



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

“Caracterización de genes de la subfamilia de factores de transcripción R2R3-MYB en el abeto *Abies religiosa*”

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGO

P R E S E N T A:

JUAN JOSÉ RAMÍREZ LERMA

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

2012





Universidad Nacional
Autónoma de México



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.

1. Datos del alumno

Ramírez

Lerma

Juan José

55811036

Universidad Nacional Autónoma de México

Facultad de Ciencias

Biología

303593207

2. Datos del tutor

Dr.

Juan Pablo

Jaramillo

Correa

3. Datos del sinodal 1

Dr.

Luis Enrique

Eguiarte

Fruns

4. Datos del sinodal 2

Dr.

Arturo Carlos II

Becerra

Bracho

5. Datos del sinodal 3

Dra.

Alicia

Gamboa

de Buen

6. Datos del sinodal 4

Dra.

Erika

Aguirre

Planter

7. Datos del trabajo escrito

“Caracterización de genes de la subfamilia de factores de transcripción R2R3-MYB en el abeto *Abies religiosa*”

104 pp

2012



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

FACULTAD DE CIENCIAS
Secretaría General
División de Estudios Profesionales

Votos Aprobatorios

DR. ISIDRO ÁVILA MARTÍNEZ
Director General
Dirección General de Administración Escolar
Presente

Por este medio hacemos de su conocimiento que hemos revisado el trabajo escrito titulado:

"Caracterización de genes de la subfamilia de factores de transcripción R2H3-MYB en el abeto *Abies milleriana*"

realizado por Ramírez Lerma Juan José con número de cuenta 303593207 quien ha decidido continuar mediante la opción de tesis en la licenciatura en Biología. Dicho trabajo cuenta con nuestros votos aprobatorios.

Propietario	Dr. Luis Enrique Eguarte Frana	
Propietario	Dr. Arturo Carlos H Becerra Bracho	
Propietario Tutor	Dr. Juan Pablo Jaramillo Correa	
Suplente	Dea. Alicia Gamboa de Buen	
Suplente	Dea. Erika Aguirre Planter	

Atentamente,

"POR MI RAZA HABLARÁ EL ESPÍRITU "

Ciudad Universitaria, D. F., a 29 de marzo de 2012

EL JEFE DE LA DIVISIÓN DE ESTUDIOS PROFESIONALES

ACT. MAURICIO AGUILAR GONZÁLEZ

Señor sinodal: antes de firmar este documento, solicite al estudiante que le muestre la versión digital de su trabajo y verifique que la misma incluya todas las observaciones y correcciones que usted hizo sobre el mismo.

MAG/CZS/mdm

“...The child wonders at the Christmas Tree:
Let him continue in the spirit of wonder
At the Feast as an event not accepted as a pretext;
So that the glitteting rapture, the amazement
Of the first-remembered Christmas Tree...”

T.S. Eliot, 1954

**Esta investigación fue realizada en el Laboratorio de Evolución
Molecular y Experimental, Instituto de Ecología, Universidad
Nacional Autónoma de México, con el apoyo del proyecto
*Genética de poblaciones molecular y filogeografía de plantas
mexicanas* PAPIIT IN224309-3**

Agradecimientos

A mi familia y a la Universidad Nacional Autónoma de México por enseñarme tanto y hacer de mi lo que soy hoy.

A todos los maestros que me han formado como Biólogo, en particular al Dr. Juan Pablo Jaramillo Correa, a la M. en C. Julia Carabias Lillo, al Dr. Luis Eguiarte Fruns, a la Dra. Alejandra Vázquez-Lobo, Dra Erika Aguirre Planter, Dr. Daniel Piñero Dalmau, Biol. David Benavides, Biol. Lauro Ayala Garduño...y tantos otros.

A los sinodales que revisaron y ayudaron a terminar y dar forma a este documento: Dr. Luis Eguiarte Fruns, Dr. Arturo Becerra Bracho, Dr. Juan Pablo Jaramillo Correa, Dra Alicia Gamboa de Buen, Dra Erika Aguirre Planter.

