calculados a partir de una distribución beta para los valores de este estadístico (Tajima, 1989).

Posteriormente se implementaron dos pruebas combinadas. La prueba HKA (Hudson *et al.*, 1987) y la prueba MK (McDonald & Kreitman, 1991), ambas también en DNAsp (Librado & Rozas, 2009). Para el primer test, se ensamblaron matrices con dos genes cada una para poder evaluar la divergencia. Cada matriz constó del gen 0\_6448\_02 y uno de los genes restantes, de tal manera que el primer gen fue comparado con cada uno de los genes restantes, resultando en 22 pares de genes analizados. Se seleccionó el gen 0\_6448\_02 como referente para la comparación por presentar un valor de la  $D$  de Tajima no significativo y cercano al promedio de todos los genes usados, lo que indica una posible neutralidad. Sin embargo es importante señalar que al tratarse de genes candidatos es difícil garantizar la ausencia total de selección. Se definió el grupo interespecífico con las secuencias de *Cedrus* y de *Pseudolarix*, mientras que el grupo intraespecífico contenía todas las secuencias de *Abies*. Los sitios con gaps fueron omitidos. En cada caso se realizó una prueba de  $\chi^2$  para determinar si las diferencias entre los parámetros  $\Theta$  estimados para cada gen fueron significativas.

La prueba MK se realizó para cada gen de manera individual y solo en aquellos casos en las que la región codificante pudo ser asignada con seguridad. Los grupos inter e intraespecífico se definieron de la misma manera que para la prueba HKA, y también se excluyeron los sitios con gaps. En cada caso se utilizó una prueba exacta de Fisher y una prueba  $G$  para determinar si las diferencias entre los coeficientes de sustituciones sinónimas y no sinónimas fueron significativas.

Para todas las pruebas a nivel filogenético se utilizó una filogenia modificada a partir de las reportadas previamente para *Abies* (Xiang *et al* 2015), obtenidas en dicho estudio mediante métodos de máxima verosimilitud (ML) e inferencia bayesiana (BI) con información de marcadores nucleares (*nrITS*), mitocondriales (*ssu-rRNA* y *nad5* intron 1), y de cloroplasto (*rpl16*, *rps12-rpl20* y *trnT-F*). Para los análisis realizados en el presente trabajo se seleccionó aquella que presentó una mejor resolución de las

relaciones entre secciones y que no incluía incongruencias debidas a introgresión o retención de caracteres ancestrales (Figura 1).

Las pruebas a nivel filogenético realizadas fueron el *Relative Rate Ratio (RRR) test* (Creevey & McInerney, 2002) implementado en CRANN versión 1.04 (Creevey & McInerney, 2003) y los modelos evolutivos por codones mediante los métodos de máxima verosimilitud implementados en el programa PAML versión 4.9 (Yang, 2007). Estos análisis fueron realizados exclusivamente sobre las zonas codificantes de los genes retenidos en las pruebas anteriores.

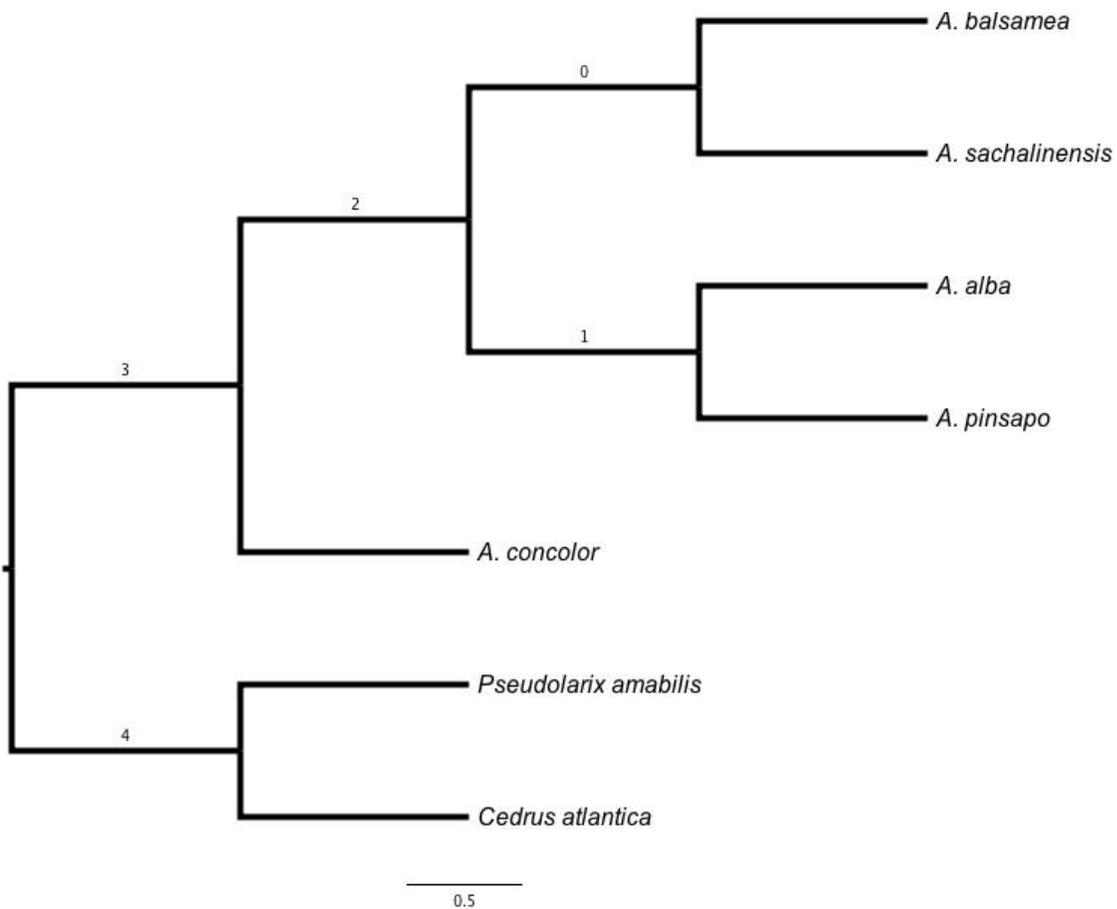


Figura 1. Filogenia del género *Abies* modificada a partir de la reportado por Xiang *et al.*, 2015 y utilizada para el análisis en CRANN y PAML. La numeración corresponde a la asignada para cada rama interna como referencia por CRANN; esta no representa ningún resultado de ninguna prueba.

Para el *RRR* test, se analizó cada gen de manera individual y en cada caso se utilizó la misma filogenia (figura 1). La significancia estadística fue determinada mediante dos pruebas: la prueba exacta de Fisher y la prueba *G*. En cada caso y por sugerencia del manual de uso, se tomó como valor verdadero el de la prueba de Fisher, y solo cuando esta resultaba imposible de calcular se utilizó el valor de la prueba *G* (Creevey & McInerney, 2002). Como parte del análisis, CRANN realiza adicionalmente un *Neutral Substitution Test*, el cual calcula el número total de sustituciones de reemplazo y sinónimas observadas dentro de cada clado de la filogenia a partir de los valores de *SI*, *SV*, *RI* y *RV* calculados para el *RRR* test. Esta prueba a su vez estima el número esperado de cada tipo de sustitución, basado en el número de cada tipo de sitio dentro de la secuencia. Esto permite inferir selección negativa, cuando el número observado de sustituciones de reemplazo es menor que el esperado, y selección positiva, cuando el número observado de sustituciones de reemplazo es mayor que el esperado, pero sin distinguir si se trata de selección direccional o no direccional (Creevey & McInerney, 2002). La significancia de este test se determinó de la misma manera que para el *RRR* test.

Para los modelos evolutivos por codones se utilizó la misma filogenia utilizada en los análisis en CRANN. En primer lugar, se evaluó para cada gen dos modelos por rama (Yang, 1998). Estos incluyeron un modelo nulo con un solo coeficiente  $\omega$  para todas las ramas de la filogenia y un modelo libre, que considera un coeficiente  $\omega$  variable entre las diferentes ramas de la filogenia (ver Introducción). Adicionalmente, se probó un modelo con dos valores de  $\omega$  para los genes que mostraron significancia en los análisis de CRANN. En estos análisis se probó específicamente un cambio en el valor de  $\omega$  en las ramas en las que el test *RRR* indicó un cambio en la dirección de la tasa de selección (ver resultados).

Finalmente se evaluaron dos pares de modelos evolutivos por sitio para cada gen por separado. Primero, se comparó un modelo neutro con un modelo de selección (Rasmus Nielsen & Yang, 1998); luego, se probó el modelo M7 contra el M8 (Yang *et al.*, 2000) (ver Introducción para más detalles sobre estos modelos). En todos los casos (modelos por rama y por sitio) los parámetros  $\kappa$  y  $\omega$  fueron estimados por

máxima verosimilitud, y las frecuencias en equilibrio de los codones fueron estimadas a partir de las frecuencias nucleotídicas promedio en las tres posiciones de codón (opción F3X4 en el programa). Igualmente, en todos los casos, se compararon los valores de verosimilitud de los modelos mediante un *Likelihood Ratio Test (LRT)* para determinar cuál modelo explicaba mejor los datos. En cada caso (modelos por rama y por sitio) el número de grados de libertad se determinó a partir de la diferencia en el número de parámetros de cada modelo.

## **Resultados**

### *Obtención de las secuencias*

De los 788 genes iniciales, 95 amplificaron exitosamente en *A. concolor* y 84 en *A. pinsapo*. Después de revisar los cromatogramas de estos genes, se seleccionaron los 67 genes que amplificaron exitosamente en ambas especies (Anexo 1). Con las búsquedas con el algoritmo BLAST se obtuvieron secuencias putativamente ortólogas para *A. alba*, *A. balsamea*, *A. sachalinensis*, *A. sibirica*, *Pseudolarix amabilis* y *Cedrus atlantica* para buena parte de los 67 genes retenidos. Sin embargo, cabe anotar que para *A. sibirica* la mayoría de las búsquedas se arrojaron fragmentos con una cobertura menor al 20% de la secuencia original y con valores del alineamiento menores a 40; este pobre ensamblado nos llevó a descartar todas las secuencias de esta especie para los análisis posteriores. Finalmente, luego de aplicar los parámetros necesarios para retener un gen (un mínimo de 20% de cobertura con respecto a la secuencia original y un E-value no mayor a 1E-10), se seleccionaron 23 genes para los análisis posteriores. Las secuencias de estos genes fueron alineadas para generar el mismo número de bases de datos con las que se hicieron los cálculos de diversidad genética. Para tres de los genes retenidos, 0\_6566\_02, 0\_7921\_01, 2\_6313\_01 y 2\_10216\_01, no se logró asignar con seguridad una región codificante ni un marco de lectura, por lo cual fueron omitidos de las pruebas para las que esta información es necesaria (*MK*, *RRR* test y análisis en PAML).

Las 23 secuencias seleccionadas para los análisis presentaron una longitud promedio de 414 bp, con una longitud máxima de 658 bp y una mínima de 146 bp. Su contenido GC promedio fue de 42.5%, yendo desde 35.06% (gen 0\_7921\_01) hasta 56.65% (gen 0\_12329\_02). De estas secuencias, 6 se clasificaron como controles, putativamente neutrales, debido a la falta de información sobre su función o similitud con alguna secuencia con función conocida. Los 17 restantes, se mantuvieron como genes candidatos. Las funciones de estos últimos incluyeron proteínas de unión a nucleótidos y otras proteínas, diferentes cinasas y proteínas con actividad de oxidorreductasas y catalíticas; por ejemplo, en síntesis de sucrosa y beta-D-glucanos (Anexo 3).

#### *Diversidad genética*

Todos los genes retenidos fueron variables entre especies. El número de haplotipos por gen para el género *Abies* estuvo entre dos (UMN\_1337\_02) y cinco (0\_7921\_01 y 2\_10216\_01; Tabla 2). Los estimados de  $\theta_T$  y de  $\theta_W$  fueron muy heterogéneos; los valores de  $\theta_T$  más bajos fueron de  $2,32 \times 10^{-3}$  para el gen 0\_16009\_02 y los más altos de 0.13796 para el gen CL708Contig1\_02; mientras que para  $\theta_W$  el valor mínimo fue de  $2,78 \times 10^{-3}$  (0\_16009\_02) y el máximo de 0.12675 (0\_12329\_02). Los valores de divergencia  $D_{xy}$  (obtenidos comparando al género *Abies* con los dos outgroups) también fueron bastante heterogéneos, yendo desde 0.0219 (UMN\_495\_01) hasta 0.15172 (0\_7921\_01).

Tabla 1. Estimados de diversidad genética ( $h$ ,  $\theta_T$  y  $\theta_w$ ), divergencia (Dxy) y de tests de selección (D, Ds y Dns) para 23 genes entre cinco especies de *Abies*. \*\*\*significancia  $p < 0.001$ . Ver anexo para la anotación y otras informaciones adicionales de cada gen.

	DIVERSIDAD			Dxy	D	Ds	Dns	Longitud (bp)
	h	$\theta_T$	$\theta_w$					
0_6448_02	3	0.00758	0.0084	0.02667	-0.66823	0.24314	-1.0938	404
0_6566_02	4	0.01617	0.01634	0.06289	-0.07339	NA	NA	521
0_7921_01	5	0.09778	0.112	0.15172	-0.94025	NA	NA	288
0_8531_01	3	0.00401	0.00413	0.04013	-0.17475	-0.17475	NA	472
0_12021_01	3	0.00497	0.00596	0.03173	-0.97256	-0.8165	-0.8165	337
0_12190_01	4	0.08119	0.09743	0.04958	-1.26241	-1.25619	-1.23588	432
0_12329_02	4	0.10736	0.12675	0.07853	-1.15755	-1.24874	-1.24046	264
0_16009_02	3	0.00232	0.00278	0.04008	-0.97256	-0.97256	NA	416
1_6493_01	4	0.0411	0.04932	0.06181	-1.21852	-1.18441	-1.14554	329
2_2937_01	4	0.01939	0.02327	0.04032	-1.22992	-1.16172	-1.18441	404
2_6313_01	4	0.00943	0.01131	0.04951	-1.16172	NA	NA	569
2_7121_02	3	0.03739	0.04383	0.0439	-1.08716	-0.8945	-1.19955	418
2_7803_01	4	0.02561	0.02488	0.02417	0.21615	-0.41017	0.4362	379
2_10216_01	5	0.01907	0.02223	0.0246	-1.04473	NA	NA	428
CL516Contig1_07	4	0.06543	0.07901	0.04257	-1.2925***	-1.29386***	-1.04849	388
CL708Contig1_02	3	0.13796	0.11111	0.0374	1.82369	-0.8165	NA	546
CL813Contig1_03	4	0.05455	0.05455	0.02857	0	0.59633	1.12397	384
CL1004Contig1_08	3	0.01295	0.01036	0.03571	1.57274	1.22474	NA	658
CL2332Contig1_01	4	0.07919	0.06443	0.04554	1.69413	1.32768	1.80808	387
CL4354Contig1_01	3	0.0094	0.00966	0.03926	-0.17475	-0.97256	NA	399
UMN_495_01	4	0.00738	0.00787	0.0219	-0.41017	-0.17475	-0.8165	561
UMN_1337_02	2	0.08511	0.10213	0.05556	-1.24614	-1.22485	-1.21039	146
UMN_5272_01	4	0.01765	0.01882	0.04375	-0.44037	1.45884	-1.14554	402

### Pruebas de neutralidad

Los valores de la  $D$  de Tajima fueron negativos y no significativos para la mayoría de los genes (Tabla 2). Solo en un caso se obtuvo un valor significativo para el gen CL516Contig1\_07, que mostró un exceso de variantes de baja frecuencia y un valor de  $D$  para los sitios sinónimos menor que el de los sitios no sinónimos (Tabla 2); esto indicaría una posible acción de la selección purificadora o una expansión poblacional.

En cuatro casos más se registraron valores positivos, pero no significativos (2\_7803\_01, CL708Contig1\_02, CL1004Contig1\_08 y CL2332Contig1\_01).

La prueba HKA no mostró diferencias significativas en los estimados de  $\theta$  entre las regiones genómicas comparadas en cada uno de los 22 pares de genes analizados (ver tabla 3). Por otro lado, la prueba MK indicó que en la mayoría de los casos había más mutaciones sinónimas variables que fijas, así como más variantes no sinónimas variables que fijas (tabla 4); sin embargo, en ningún caso estas diferencias fueron significativas, por lo cual tampoco se sugiere evidencia de algún tipo de selección.

**Tabla 2. Valores de  $\theta$  para cada uno de los 22 pares de genes analizados con la prueba HKA entre especies de *Abies*. En ningún caso se encontraron diferencias significativas.**

LOCI COMPARADOS	$\theta$ 1	$\theta$ 2
6448/6566	0.00669	0.01606
6448/7921	0.01362	0.09907
6448/8531	0.00591	0.00646
6448/12021	0.00808	0.00824
6448/12190	0.01343	0.05403
6448/12329	0.01588	0.13647
6448/16009	0.00523	0.00617
6448/6493	0.01295	0.06881
6448/2937	0.00732	0.00896
6448/6313	0.00709	0.01068
6448/7121	0.01365	0.04767
6448/7803	0.00814	0.00696
6448/10216	0.00982	0.01209
6448/CL516	0.01413	0.06204
6448/CL708	0.00713	0.0067
6448/CL813	0.01063	0.02482
6448/CL1004	0.00589	0.00497
6448/CL2332	0.0067	0.00811
6448/CL4354	0.00796	0.00815
6448/U495	0.00637	0.00305
6448/U1337	0.01546	0.09317
6448/U5272	0.01098	0.028

Tabla 3. Resultados de la prueba MK entre especies de *Abies*. Número de variantes sinónimas fijas y variables y no sinónimas fijas y variables para cada gen. Ningún caso fue significativo.

	Sin Fij	Sin Var	NoSin Fij	NoSin Var
0_6448_02	2	8	0	4
0_8531_01	1	11	2	9
0_12021_01	1	2	1	0
0_12190_01	1	14	0	5
0_12329_02	1	29	0	21
0_16009_02	1	8	3	5
1_6493_01	2	11	0	7
2_2937_01	1	12	0	8
2_7121_02	0	12	0	12
2_7803_01	0	4	0	3
CL516Contig1_07	0	17	0	2
CL708Contig1_02	1	4	2	0
CL813Contig1_03	0	4	0	4
CL1004Contig1_08	1	4	0	0
CL2332Contig1_01	2	3	1	0
CL4354Contig1_01	2	6	0	1
UMN_495_01	1	3	0	1
UMN_1337_02	0	20	0	13
UMN_5272_01	0	5	0	6

*Pruebas filogenéticas: RRR test y Neutral Substitution Test*

Para la mayoría de los genes no se encontró evidencia de selección natural en ninguna de las ramas de la filogenia. Solo en el gen 0\_6448\_02 mostró evidencia de selección ni direccional, ya que se observó un número significativamente mayor de sustituciones no sinónimas variables sobre la rama 3 de la filogenia. (Fisher's  $p= 0.035291$ ; Figura 1). Adicionalmente, se detectaron dos casos significativos con la prueba  $G$ : por un lado,

selección no direccional para el gen 0\_8531\_01 sobre la rama interna que define al grupo *A.Balsamea-sachalinensis-alba-pinsapo* (rama 2). Por otro lado, se infirió selección direccional en el gen CL2332Contig1\_01 sobre la rama interna que define al grupo *Balsamea-Sachalinensis* (rama 0). Cabe anotar que estos dos casos deben ser confirmados en estudios futuros ya que solo se detectó significancia con la prueba *G* con valores bajos en al menos alguna de las cuatro categorías de mutaciones.

El *Neutral Substitution Test* detectó nueve genes para los cuales el número observado de sustituciones de reemplazo fue significativamente menor de lo esperado y por lo tanto, en los cuales estaría actuando la selección negativa. Este patrón fue en general consistente, presentándose en la mayoría de los casos sobre las ramas 0, 2 y 3 (fig. 1; Tabla 5).

Tabla 4. Resultados del *Neutral Substitution Test* para especies de *Abies* en un contexto filogenético. Solo se indican los genes y ramas en donde se encontró un número mayor (selección positiva) o menor (selección negativa) de sustituciones de reemplazo con respecto al número esperado en cada caso.