A toda las personas (investigadores, técnicos, académicos y compañeros) del Instituto de Ecología de la UNAM: Dra. Laura Espinosa Asuar, Dra. Valeria Souza, Biol. Jonás Aguirre, Julia Múgica, Biol. Santiago Ramírez, Laura Corona Figueroa, Sandra Lorena Ament, M. en C. Eria Rebollar, M. en C. Morena Avitia, Biol. Jaime Gasca Pineda, M. en C. Enrique Scheinvar, Natalia Martínez, Carolina Durán, Rafael Silvio, Manuel Rosas, Eva Piedra.

A mi segunda familia por darme aliento y compartir tantos buenos momentos conmigo en especial a Rodrigo Elizalde, Daniel Martínez, Martín Zumaya, Guillermo Olivo, Raúl Ortiz, Daniel Badillo, José Luis Becherel, Eduardo Piña, Elizabeth Gómez, David Gómez, Pablo Gómez, Jorge Eduardo Sánchez, Noé Rodríguez, Jesús Domínguez.

A todas las personas (y entes) increíbles que he conocido en los últimos años, muchas de las cuales hoy puedo estar orgulloso de llamarlos amigos: Alejandra Martínez, Rodrigo León, Nuria Rubio, Lucia Magis, Thalia López, La Nutria Voladora, Andrea Navarrete, Fiorella Ortiz, Andrés Ojeda, Beatriz Cruz, La Tortuga Cecilio, Elisa Castro, Marina Rivero, Alicia Mastretta, Lucia Ruiz, Alicia Barceinas, Ardilla Maléfica, Renata Cao, Barbara Casillas, Leticia Turcio, Gabriela Mariscal, Fernanda Herce, Santiago Herce, Jesús Salcedo, Rogelio el Ninja, Enrique Limón, Diego Noriega, Sara Straffon, Roberto Ortiz, Adriana Nava, Ana Coll, David Chibras, Ingrid Rendón, Violeta Valadez, Eduardo Contreras, Patricia Macias, Carolina Magis, Mariana Flores, Serafín Pérez, Melisa Ramírez, Daniel Méndez, Fernando Arruti...Y todos los que me faltan (reales e imaginarios).

A Natura y Ecosistemas Mexicanos y todo su personal que me ha enseñado otro lado de la Biología y adoptarme dentro de un proyecto trascendental del cual puedo sentirme orgulloso, empezando por el Biol. Javier de la Maza, y toda la gente que labora en las estaciones de campo Chajul y Tzendales (lugares donde escribí gran parte de la presente tesis).

...Y por supuesto a la Vida, ¡si, esa que se escribe con V mayúscula!

Resumen

Los genes R2R3-MYB forman una subfamilia de factores de transcripción ampliamente diversificada en plantas. Estos han sido relacionadas funcionalmente con la regulación de numerosos procesos de la biología vegetal como el metabolismo primario y secundario, el desarrollo ontogénico, la diferenciación e identidad celular, y la respuesta a condiciones de estrés biótico y abiótico.

En el caso particular de las coníferas, los genes MYB han sido involucrados con procesos adaptativos al clima, principalmente con precipitación y temperaturas bajas. Un subgrupo (Sg4) de esta subfamilia génica ha sido asociado con respuestas a estrés que involucran la producción de componentes del xilema secundario y pared celular.

Sin embargo, todos los trabajos previos sobre genes MYB en coníferas se han centrado en especies boreales de la subfamilia Pinoidea. La presente tesis explora la evolución molecular de los genes R2R3-MYB del subgrupo 4 en *Abies religiosa*, una especie subtropical perteneciente a la subfamilia Abietoideae distribuida en el centro de México.

Para este trabajo se obtuvieron y analizaron 15 secuencias R2R3-MYB Sg4 de copia única. Se generaron reconstrucciones filogenéticas utilizando Máxima Parsimonia y métodos Bayesianos que sugieren una evolución relativamente independiente entre estos genes y sus homólogos en otras coníferas. Asimismo, se detectaron procesos moleculares que pueden haber intervenido en la evolución y diversificación de esta subfamilia génica tales como la recombinación no homóloga entre loci parálogos, y la modulación de elementos repetitivos en tandem.

Abstract

R2R3-MYB is a highly diversified subfamily of transcription factors that has been associated with the regulation of various functional processes in plants, including the regulation of primary and secondary metabolism, the ontogenic development, the cell identity and differentiation, and the responses to biotic and abiotic stress conditions.

In Conifers, MYB genes are involved in the adaptive response to climate conditions, particularly to precipitation and low temperature. The subgroup 4 of the R2R3-MYB subfamily has been previously associated to the cell wall components and secondary xylem formation, as a response to stress.