	Rama de la filogenia				
	0	1	2	3	4
<b>0_6448_02</b>	-	-	-	-	neg
<b>0_12190_01</b>	neg	-	neg	neg	-
<b>0_12329_02</b>	neg	-	neg	neg	-
<b>1_6493_01</b>	-	-	-	neg	-
<b>2_7121_02</b>	-	-	-	-	neg
<b>CL516Contig1_07</b>	neg	-	neg	neg	-
<b>CL813Contig1_03</b>	neg	-	neg	neg	-
<b>UMN_495_01</b>	-	-	-	neg	-
<b>UMN_1337_02</b>	neg	-	neg	neg	-

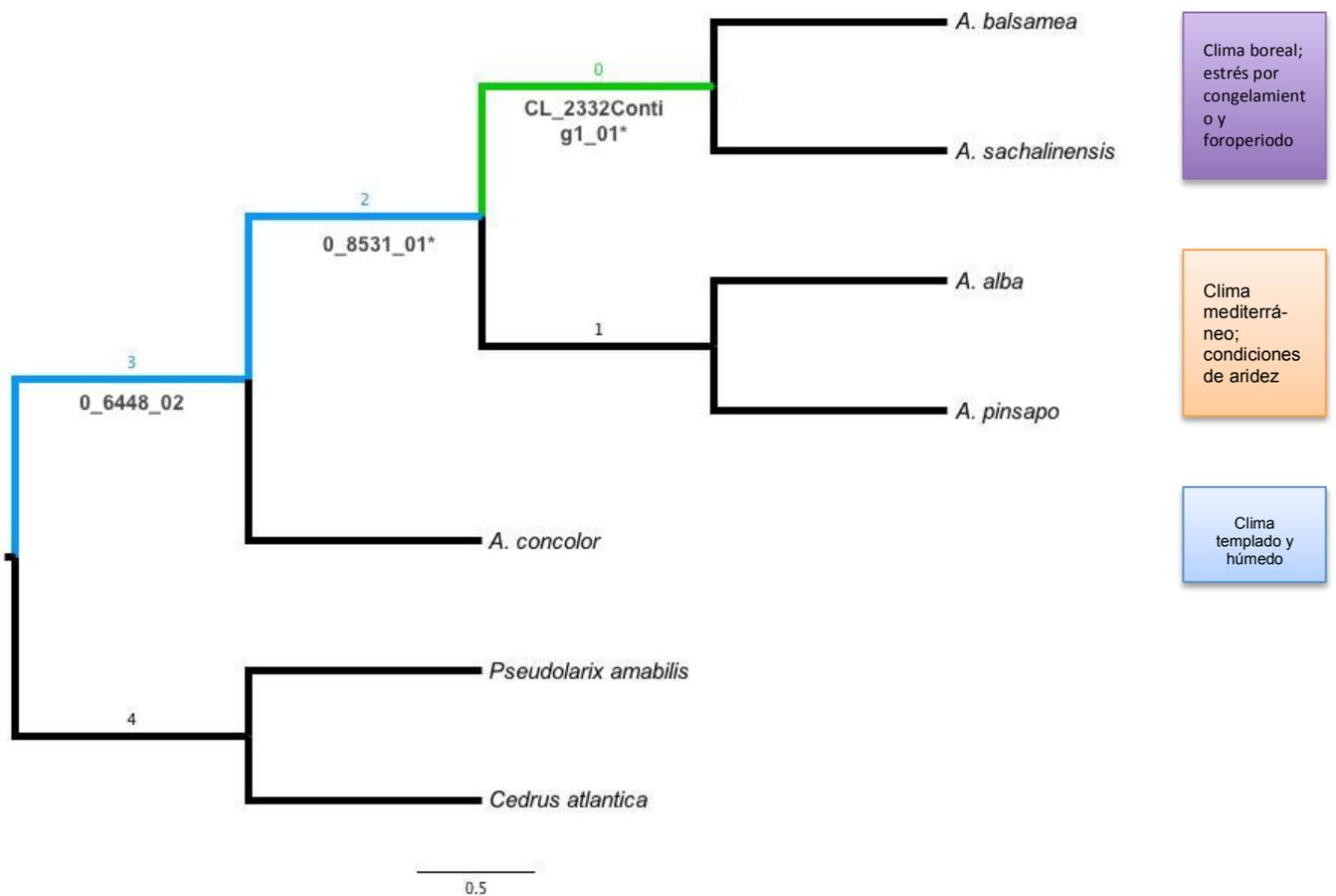


Figura 1. Filogenia utilizada en los análisis de CRANN, modificada a partir de Xiang *et al.* (2014). En azul se muestra la rama sobre la que se encontró evidencia de selección balanceadora para los genes 0\_6448\_02 (rama 3) y 0\_8531\_01\* (rama 2). En verde se muestra la rama que presentó evidencia de selección direccional en el gen CL2332Contig1\_01\* . \*Significancia solamente con la prueba G. Los recuadros de colores indican las condiciones ambientales generales de cada grupo.

### Pruebas filogenéticas: modelos evolutivos por codones

Para ninguno de los genes se encontraron diferencias significativas en los modelos por rama entre el modelo con un solo coeficiente  $\omega$  y el modelo libre (i.e. con tantos valores de  $\omega$  como ramas en el árbol). De igual manera, el modelo con dos coeficientes probado para los genes 0\_6448\_02, 0\_8531\_01 y CL2332Contig1\_01, en las ramas indicadas en cada caso por el RRR test, no explicó los datos significativamente mejor que el modelo con un solo coeficiente (Tabla 6). Sin embargo, se encontró un patrón

general consistente con selección negativa, ya que los coeficientes  $\omega$  a través de la filogenia, fueron menores a uno para todos los genes de acuerdo con el modelo nulo o de un solo coeficiente (Tabla 6).

Tabla 5. Valores de verosimilitud (log) para los modelos por rama con un (uno) o dos coeficientes (dos) y para el modelo libre (libre; tantos valores de  $\omega$  como ramas en la filogenia). En ningún caso el LRT indicó diferencias significativas entre los modelos.  $\omega_0$  corresponde a los valores obtenidos por el modelo nulo (de un solo coeficiente) para cada gen.

	Modelo de evolución			
	Uno	Dos	Libre	$\omega_0$
<b>0_6448_02</b>	-390.45621	-390.456188	-386.141698	0.16415
<b>0_8531_01</b>	-514.097736	NA	-512.151006	0.20108
<b>0_12021_01</b>	-150.831981	NA	-149.972392	0.17113
<b>0_12190_01</b>	-230.599814	NA	-229.553578	0.06952
<b>0_12329_02</b>	-413.007601	NA	-409.500957	0.10728
<b>0_16009_02</b>	-414.460292	NA	-410.719965	0.35266
<b>1_6493_01</b>	-275.167932	NA	-270.646436	0.13476
<b>2_2937_01</b>	-449.381571	NA	-443.785568	0.23556
<b>2_7121_02</b>	-332.526857	NA	-328.639647	0.33961
<b>2_7803_01</b>	-179.328294	NA	-179.219864	0.1099
<b>CL516Contig1_07</b>	-188.308735	NA	-188.298725	0.00114
<b>CL708Contig1_02</b>	-194.788489	NA	-192.527805	0.15363
<b>CL813Contig1_03</b>	-174.455159	NA	-171.483994	0.50272
<b>CL1004Contig1_08</b>	-128.169047	NA	-128.168984	0.0001
<b>CL2332Contig1_01</b>	-164.861583	-164.698064	-163.941338	0.06571
<b>CL4354Contig1_01</b>	-213.230285	NA	-213.230325	0.0001
<b>UMN_495_01</b>	-206.134518	NA	-204.959255	0.05224
<b>UMN_1337_02</b>	-289.494445	NA	-288.410429	0.10523
<b>UMN_5272_01</b>	-184.254529	NA	-177.463005	0.21298

Por otro lado, los modelos por sitio tampoco lograron detectar selección en ningún caso, ya que los modelos de selección y M8, no fueron significativamente mejores que los modelos neutral y M7, respectivamente (tabla 7). Sin embargo, en el gen 2\_7121\_02 el análisis BEB (Bayes Empirical Bayes) implementado como parte de los análisis por sitio, logró detectar dos sitios posiblemente bajo selección positiva ( $\omega > 1$ ) con una probabilidad posterior de más de 95% en las posiciones 45 y 47 de la secuencia. Ambos cambios son de reemplazo y sustituyen una prolina por una leucina y por una metionina, respectivamente.

Tabla 6. Valores de verosimilitud (log) obtenidos para los modelos por sitio para cada gen. Modelo neutral (NEU), modelo de selección positiva (POS) y modelos M7 y M8. En ningún caso se detectaron diferencias significativas entre los modelos.

	NEU	POS	M7	M8
<b>0_6448_02</b>	-390.456236	-390.45621	-390.460976	-390.461002
<b>0_8531_01</b>	-514.097763	-514.097736	-514.103839	-514.103866
<b>0_12021_01</b>	-150.831982	-150.832228	-150.832075	-150.832076
<b>0_12190_01</b>	-230.599863	-230.599814	-230.603081	-230.603457
<b>0_12329_02</b>	-412.905499	-412.905499	-412.042784	-412.042835
<b>0_16009_02</b>	-414.460299	-414.460292	-414.462605	-414.462612
<b>1_6493_01</b>	-270.646436	-269.734597	-270.835527	-269.693856
<b>2_2937_01</b>	-446.796129	-446.291214	-446.799248	-446.291213
<b>2_7121_02</b>	-327.352833	-325.53677	-327.359554	-325.53677
<b>2_7803_01</b>	-179.328334	-179.328294	-179.33501	-179.335048
<b>CL516Contig1_07</b>	-182.420246	-182.420246	-183.371098	-182.416524
<b>CL708Contig1_02</b>	-194.788502	-194.788489	-194.79061	-194.790623
<b>CL813Contig1_03</b>	-173.758999	-173.405349	-173.759734	-173.405348
<b>CL1004Contig1_08</b>	-128.168157	-128.168079	-128.16807	-128.168148
<b>CL2332Contig1_01</b>	-164.861608	-164.861583	-164.863643	-164.863668
<b>CL4354Contig1_01</b>	-213.227084	-213.226857	-213.226823	-213.22705
<b>UMN_495_01</b>	-206.134408	-206.13438	-206.136048	-206.136075
<b>UMN_1337_02</b>	-281.751947	-281.751947	-281.753268	-281.751899
<b>UMN_5272_01</b>	-183.751408	-183.751408	-184.010412	-183.668241

## Discusión

De los 788 genes inicialmente amplificados para *A. pinsapo* y *A. concolor*, solo lograron amplificarse 67 exitosamente en ambas especies. Cabe recordar que los cebadores de estos genes fueron transferidos desde *Pinus taeda* y que estos 67 son probablemente los mejor conservados entre especies de coníferas, lo que es consistente con el patrón encontrado en los análisis realizados en este trabajo, en donde predomina la acción de la selección negativa sobre los genes estudiados. Adicionalmente, la dificultad para encontrar una cobertura adecuada en el transcriptoma de al menos una de las especies de *Abies* utilizadas y/o para los dos grupos externos también puede deberse a que solo aquellos genes mejor conservados pudieron encontrarse entre especies. Esta limitación inicial puede llevar a estimaciones sesgadas hacia niveles de diversidad y divergencia más bajos de lo esperado a nivel genómico, así como a dificultades para detectar eventos de selección positiva y una predominancia de selección negativa.

Las funciones de los genes retenidos se relacionan principalmente con aspectos regulatorios de eventos del desarrollo y la fenología. Estos codificaron para proteínas cinasas y de unión a nucleótidos y proteínas (DCL1, CDPK6, proteínas Argonautas y SRF8), así como proteínas involucradas en procesos de oxidorreducción y síntesis de azúcares (sucrosa sintasa, alfa-trehalosa-fosfato sintasa, 1-3-beta-glucano sintasa, etc.). Todas ellas son funciones generales o esenciales para todas las plantas, por lo que no es sorprendente que los *primers* para amplificarlas sean fáciles de transferir entre especies y/o que sus secuencias se encuentren relativamente conservadas entre especies.

### *Diversidad y divergencia*

Los niveles de diversidad y divergencia encontrados en los genes analizados fueron en general bajos y heterogéneos. Esto corresponde con lo encontrado en estudios previos, los cuales reportan niveles igualmente bajos de divergencia génica entre

especies de coníferas (Bouillé & Bousquet, 2005; Mosca *et al.*, 2012). De manera similar, los niveles de diversidad intraespecífica encontrados en coníferas se consideran bajos con respecto a lo esperado dadas sus características, como sus grandes tamaños poblacionales y altos niveles de entrecruzamiento (Eckert *et al.*, 2013; Heuertz *et al.*, 2006; Pyhäjärvi *et al.*, 2007). Se han propuesto diferentes explicaciones para estos resultados. Por un lado, es posible que los grandes eventos demográficos, tales como las migraciones y los cuellos de botella experimentados por las coníferas a lo largo de su historia evolutiva, hayan sido un factor determinante en los niveles y la distribución de la variación genética que estas presentan. Como resultado de estos eventos, las poblaciones de coníferas presentan también bajos niveles de diferenciación poblacional, en algunos casos de la misma magnitud que la encontrada a nivel interespecífico (Bouillé & Bousquet, 2005), así como mayores niveles de polimorfismo al interior de las poblaciones que entre ellas (Namraud *et al.*, 2012; Wang *et al.*, 2014; Pyhajarvi *et al.*, 2007). En este aspecto, es importante considerar que las coníferas son particularmente susceptibles a estas fluctuaciones poblacionales, ya que debido a sus largos tiempos generacionales, el tiempo requerido por estas para recuperar los niveles de diversidad en equilibrio son demasiado largos y por lo tanto, los genomas de estos organismos retienen las huellas de los eventos demográficos por más tiempo (Pyhäjärvi *et al.*, 2007; Savolainen & Pyhäjärvi, 2007). Por otro lado, se ha establecido que las coníferas presentan bajas tasas de mutación, en algunos casos menores en hasta un orden de magnitud que las de las angiospermas (Brown *et al.*, 2004). Las bajas tasas de mutación darían cuenta tanto de los bajos niveles de diversidad intraespecífica así como de la poca divergencia interespecífica. Sin embargo, otros autores han obtenido tasas más altas y se ha sugerido que los tiempos de divergencia en coníferas han sido sobreestimados, lo cual descartaría las bajas tasas de mutación como causas de la baja diversidad genética de las coníferas (Bouillé & Bousquet, 2005; Willyard *et al.*, 2007). En los resultados de este trabajo a nivel interespecífico, la baja diversidad encontrada posiblemente esté relacionada con el sesgo hacia genes conservados con los que se trabajó, lo cual a su vez implicaría que la selección negativa ha estado eliminando de manera consistente la variación en estos genes.

Los valores principalmente negativos (aunque no significativos) de la  $D$  de Tajima obtenidos para la mayoría de los genes también sugieren la acción recurrente de la selección negativa sobre ellos. Valores similares se han encontrado a nivel poblacional, pero en estos casos se ha interpretado como consecuencia de eventos demográficos (Mosca *et al.*, 2012; Namroud *et al.*, 2008; Pyhäjärvi *et al.*, 2007). El único caso significativamente negativo fue el gen CL516Contig1\_07, el cual además presentó menor variación en los sitios sinónimos, lo cual sugiere un escenario de barrido selectivo parcial para este gen, en el cual se estarían manteniendo bajos niveles de variación neutral alrededor de los sitios seleccionados positivamente. Este gen codifica a la proteína AVP1, una pirofosfatasa transportadora de hidrógeno, relacionada con el aumento en el crecimiento vegetal bajo diferentes tipos de estrés abiótico, por lo cual es plausible que haya sufrido eventos de selección positiva al interior del género *Abies* (Schilling *et al.*, 2017).

La utilización de una muestra no aleatoria de genes, puede haber llevado a estimados sesgados de diversidad y divergencia, así como a valores generalmente negativos de la  $D$  de Tajima, debido al grado de conservación de los genes analizados. Este tipo de sesgo deriva en general del proceso de identificación de los polimorfismos (ascertainment bias), y es posible que este esté presente en otros estudios en coníferas tanto a nivel interespecífico como a nivel poblacional. En el futuro, la utilización de secuenciación de nueva generación (NGS) podría ayudar a eliminar este sesgo, ya que no depende de amplificaciones con blancos específicos mediante primers transferibles entre especies ni de la utilización de genomas externos de referencia (Eckert *et al.*, 2013).

#### *Patrones de selección a nivel filogenético*

Los resultados de las pruebas de selección a nivel filogenético indican un papel predominante de la selección negativa en los genes analizados. Los valores de  $\omega$  para todos los genes estimados por el modelo por sitios de un solo coeficiente fueron menores que uno en todos los casos, lo cual sugiere que estos se encuentran bajo un régimen de selección negativa. Por su parte el *Neutral Substitution Test* detectó huellas

de este tipo de selección en 9 de los 19 genes analizados (47%). En comparación, 15 de 22 genes (68%) fueron detectados como sujetos de selección negativa en un análisis con este mismo test en una escala filogenética amplia, mientras que solo 6% de los 21 genes analizados en este mismo género a una escala taxonómica más reducida (solo especies mediterráneas) presentó evidencia de este tipo de selección mediante esta prueba (Grivet *et al.*, 2013; Palmé *et al.*, 2009). En estos estudios, así como en el presente trabajo, se utilizaron muestras no aleatorias de genes. Sin embargo, es posible que las diferencias en el porcentaje de genes sujetos a selección negativa se deba principalmente a la diferente cobertura filogenética entre los estudios. En el primer caso se tuvo una buena representación de los diferentes grupos de pinos, mientras que en el segundo, solo se utilizaron especies mediterráneas. En el presente trabajo, si bien se buscó tener representación de todas las secciones del género *Abies*, esto no se logró completamente, principalmente debido a la poca disponibilidad de secuencias para este género en las bases de datos. Debido a esto, es difícil comparar estos resultados con los de otros estudios; sin embargo, al menos en el primer caso, se coincide en que la selección negativa tiene un papel predominante en un porcentaje importante de los genes analizados.

De los nueve casos de selección negativa detectados por el *RRR* test para *Abies*, cinco se presentaron sobre las ramas 0, 2 y 3 de la filogenia, las cuales son ramas secuencialmente derivadas una de otra. En estos casos, es posible que los efectos de las ramas externas se reflejen en las ramas más internas como efecto residual de la adición de los coeficientes los cuales se van sumando desde las ramas más externas hacia las más internas (Creevey & McInerney, 2002). Por lo tanto, aunque es posible que la selección negativa haya actuado de manera constante a lo largo de todas las ramas que conducen a la sección *Balsamea* en estos 5 casos, también es probable que el evento de selección negativa sea exclusivo de la rama externa que conduce a este grupo. Se puede considerar que en esta situación, el primer caso sea el más probable, debido en parte a que se encontraron otros casos de selección negativa en ramas específicas que no afectaron los cálculos de las ramas vecinas. Las especies de la sección *Balsamea* son las que tienen una distribución más boreal dentro del género

*Abies*, y pueden estar sujeta a diferentes tipos de estrés específicos, tales como congelamiento, desecación y variación en el fotoperiodo que las diferencian de las demás. Algunas de las funciones de estos cinco genes candidatos podrían estar relacionadas a la adaptación a estas condiciones, por lo cual no es sorprendente que estas hayan estado constreñidas a lo largo de la evolución de esta sección. Esto podría comprobarse con estudios de adaptación a nivel local en el futuro.

Por otro lado, se detectaron señales de selección positiva en el gen 0\_6448\_02, el cual codifica a la proteína DCL1 (*Dicer-like protein 1*). De acuerdo con el *RRR* test, este gen presentó un valor significativamente mayor de sustituciones no sinónimas variables en la rama que conduce a todo el grupo de *Abies*, lo cual sugiere la acción de selección no direccional actuando en todo el género. Es importante señalar que de acuerdo con el *Neutral Substitution Test*, el mismo gen mostró evidencia de selección negativa en la rama que define al grupo externo, lo cual refuerza la evidencia del cambio de régimen selectivo dentro del grupo de *Abies*. El *RRR* test detectó otros dos casos cuya significancia solo fue detectada por la prueba *G*, por lo cual serían necesarios tests adicionales con secuencias de otras especies para confirmar la tendencia. El gen 0\_8531\_01, que codifica una proteína de unión a iones de cobre, presentó señales de selección no direccional sobre la rama que define al grupo *A.balsamea-sachalinensis-alba-pinsapo* (rama 2); mientras que el gen CL2332Contig1\_01, que codifica a la proteína CDPK6, presentó evidencia de selección direccional sobre la rama 0 (*Balsamea-Sachalinensis*). Es interesante señalar que, en estos dos casos, el modelo libre en PAML detectó resultados que, si bien no fueron significativos, sí reflejaron una gran heterogeneidad del coeficiente  $\omega$  entre ramas, particularmente en aquellas sobre las cuales el *RRR test* detectó casos significativos, reforzando la necesidad de hacer pruebas complementarias con secuencias de especies adicionales. En estos casos, las ramas contenidas dentro de los grupos definidos por cada rama detectada por el *RRR* test presentaron valores de  $\omega$  diferentes a los de las ramas de los otros grupos. Por ejemplo, para el gen 0\_6448\_02, los coeficientes  $\omega$  de todas las ramas contenidas dentro del grupo definido por la rama 3 (detectada por el *RRR* test), mostraron valores mayores que uno (señal de selección positiva), mientras que las otras ramas

mostraron coeficientes por debajo de uno. De manera similar, para el gen 0\_8531\_01, la rama 2 (Balsamea-Sachalinensis-Alba-Pinsapo) presentó valores de  $\omega$  mayores que uno mientras que el resto tuvieron valores menores que uno. Esto, si bien no constituyen resultados significativos, sí reflejan un patrón consistente con lo detectado por el RRR test, en el cual, estos genes estarían sufriendo un cambio en las tasas de sustituciones sinónimas y no sinónimas al interior de ciertos linajes, probablemente como consecuencia de la acción de selección positiva, aunque también podría tratarse de diferencias en la intensidad de la selección purificadora en estos linajes.