Given that previous studies in conifer MYB genes focused only in boreal species that belong to Pinoideae subfamily, this thesis explores the molecular evolution of the R2R3-MYB Sg4 genes in *Abies religiosa*, a subtropical species from the Abietoideae subfamily distributed in central Mexico.

Fifteen different single-copy R2R3-MYB Sg4 genes were obtained and analyzed. New phylogenetic reconstructions performed with Maximum Likelihood and Bayesian methods suggested a relatively independent evolution between the *Abies* genes and their counterparts in other conifers. Further molecular processes, such as non-homologous recombination among paralogs and modulation of tandem repetitive elements, were also inferred; thus implying that they play a key role in the evolution and diversification of this gene family in *Abies religiosa*.

Índice de contenidos

Resumen	i
Abstract	ii
1 Introducción	1
Prefacio	1
1.2 Características genéticas y de la historia natural de Coníferas en el contexto de la genómica y la evolución molecular.	2
1.1.1 Breve descripción del género <i>Abies</i>	3
1.1.2 Particularidades de los genomas de coníferas.	7
1.1.3 Mecanismos moldeadores de estructura y tamaño de genomas en coníferas.	7
a) Poliploidización	8
b) Duplicación génica	8
c) Modulación de secuencias de DNA repetitivo y elementos móviles	10
1.2 Familias génicas y factores de transcripción en plantas	11
1.2.1 Métodos filogenéticos y análisis evolutivo de secuencias	15
1.3 Factores de transcripción MYB	15
1.3.1 Factores de transcripción en el contexto genómico	15
1.3.2 Factores de transcripción MYB (origen, estructura, evolución y funciones)	18
1.3.3 Avances actuales en la caracterización de R2R3-MYB en coníferas.	24
2 Justificación y Objetivos	27
2.1 Justificación	27
2.2 Objetivos	28
2.2.1 Objetivo General	28
2.2.2 Objetivos Particulares	29

3 Materiales y Métodos	30
3.1 Obtención de tejido y extracción de DNA	30
3.2 Amplificación (PCR)	30
3.3 Clonación	32
3.4 Extracción del plásmido	32
3.5 Recopilación de secuencias	33
3.6 Alineación	35
3.7 Reconstrucción filogenética.	36
3.7.1 Reconstrucción utilizando ML	36
3.7.2 Reconstrucción con metodos Bayesianos	36
3.8 Pruebas de recombinación	37
3.9 Análisis de adquisición de intrones.	38
3.10 Búsqueda de secuencias repetitivas y/o móviles	38
4 Resultados	40
4.1 Obtención de secuencias	40
4.2 Relaciones filogenéticas de R2R3-MYB	41
4.3 Señales de recombinación	43
4.4 Adquisición de intrones	46
4.5 Secuencias repetitivas y/o móviles	48
5 Discusión	49
5.1 Relaciones filogenéticas de R2R3-MYB	49
5.1.1 Agrupamiento y posible evolución de secuencias de R2R3-MYB	49
5.1.2 Procesos que pueden afectar la reconstrucción filogenética	51
5.2 Recombinación no homóloga de genes R2R3-MYB en <i>A. religiosa</i>	53
5.3 Adquisición de intrones y secuencias repetitivas en los ArMYB.	54
6 Conclusiones y Perspectivas	57

6.1 Conclusiones	57
6.2 Perspectivas	58
Referencias	60
Apéndices	I
Apéndice 1: Información relevante sobre Secuencias utilizadas en los análisis	II
Apéndice 2: Secuencias nucleotídicas completas de ArMYB	IV
Apéndice 3: Posición y longitud de regiones génicas de secuencias ArMYBs	VIII
Apéndice 4: Alineación, estructura secundaria y grado de conservación de secuencias de aminoácidos utilizadas en los análisis.	IX
Apéndice 5: Análisis filogenéticos adicionales de genes R2R3-MYB	X