Aunque no significativa, la prueba MK reveló un patrón que refleja una preponderancia de variantes no sinónimas segregantes con respecto al número de variantes no sinónimas fijas. Esto posiblemente refleje por un lado los diferentes eventos demográficos sufridos por las especies de *Abies*, los cuales podrían provocar que variantes deletéreas sean retenidas dentro de las especies, así como variaciones en la intensidad de los regímenes selectivos (Fay & Wu, 2001; Hasegawa *et al.*, 1998). Por otro lado, este patrón puede estar asociado con la acción recurrente de la selección negativa en estos genes, la cual limita el número de sustituciones no sinónimas que ocurren en los genes estudiados y dificulta la detección de eventos de selección positiva cuando no hay suficientes sitios seleccionados de esta manera (Fay *et al.*, 2002).

En el caso de la prueba HKA, es posible que esta no haya arrojado resultados significativos dada la poca certeza al asignar una región neutral para realizar las comparaciones. Tradicionalmente, las regiones asignadas como neutrales en estas pruebas han sido promotores génicos o regiones intrónicas (Hudson *et al.*, 1987). Sin embargo, diferentes estudios han demostrado que las regiones promotoras de los genes también pueden estar sujetas a una fuerte selección purificadora y que los intrones pueden llegar a tener grandes diferencias en sus patrones y niveles de polimorfismo así como historias evolutivas contrastantes (Nachman & Crowell, 2000; Wray *et al.*, 2003). En este estudio, se utilizó un gen con valores de  $D$  de Tajima no

significativos y cercanos al promedio de todos los genes. Esto no garantiza su neutralidad, pero creo que es un buen aproximado para detectar desviaciones de esta. El porcentaje de genes bajo selección positiva a nivel filogenético detectados en este estudio está entre el 5% si se considera el único caso significativo y 17% si se consideran los casos pendientes de confirmación. Esto está dentro del rango encontrado en otros estudios de coníferas (Palmé *et al.*, 2009; Savolainen & Pyhäjärvi, 2007). Sin embargo, al haber evaluado genes a nivel filogenético, se esperaría haber encontrado una mayor proporción de eventos selectivos con respecto a los estudios realizados en especies individuales. Por un lado, esto está relacionado con el nivel de conservación de los genes analizados, los cuales no necesariamente reflejan los patrones de divergencia de las especies de *Abies* y menos aún, los eventos de divergencia adaptativa. Por otro lado, es posible que la baja cobertura a nivel filogenético también haya dificultado la detección de selección por las pruebas utilizadas, lo cual se volvería particularmente problemático en las ramas más externas de la filogenia, debido a la baja divergencia entre especies (Palmé *et al.*, 2009). En este aspecto es importante considerar los de por sí bajos niveles de divergencia entre las especies de *Abies*, las cuales no solo se reflejan en este estudio, si no en las dificultades para determinar las relaciones filogenéticas en algunos casos, como al interior de la sección Grandis & Oiamel, o entre ellas, como el caso de la posición filogenética de la sección Balsamea; así como en las bajas tasas de diversificación encontradas para el género (Q-P. Xiang *et al.*, 2015; Aguirre-Planter *et al.*, 2012). Por lo tanto, en el futuro serán necesarios estudios en los que se evalúen datos multigénicos, seleccionados aleatoriamente o datos a nivel genómico con una mayor representación filogenética que permitan evaluar mejor la divergencia entre secciones.

#### *Genes candidatos bajo selección positiva*

Se encontró evidencia de selección no direccional dentro de todo el grupo de *Abies* en la proteína DCL1, miembro de la familia de proteínas DCL (Dicer-like proteins). Se trata de una ribonucleasa multi-dominio, cuya función es la biogénesis y procesamiento de siRNA's y miRNA's a partir de RNA's de doble cadena. Estos

miRNA's regulan la expresión génica en diferentes procesos fisiológicos y de desarrollo en las plantas, tales como la morfogénesis y desarrollo de la hoja, el tiempo de floración, la embriogénesis y la determinación de la identidad de los órganos florales (He & Hannon, 2004). Las proteínas DCL presentan diferentes patrones de expresión durante eventos del desarrollo del tallo y las hojas, así como en diferentes etapas del desarrollo floral y de la semilla. De igual manera, forman parte de la respuesta de las plantas al estrés hídrico, salino y por congelamiento, así como a la acción de infecciones fúngicas (He & Hannon, 2004). Es interesante observar un exceso de variantes segregantes de reemplazo en el gen que codifica a DCL1 al interior del grupo de *Abies*, ya que puede indicar un proceso de diversificación de este gen dentro de este linaje que tal vez estaría asociado a alguno de estos tipos de estrés. Esto es consistente con estudios comparativos de las secuencias de estas proteínas entre diferentes especies de plantas (*Arabidopsis thaliana*, *Vitis vinifera*, *Oryza sativa* y *Populus trichocarpa*), los cuales han demostrado que existen importantes niveles de divergencia funcional entre ellas. Concretamente, se ha encontrado que DCL1 es la más divergente de estas proteínas con respecto a las otras tres (He & Hannon, 2004). En el caso de *Abies*, es posible que las características ambientales contrastantes en las que se desarrollan las especies de las diferentes secciones del género, tales como condiciones más boreales para la sección Balsamea, más áridas para los grupos mediterráneos y más templadas y húmedas para la sección Grandis & Oiamel (grupo Mesoamericano; Figura 2), funcionen como presiones selectivas, provocando el patrón diversificador observado en este gen.

Los otros dos casos de selección positiva, a pesar de la baja confiabilidad estadística discutida anteriormente, incluyen una proteína cinasa dependiente de calcio miembro de la familia de proteínas CDPK's (CDPK6; Mosca *et al.*, 2012). Esta estaría bajo un régimen de selección direccional dentro del grupo *A.balsamea-sachalinensis*. De manera similar a las DCL, las CDPK's están involucradas en la regulación de las respuestas de las plantas a diferentes estímulos lumínicos, hormonales, y por congelamiento, desecación y déficit de sucrosa (Li *et al.*, 2008). Los resultados aquí encontrados pueden indicar un proceso incipiente de divergencia adaptativa de la

sección Balsamea, como respuesta a sus requerimientos ecológicos específicos. Esta se encuentra distribuida en las latitudes más boreales para el género, lo que indicaría presiones por congelamiento, desecación y cambios drásticos del fotoperiodo (figura 2); todas estas funciones en las que las proteínas CDPK están involucradas.

Con respecto a estos dos casos (DCL1 y CDPK6) es importante señalar que en otros estudios ya se han encontrado señales de selección positiva en genes reguladores de respuestas a estrés y de eventos del desarrollo. En coníferas y a nivel local, se han detectado huellas de selección positiva en genes relacionados con la señalización mediante giberelinas (E3), la respuesta a estrés oxidativo (zf-Box) y en reguladores del desarrollo del meristemo y la emriogénesis (HD-III zip), mientras que a nivel filogenético se encontraron huellas de selección positiva en los genes que codifican proteínas de la familia *ASR*, involucrados en la regulación de la respuesta mediante ácido abscísico a estrés hídrico, salino y por congelamiento (Mosca *et al.*, 2012; Namrud *et al.*, 2008, 2010; Grivet *et al.*, 2013). Estos descubrimientos sugieren que las rutas reguladoras de diferentes eventos del desarrollo y fenológicos, así como de respuestas a estrés, podrían haber sido blancos importantes de la selección natural a lo largo de la evolución de las coníferas.

El segundo caso por confirmar corresponde a una proteína desconocida de unión a iones de cobre y con acción de oxidorreductasa, la cual estaría sujeta a selección no direccional en el grupo definido por la rama 2 de la filogenia. En plantas, los iones de cobre funcionan como cofactores para que diferentes proteínas realicen correctamente sus funciones, entre las que se encuentran la fotosíntesis, la respiración, la respuesta a estrés oxidativo y la señalización mediante etileno. Sin embargo, los iones de cobre prácticamente no existen de forma libre en las células vegetales debido a su alta toxicidad, lo cual implica que existen sustancias queladoras, entre ellas proteínas de unión a iones de cobre y chaperonas de cobre, cuya función es contener y transportar a estos iones para reducir su toxicidad (Kung *et al.*, 2006). Debido a lo anterior, es posible que estas funciones de las proteínas de unión a cobre estén bajo efectos selectivos entre especies. Sin embargo, esto queda como una hipótesis a ser probada en estudios futuros.

En el futuro, será necesario realizar estudios para confirmar el rol adaptativo de los genes aquí encontrados, particularmente en los dos casos donde la significancia no pudo determinarse con certeza. Para esto, sería necesario evaluar la historia evolutiva de estas familias de proteínas al interior del género *Abies*, para lo cual se requeriría aumentar el muestreo a nivel filogenético así como obtener secuencias de las demás proteínas de cada familia. Estudios de evolución molecular de este tipo ya han revelado los altos niveles de divergencia funcional, así como eventos de duplicación y subfuncionalización presentados por las proteínas DCL en otros grupos de plantas; por lo cual sería interesante evaluarlas dentro de las coníferas (Liu *et al.*, 2009). Adicionalmente, también serían necesarios estudios interespecíficos de asociación genética, con los cuales se podría determinar si efectivamente, los polimorfismos en estos genes se encuentran asociados con características fenotípicas relacionadas con los diferentes tipos de estrés que pueden sufrir las especies de *Abies*. Mediante estos estudios ya se han logrado detectar SNP's en genes candidatos en pinos, asociados con características relacionadas con la resistencia al estrés por desecación y congelamiento (Eckert *et al.*, 2009, 2010).

## Conclusiones

Se obtuvieron 23 secuencias para cinco especies de *Abies* y para dos grupos externos, *Cedrus atlantica* y *Pseudolarix amabilis*. Se trata de un número reducido de genes comparado con el número de genes inicialmente amplificados. Posiblemente se trate de genes conservados entre especies, constreñidos evolutivamente mediante selección negativa. Esto resalta la necesidad de generar más recursos genéticos para especies no modelo, de importancia ecológica y económica, como son las especies del género *Abies*. Esto ayudará a mejorar los trabajos ecológicos y evolutivos, así como las estrategias de conservación.

A partir de las anotaciones de los genes de *A. alba*, se logró identificar los homólogos de 17 de estos genes que codifican proteínas con función conocida. Entre las funciones de dichos genes destacaron aspectos regulatorios y de señalización mediante unión a nucleótidos y proteínas, así como funciones de cinasas y oxidorreductasas.

Se encontró que selección negativa es predominante a lo largo de la filogenia de *Abies* para estos genes. Esta estaría actuando en al menos 47% de los genes estudiados. Esto sugiere que las funciones de estos genes, como aspectos metabólicos y de regulación, son relevantes para la supervivencia de estos organismos y que, por lo tanto, se encuentran constreñidas evolutivamente. Además refuerza la idea de que se trabajó con un subconjunto de genes, sesgado hacia genes conservados entre especies, lo cual puede ser un factor que dificulte la detección de selección positiva, principalmente con pruebas como la MK y las basadas en coeficientes  $dn/ds$ .

Se detectó selección positiva no direccional en el gen DCL1, y posiblemente en el gen CDPK6. También hubo señales de selección direccional en un gen que codifica una proteína de unión a iones de cobre. Estos resultados sugieren que, a nivel filogenético, la selección natural estaría actuando sobre puntos específicos de las rutas reguladoras

de las respuestas a diferentes presiones ambientales, como el estrés hídrico, por congelamiento y lumínico, así como sobre aspectos reguladores del desarrollo y la fenología. Al detectar a estos genes como candidatos a ser sujetos de selección natural, se vuelve evidente la necesidad de realizar estudios moleculares sobre la evolución de estas familias génicas (DCL's, CDPK's) en el género *Abies* y otras coníferas para arrojar luz sobre el papel de estas proteínas en la evolución adaptativa de estas especies.

En el futuro será necesario buscar señales de selección en muestras de datos a nivel genómico, así como mediante estudios de expresión génica y transcriptómica, y estudios de asociación genómica, los cuales permitirán en conjunto evaluar el papel de la selección natural en el género *Abies* y otras coníferas.

## Anexos

Tabla anexa 1. Disponibilidad de datos por gen y por especie, cobertura, scoring y E-value críticos utilizados en cada caso para seleccionar los genes

Locus	Pinsapo	Concolor	Alba	Balsamea	Sachalinensis	Sibirica	Cedrus	Pseudolarix	Score	E-value (Cedrus)	E-value (Pseudolarix)
0_3969_01	SI	SI	< 20%	Si	< 20%	46%	SI	100%			
0_4032_02	SI	NO	100%	87%	100%	70%	87.85%	90.39%			
0_4105_01	SI	SI	< 20%	100%	<20%	51%	47%	58.16%			
0_6448_02*	SI	SI	100%	73.80%	64.90%	100%	48%	50.90%	80-200	2.00E-44	4.00E-43
0_6566_02*	SI	SI	100%	45%	45%	<25%	33%	34%	80-200	1.00E-30	4.00E-33
0_7009_01	SI	NO	100%	100%	<25%	63%	92.20%	94%			
0_7171_01	SI	NO	100%	100%	56.28%	<50%	71%	93.71%			
0_7921_01*	SI	SI	100%	62%	53%	40%	33%	34%	80-200	1.00E-27	2.00E-38
0_8479_01	SI	SI	<20%	63%	44%	< 20%	66.10%	57.29%			
0_8531_01*	SI	SI	100%	93%	81%	42%	72%	69.90%	80-200	2.00E-36	1.00E-38
0_9383_01	SI	SI	100%	64%	10%	6%	78.30%	100%			
0_9462_01	SI	SI	12%	100%	NO	38%	86.90%	60.65%			
0_12021_01*	SI	SI	100%	61%	61%	25%	48.30%	50.56%	80-200	7.00E-40	4.00E-40
0_12156_02	SI	SI	100%	100%	< 20%	< 30%	74.80%	97.50%			
0_12190_01*	SI	SI	100%	100%	77%	< 20%	46%	78%	80-200	2.00E-41	7.00E-40
0_10054_01	SI	NO	<15%	<10%	<10%	<10%	63.80%	??			
0_12329_02*	SI	SI	100%	94.30%	93.18%	<40%	76.50%	SI	80-200	6.00E-40	4.00E-40
0_3723_01	SI	NO	<15%	45%							
0_10162_01	SI	NO	<20%	67%	66%	<25%	72.46%	71.83%			
0_14221_01	SI	SI	100%	45.50%	<25%	<25%	62.18%	65.50%			
0_16009_02*	SI	SI	100%	100%	83.17%	<20%	77.40%	94.71%	80-200	1.00E-38	3.00E-38
0_16068_01	SI	SI	100%	45.45%	<25%	<20%	44.40%	46.15%			
0_16459_01	SI	NO	<20%	43.14%	60%	<20%	96.62%	90.33%			
0_18296_01	SI	SI	100%	75.77%	<25%	<25%	71%	62.43%			
0_18644_02	SI	SI	<10%	<10%	NO	NO	62.24%				
1_6493_01*	SI	SI	100%	89%	48.63%	51.67	94.20%	80.85%	80-200	7.00E-31	5.00E-35
2_2937_01*	SI	SI	100%	99%	99%	<20%	81.68%	63.86%	80-200	5.00E-38	5.00E-40
2_3852_01	SI	SI	100%	<20%	<25%	<25%	60.85	68.97%			

	%										
Locus	Pinsap o	Concol or	Alba	Balsam ea	Sachalinen sis	Sibiric a	Cedru s	Pseudola rix	Scor e	E- value (Cedru s)	E-value (Pseudola rix)
2_4107_01	SI	NO	<20%	49%	0%	<15%	64%	29%			
2_4724_01	SI	SI	<20%	53%	39.50%	50.36 %	49.63 %	43.17%			
2_5668_01	SI	NO	100%	61.13%	55%	<20%	30.49 %	52.19%			
2_6052_01	SI	SI	100%	44.38%	<20%	<20%	87.72 %	95.30%			
2_6313_01*	SI	SI	100%	67.13%	54.22%	<20%	53.07 %	50.35%	80- 200	3.00E- 36	2.00E-35
2_6351_01	SI	NO	<20%	98.40%	48.86%	<20%	38%	42.30%			
2_7121_02*	SI	SI	100%	87.79%	67.46%	<20%	55.26 %	58.61%	80- 200	3.00E- 40	6.00E-45
2_7725_01	SI	SI	<20%	66%	66%	<20%	51.27 %	34.28%			
2_7803_01*	SI	SI	100%	100%	98.68%	93.66 %	34.82 %	38.72%	80- 200	6.00E- 42	7.00E-39
2_8011_02	SI	SI	100%	85.40%	56.64%	<20%	56.86 %	44.44%			
2_9930_01	SI	SI	100%	<20%	<20%	<20%	76.19 %	60.90%			
2_9940_01	SI	SI	<20%	100%	<20%	<20%	97.87 %	88.27%			
2_10212_01	SI	NO	100%	100%	<20%	<20%	70.32 %	51.40%			
2_10216_01*	SI	SI	100%	100%	99%	<20%	40.81 %	67.83%	80- 200	4.00E- 38	5.00E-40
CL415CONTIG1_04	SI	SI	<20%	48.51%	48.30%	<20%	26.24 %	28.17%			
CL516Contig1_07*	SI	SI	100%	73.71%	68.29%	50%	49.22 %	26.03%	80- 200	5.00E- 43	6.00E-45
CL708CONTIG1_02*	SI	SI	100%	40.65%	40.84%	<20%	22.89 %	23.62%	80- 200	3.00E- 36	4.00E-38
CL813Contig1_03*	SI	SI	83.59 %	60.67%	36.71%	<20%	29.94 %	26%	80- 200	1.00E- 34	2.00E-39
CL1004Contig1_08*	SI	SI	100%	22.18%	22.18%	13.22 %	12.91 %	13.22%	80- 200	3.00E- 32	1.00E-28
CL1430CONTIG1_06	SI	SI	NO	19.89%	NO	NO	25%	19.21%			
CL1489Contig1_01	SI	NO	NO	100%	84.44%	NO	72%	44.88%			
CL1634Contig1_01	SI	SI	NO	100%	73.65%	76.88 %	62.36 %	60.21%			
CL1848Contig1_01	SI	SI	NO	100%	70.37%	NO	85.80 %	96.29%			
CL2332Contig1_01*	SI	SI	100%	46.77%	39%	31.55 %	28.68 %	26%	80- 200	2.00E- 36	1.00E-36
CL3539Contig1_01	SI	NO	NO	42%	40.57%	NO	28.40 %	27.44%			
CL4257CONTIG1_01	SI	NO	NO	20.30%	62.99%	NO	41.48 %	20.48%			
CL4354Contig1_01*	SI	SI	100%	37.34%	37%	NO	34.33 %	37%	80- 200	2.00E- 41	1.00E-36
CL4470Contig1_01	SI	NO	NO	32.29%	30.12%	NO	22.45 %	35.65%			
UMN_495_01*	SI	SI	100%	65.59%	44.38%	52.40 %	24.77 %	59.53%	80- 200	3.00E- 41	9.00E-40

<b>UMN_1337_02*</b>	SI	SI	100%	100%	100%	NO	94.52%	100%	80-200	1.00E-44	7.00E-40
<b>Locus</b>	Pinsap o	Concol or	Alba	Balsam ea	Sachalinen sis	Sibiric a	Cedru s	Pseudola rix	Scor e	E- value (Cedru s)	E-value (Pseudola rix)
<b>UMN_1483_01</b>	SI	NO	NO	78.19%	54.58%	NO	23.24%	36.75%			
<b>UMN_1911_01</b>	SI	SI	100%	NO	NO	NO	NO	NO			
<b>UMN_2789_01</b>	SI	SI	NO	88.76%	NO	NO	24.40%	58.96%			
<b>UMN_4361_01</b>	SI	SI	NO	99%	97%	NO	90%	66.82%			
<b>UMN_5272_01*</b>	SI	SI	100%	100%	50.74%	NO	33.83%	28.85%	80-200	5.00E-38	6.00E-45
<b>UMN_5833_01</b>	SI	SI	NO	100%	52.12%	55.89%	62.26%	64.15%			
<b>UMN_6426_02</b>	SI	NO	100%	54.69%	72.74%	NO	22.09%	52.11%			

Tabla anexa 2. Resultados del RRR test. Número de sustituciones sinónimas fijas y variables (SI y SV) y no sinónimas fijas y variables (RI y RV) para cada gen en cada una de las 4 ramas internas de la filogenia. \*p<0.05 para el la prueba exacta de Fisher. °p<0.05 para la prueba G

	0				1				2				3				4			
	RI	R	S	S	R	R	S	S	R	R	S	S	R	R	S	S	R	R	S	S
		V	I	V	I	V	I	V	I	V	I	V	I	V	I	V	I	V	I	V
0_6448_02	0	8	0	1	0	1	1	0	0	10	1	1	0	15*	3	1	0	2	0	7
0_6566_02																				
0_7921_01																				
0_8531_01	0	3	0	0	0	0	2	0	0	4°	2	0	4	5	6	2	0	15	0	10
0_12021_01	1	1	0	0	0	0	0	0	1	1	1	0	2	2	2	0	0	2	0	1
0_12190_01	1	26	0	44	1	0	0	0	2	26	0	44	3	28	3	46	0	4	0	3
0_12329_02	0	33	0	34	1	0	1	0	1	33	1	34	1	37	1	39	0	3	0	4
0_16009_02	0	1	0	2	0	0	0	1	0	1	0	3	6	1	3	5	0	10	0	8
1_6493_01	0	13	0	9	0	0	0	2	0	15	0	11	0	17	2	17	0	9	0	8
2_2937_01	0	11	0	3	0	3	0	4	0	14	0	7	6	15	2	9	0	9	0	8
2_6313_01																				
2_7121_02	1	18	0	7	0	1	0	3	1	19	0	10	3	21	0	12	0	4	0	7
2_7803_01	6	1	0	0	1	11	1	4	9	14	1	4	1	15	1	4	0	2	0	3
2_10216_01																				
CL516Contig1_07	0	6	0	32	0	0	1	0	0	6	1	32	0	7	1	34	0	1	0	2
CL708Contig1_02	2	0	0	1	1	0	0	0	3	0	0	1	5	2	1	1	0	0	0	3
CL813Contig1_03	2	0	5	1	0	3	0	5	2	3	5	6	3	4	5	8	0	2	0	1
CL1004Contig1_08	2	0	0	0	3	0	0	0	5	0	0	0	6	2	2	0	0	1	0	4
CL2332Contig1_01	15°	0	4	2	2	2	0	0	1	2	5	3	2	7	7	3	0	0	0	2
CL4354Contig1_01	1	0	1	0	1	1	0	0	4	1	1	0	6	5	2	0	0	6	0	0
UMN_495_01	0	2	1	1	0	0	0	0	0	2	2	1	2	2	5	6	0	2	0	3
UMN_1337_02	0	14	0	16	0	0	0	0	0	14	0	16	0	14	2	16	0	1	0	3
UMN_5272_01	0	7	0	0	2	0	3	1	2	7	3	1	3	10	3	4	0	2	0	4

Tabla anexa 3. Anotación de los genes retenidos para los estudios de acuerdo con Mosca *et al.*, 2012

Gen	Proteína	Función biológica
<b>0_644</b> <b>8_02</b>	Helicasa dependiente de ATP (DCL1)	GO:0003725 Unión a RNA de doble cadena
<b>0_656</b> <b>6_02</b>	Proteína desconocida	
<b>0_792</b> <b>1_01</b>	Unión/catálisis/oxidoreductasa	GO:0005488 Unión GO:0016491 actividad oxidoreductasa GO:0003824 actividad catalítica
<b>0_853</b> <b>1_01</b>	Unión a iones de cobre/oxidoreductasa	GO:0016491 actividad oxidoreductasa GO:0005507 unión a iones de cobre
<b>0_120</b> <b>21_01</b>	SRF8 (STRUBBELIG-RECEPTOR FAMILY 8)	GO:0016301 actividad cinasa GO:0004672 proteína cinasa GO:0005524 Unión a ATP GO:0004872 actividad receptora GO:0000166 unión a nucleótidos GO:0005515 Unión a proteínas
<b>0_121</b> <b>90_01</b>	Proteína desconocida	
<b>0_123</b> <b>29_02</b>	Proteína de unión a poli A	GO:0003676 Unión a ácidos nucleicos GO:0003723 Unión a RNA GO:0000166 Unión a nucleótidos
<b>0_160</b> <b>09_02</b>	Proteína argonauta	GO:0003676 Unión a ácidos nucleicos GO:0005515 unión a proteínas
<b>1_649</b> <b>3_01</b>	Proteína desconocida	
<b>2_293</b> <b>7_01</b>	Antiportador de ADP:ATP	GO:0005524 Unión a ATP GO:0005471 Actividad antiportadora de ADP:ATP
<b>2_631</b> <b>3_01</b>	Proteína desconocida	
<b>2_712</b> <b>1_02</b>	Proteína desconocida	
<b>2_780</b> <b>3_01</b>	Proteína desconocida	
<b>2_102</b> <b>16_01</b>	Unión a ubiquitina (ELC)	División celular. Ramificación de tricomas
<b>CL516</b> <b>Contig</b> <b>1_07</b>	ATPasa/ Pirofosfatasa transportadora de hidrógeno (AVP1)	GO:0009678 Pirofosfatasa transportadora de hidrógeno GO:0004427 actividad difosfatasa inorgánica
<b>CL708</b> <b>Contig</b> <b>1_02</b>	1-deoxy-D-xilulosa-5-fosfato reductoisomerasa	GO:0016491 actividad oxidoreductasa GO:0030604 actividad 1-deoxy-D-xilulose-5-fosfato reductoisomerasa GO:0003824 actividad catalítica GO:0016853 actividad isomerasa GO:0046872 unión a iones metálicos
<b>CL813</b> <b>Contig</b> <b>1_03</b>	sucrosa sintasa	GO:0016157 Síntesis de sucrosa
<b>CL100</b> <b>4Conti</b> <b>g1_08</b>	proteína del complejo t, subunidad 1 epsilon	GO:0051082 Unión a proteínas no plegadas GO:0005524 unión a ATP GO:0000166 unión a nucleótidos GO:0005515 unión a proteínas
<b>CL233</b> <b>2Conti</b> <b>g1_01</b>	Cinasa dependiente de calcio 6 (CDPK6)	GO:0016301 actividad cinasa GO:0004672 proteína con actividad cinasa GO:0005524 unión a ATP GO:0005509 unión a iones de calcio GO:0016740 actividad tranferasa GO:0000166 unión a nucleótidos GO:0004674 cinasa de serina/treonina
<b>CL435</b> <b>4Conti</b> <b>g1_01</b>	Fosfatasa de serina/treonina (PP2A-3)	GO:0005506 unión a iones de hierro GO:0004721 fosfoproteína con actividad fosfatasa
<b>UMN_</b> <b>495_0</b> <b>1</b>	Transferencia de grupos glucósilo	GO:0003843 1,3-beta-D-glucano sintasa
<b>UMN_</b> <b>1337_</b> <b>02</b>	alfa-trehalosa-fosfato sintasa	GO:0003824 actividad catalítica
<b>UMN_</b> <b>5272_</b> <b>01</b>	Familia 6-fosfogluconato deshidrogenasa	GO:0050661 Unión a NADP GO:0004616 actividad fosfogluconato deshidrogenasa (descarboxilación) GO:0070401 unión a NADP+ GO:0050662 unión a coenzimas GO:0016491 actividad oxidoreductasa GO:0003824 actividad catalítica GO:0005488 unión

## Bibliografía

- Behringer, D., Zimmermann, H., Ziegenhagen, B., & Liepelt, S. (2015). Differential gene expression reveals candidate genes for drought stress response in *Abies alba* (Pinaceae). *PLoS ONE*, *10*(4), 1–18. <https://doi.org/10.1371/journal.pone.0124564>
- Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. (2012). Impacts of climate change on the future of biodiversity. *Ecology Letters*, *15*(4), 365–377. <https://doi.org/10.1111/j.1461-0248.2011.01736.x>
- Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. *Trends in Genetics*, *22*(8), 437–446. <https://doi.org/10.1016/j.tig.2006.06.005>
- Buschiazzo, E., Ritland, C., Bohlmann, J., & Ritland, K. (2012). Slow but not low: genomic comparisons reveal slower evolutionary rate and higher  $dn/ds$  in conifers compared to angiosperms. *BMC Evolutionary Biology*, *12*(1), 8. <https://doi.org/10.1186/1471-2148-12-8>
- Chen, J., Gyllenstrand, N., Zaina, G., Morgante, M., Bousquet, J., Eckert, A., ... Lascoux, M. (2012). Disentangling the Roles of History and Local, *191*(July), 1–17. <https://doi.org/10.5061/dryad.82201.1>
- Creevey, C. J., & McInerney, J. O. (2003). CRANN: Detecting adaptive evolution in protein-coding DNA sequences. *Bioinformatics*, *19*(13), 1726. <https://doi.org/10.1093/bioinformatics/btg225>
- Creevey, & McInerney. (2002). An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences. *Gene*, *300*(1–2), 43–51. [https://doi.org/10.1016/S0378-1119\(02\)01039-9](https://doi.org/10.1016/S0378-1119(02)01039-9)
- Eckert, A. J., Bower, A. D., Jermstad, K. D., Wegrzyn, J. L., Knaus, B. J., Syring, J. V., & Neale, D. B. (2013). Multilocus analyses reveal little evidence for lineage-wide adaptive evolution within major clades of soft pines (*Pinus* subgenus *Strobus*). *Molecular Ecology*, *22*(22), 5635–5650. <https://doi.org/10.1111/mec.12514>
- Eckert, A. J., Van Heerwaarden, J., Wegrzyn, J. L., Nelson, C. D., Ross-Ibarra, J., González-

- Martínez, S. C., & Neale, D. B. (2010). Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics*, *185*(3), 969–982. <https://doi.org/10.1534/genetics.110.115543>
- Eckert, A. J., Wegrzyn, J. L., Pande, B., Jermstad, K. D., Lee, J. M., Liechty, J. D., ... Neale, D. B. (2009). Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold hardiness in coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*). *Genetics*, *183*(1), 289–298. <https://doi.org/10.1534/genetics.109.103895>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Eveno, E., Collada, C., Guevara, M. A., Léger, V., Soto, A., Díaz, L., ... Garnier-Géré, P. H. (2008). Contrasting patterns of selection at *Pinus pinaster* Ait. Drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution*, *25*(2), 417–437. <https://doi.org/10.1093/molbev/msm272>
- Farjon, A., & Rushforth, K. (1989). A classification of *Abies* Miller (Pinaceae). *Notes of the Royal Botanical Garden of Edinburgh*, *46*(1), 59–79. Retrieved from <http://www.cabdirect.org/abstracts/19900641693.html>
- Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, *155*(3), 1405–1413.
- Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, *133*(3), 693–709. <https://doi.org/evolution>
- Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*. <https://doi.org/10.1093/oxfordjournals.molbev.a040153>
- Grivet, D., Climent, J., Zabal-Aguirre, M., Neale, D. B., Vendramin, G. G., & González-Martínez, S. C. (2013). Adaptive evolution of Mediterranean pines. *Molecular Phylogenetics and Evolution*, *68*(3), 555–566. <https://doi.org/10.1016/j.ympev.2013.03.032>

Grivet, D., Sebastiani, F., Alía, R., Bataillon, T., Torre, S., Zabal-Aguirre, M., ... González-Martínez, S. C. (2011). Molecular footprints of local adaptation in two mediterranean conifers. *Molecular Biology and Evolution*, 28(1), 101–116. <https://doi.org/10.1093/molbev/msq190>

Hahn, M. W. (2008). Toward a selection theory of molecular evolution. *Evolution*, 62(2), 255–265. <https://doi.org/10.1111/j.1558-5646.2007.00308.x>

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W., & Medina-Elizade, M. (2006). Global temperature change. *Proceedings from the National Academy of Sciences of the United States of America*, 103(39), 14288–14293.

He, L., & Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews. Genetics*, 5(7), 522–531. <https://doi.org/10.1038/nrg1415>

Hedrick, P.W. (2005) Genetics of populations. Jones and Bartlett Publisers, 3º Edición.

Hudson, R. R., Kreitman, M., & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics*, 116(1), 153–159. <https://doi.org/10.1093/molbev/msv035>

Jaramillo-Correa, J. P., Aguirre-Planter, E., Khasa, D. P., Eguiarte, L. E., Piñero, D., Furnier, G. R., & Bousquet, J. (2008). Ancestry and divergence of subtropical montane forest isolates: Molecular biogeography of the genus *Abies* (Pinaceae) in southern México and Guatemala. *Molecular Ecology*, 17(10), 2476–2490. <https://doi.org/10.1111/j.1365-294X.2008.03762.x>

Jaramillo-Correa, J. P., Rodriguez-Quilon, I., Grivet, D., Lepoittevin, C., Sebastiani, F., Heuertz, M., ... Gonzalez-Martinez, S. C. (2015). Molecular proxies for climate maladaptation in a long-lived tree (*Pinus pinaster*, Aiston, Pinacea). *Genetics*, 199(3), 793–807. <https://doi.org/10.5061/dryad.fb436.the>

Jensen, J. D., Wong, A., & Aquadro, C. F. (2007). Approaches for identifying targets of positive selection. *Trends in Genetics*, 23(11), 568–577. <https://doi.org/10.1016/j.tig.2007.08.009>

Kimura, M. (1991). The neutral theory of molecular evolution: A review of recent

evidence. *The Japanese Journal of Genetics*, 66, 367–386.

KIMURA, M. (1968). Evolutionary Rate at the Molecular Level. *Nature*.  
<https://doi.org/10.1038/217624a0>

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>

Li, A. L., Zhu, Y. F., Tan, X. M., Wang, X., Wei, B., Guo, H. Z., ... Mao, L. (2008). Evolutionary and functional study of the CDPK gene family in wheat (*Triticum aestivum* L.). *Plant Molecular Biology*, 66(4), 429–443. <https://doi.org/10.1007/s11103-007-9281-5>

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>

Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., ... Marchetti, M. (2010). Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. *Forest Ecology and Management*, 259(4), 698–709. <https://doi.org/10.1016/j.foreco.2009.09.023>

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, 351(6328), 652–654. <https://doi.org/10.1038/351652a0>

Mosca, E., Eckert, A. J., Liechty, J. D., Wegrzyn, J. L., La Porta, N., Vendramin, G. G., & Neale, D. B. (2012). Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evolutionary Applications*, 5(7), 762–775. <https://doi.org/10.1111/j.1752-4571.2012.00256.x>

Nachman, M. W., & Crowell, S. L. (2000). Contrasting evolutionary histories of two introns of the Duchenne muscular dystrophy gene, Dmd, in humans. *Genetics*, 155(4), 1855–1864.

Namroud, M. C., Beaulieu, J., Juge, N., Laroche, J., & Bousquet, J. (2008). Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology*, 17(16), 3599–

3613. <https://doi.org/10.1111/j.1365-294X.2008.03840.x>

Namroud, M. C., Guillet-Claude, C., MacKay, J., Isabel, N., & Bousquet, J. (2010). Molecular evolution of regulatory genes in spruces from different species and continents: Heterogeneous patterns of linkage disequilibrium and selection but correlated recent demographic changes. *Journal of Molecular Evolution*, *70*(4), 371–386. <https://doi.org/10.1007/s00239-010-9335-1>

Neale, D. B., & Kremer, A. (2011). Forest tree genomics: growing resources and applications. *Nature Reviews Genetics*, *12*(2), 111–122. <https://doi.org/10.1038/nrg2931>

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press (Vol. 17). Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Molecular+Evolutionary+Genetics+Analysis#8>

Nei, M. (2005). Selectionism and neutralism in molecular evolution. *Molecular Biology and Evolution*, *22*(12), 2318–2342. <https://doi.org/10.1093/molbev/msi242>

Nei, M., Suzuki, Y., & Nozawa, M. (2010). The Neutral Theory of Molecular Evolution in the Genomic Era. *Annual Review of Genomics and Human Genetics*, *11*(1), 265–289. <https://doi.org/10.1146/annurev-genom-082908-150129>

Nielsen, R. (2001). Statistical tests of neutrality at the age of genomics. *Heredity*, *86*(January), 641–647.

Nielsen, R. (2005). Molecular Signatures of Natural Selection. *Annual Review of Genetics*, *39*(1), 197–218. <https://doi.org/10.1146/annurev.genet.39.073003.112420>

Nielsen R. & Slatkin M. (2013). An introduction to population genetics: theory and applications. Sinauer Associates, Inc. Publishers.

Nielsen, R., & Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, *148*(3), 929–936.

- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402. <https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- Palmé, A. E., Pyhäjärvi, T., Wachowiak, W., & Savolainen, O. (2009). Selection on nuclear genes in a Pinus phylogeny. *Molecular Biology and Evolution*, 26(4), 893–905. <https://doi.org/10.1093/molbev/msp010>
- Roschanski, A. M., Csilléry, K., Liepelt, S., Oddou-Muratorio, S., Ziegenhagen, B., Huard, F., ... Fady, B. (2016). Evidence of divergent selection for drought and cold tolerance at landscape and local scales in *Abies alba* Mill. in the French Mediterranean Alps. *Molecular Ecology*, 25(3), 776–794. <https://doi.org/10.1111/mec.13516>
- Roschanski, A. M., Fady, B., Ziegenhagen, B., & Liepelt, S. (2013). Annotation and Re-Sequencing of Genes from De Novo Transcriptome Assembly of *Abies alba* (Pinaceae). *Applications in Plant Sciences*, 1(1), 1200179. <https://doi.org/10.3732/apps.1200179>
- Savolainen, O., & Pyhäjärvi, T. (2007). Genomic diversity in forest trees. *Current Opinion in Plant Biology*, 10(2), 162–167. <https://doi.org/10.1016/j.pbi.2007.01.011>
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595. <https://doi.org/PMC1203831>
- Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting Natural Selection in Genomic Data. *Annual Review of Genetics*, 47(1), 97–120. <https://doi.org/10.1146/annurev-genet-111212-133526>
- Wang, X.-Q., Tank, D. C., & Sang, T. (2000). Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular Biology and Evolution*, 17(5), 773–781. <https://doi.org/10.1093/oxfordjournals.molbev.a026356>
- Wang, X. Q., & Ran, J. H. (2014). Evolution and biogeography of gymnosperms. *Molecular Phylogenetics and Evolution*, 75(1), 24–40. <https://doi.org/10.1016/j.ympev.2014.02.005>

- Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7(2), 256–276. [https://doi.org/10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9)
- Willyard, A., Syring, J., Gernandt, D. S., Liston, A., & Cronn, R. (2007). Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for Pinus. *Molecular Biology and Evolution*, 24(1), 90–101. <https://doi.org/10.1093/molbev/msl131>
- Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., & Romano, L. A. (2003). The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution*, 20(9), 1377–1419. <https://doi.org/10.1093/molbev/msg140>
- Xiang, Q.-P., Xiang, Q.-Y., Guo, Y.-Y., & Zhang, X.-C. (2009). Phylogeny of Abies(Pinaceae) inferred from nrITS sequence data. *Taxon*, 58(1), 141–152.
- Xiang, Q. P., Wei, R., Shao, Y. Z., Yang, Z. Y., Wang, X. Q., & Zhang, X. C. (2015). Phylogenetic relationships, possible ancient hybridization, and biogeographic history of Abies (Pinaceae) based on data from nuclear, plastid, and mitochondrial genomes. *Molecular Phylogenetics and Evolution*, 82(PA), 1–14. <https://doi.org/10.1016/j.ympev.2014.10.008>
- Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, 15(5), 568–573. <https://doi.org/10.1093/oxfordjournals.molbev.a025957>
- Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*, 15(5), 568–573. <https://doi.org/10.1093/oxfordjournals.molbev.a025957>
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Ziheng Yang, Rasmus Nielsen, Nick Goldman, A.-M. K. P. (2000). Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Molecular Biology and Evolution*, 19(1), 49–57. <https://doi.org/10.1093/oxfordjournals.molbev.a003981>