Índice de tablas y figuras

Figura 1	Figura 1a: Mapa de localización de poblaciones de <i>Abies religiosa</i> Figura 1b: Fotografía de bosque de oyameles	6
Figura 2	Estructura de genes MYB	20
Figura 3	Árbol filogenético de proteínas R2R3 (Bedon <i>et al.</i> , 2007)	25
Figura 3	Reconstrucción filogenética de genes de la subfamilia R2R3-Myb	42
Figura 4	Representación de eventos únicos de recombinación no homóloga en genes ArMYB	45
Figura 5	Reconstrucción filogenética de intrones de genes R2R3-MYB	47
Tabla 1	Principales dominios funcionales de factores de transcripción en plantas	17
Tabla 2	Funciones de genes R2R3-MYB en diferentes especies de plantas	21
Tabla 3	Fórmula de Reacción de PCR	31
Tabla 4	Posibles eventos de recombinación no homóloga detectados en genes ArMYB	44
Tabla 5	Secuencias repetitivas encontradas en genes ArMYB	48

1 Introducción

Prefacio

Esta tesis surge dentro del contexto de la genómica, en particular del creciente interés por comprender los ecosistemas forestales a partir de los mecanismos moleculares que subyacen a la biología del desarrollo, la fisiología y la adaptación. Es por ello que en la introducción se plantean los conceptos básicos y el contexto general en que se da este trabajo, para justificar así sus posibles aportes en el campo de la ecología molecular y la biología evolutiva.

El primer apartado es una revisión de aquellas características de las coníferas y en particular del género *Abies*, que nos lleva a entender la importancia de estudiarlo desde una perspectiva molecular. Entre dichas características se mencionan las particularidades de sus genomas (arreglo, tamaño, contenido relativo de diferentes tipos de secuencias, presencia de familias génicas, etc.), la distribución geográfica, así como las perspectivas de usar estas plantas como objeto de estudio en contextos genómicos, ecológicos y otros campos aplicados.

La segunda parte se adentra aún más en el nivel molecular, presentando primero la definición de las familias génicas y los mecanismos que las originan. Luego se exponen algunos ejemplos de estudios que ilustran los posibles alcances que se pueden lograr en el conocimiento genómico analizando en conjunto a grupos de genes, incluyendo a los factores de transcripción en plantas. Esto últimos tienen un papel importante en la regulación génica, y por lo tanto, en la biología funcional de los organismos.

Por último, se da paso a una reseña de la familia de factores de transcripción MYB y similares (MYB-related), en particular de la subfamilia R2R3. Se revisará parte de la información generada por diversos estudios: su estructura, sus

relaciones filogenéticas, la caracterización funcional de algunos de sus representantes, y los patrones de expresión en diferentes tejidos. También se tratará el papel de estas moléculas, en especial las del subgrupo 4 (Sg4) de R2R3-MYB, en procesos adaptativos en coníferas.

1.1 Características genéticas y de la historia natural de Coníferas en el contexto de la genómica y la evolución molecular.

En los últimos años, gracias a las innovaciones en las técnicas de biología molecular y al poder de las herramientas analíticas de la bioinformática, se ha podido transitar desde una aproximación gen-céntrica a un abordaje genómico del funcionamiento de los seres vivos. Esta visión es indispensable dada la complejidad de la estructura del genoma y de la regulación de la expresión de los genes. Con todo esto podemos empezar a responder preguntas acerca del funcionamiento del aparato molecular de los seres vivos, las correlaciones existentes entre esta maquinaria y las características de la biología, la ecología, y los patrones evolutivos entre organismos, ya sea nivel de individuo, poblaciones, especies y superiores (**Riechmann, 2000**). Si bien las bases genéticas de algunas características fenotípicas han sido descubiertas por aproximaciones de la genética mendeliana (p.e., las enfermedades metabólicas asociadas al mal funcionamiento de una enzima) (**Strauss, 1971**), queda por analizar cómo están determinadas otras características más complejas, como lo son el crecimiento, y los mecanismos de adaptación al medio (**Neale y Savolainen, 2004**).

En las especies arbóreas encontramos diversas características de desarrollo, fisiológicas y ecológicas que los diferencian de otras plantas y tienen una influencia directa sobre la evolución de su genoma (**Petit y Hampe, 2006**). Por ejemplo, estas: pueden alcanzar tamaños gigantescos, algunas especies llegan a rebasar 100 m como *Sequoia gigantea* (**Farjon, 2010**), lo cual las convierte en las mayores contribuyentes en la producción de biomasa terrestre (**Connor y Lanner,**

1990); tienen tiempos generacionales y tamaños efectivos poblacionales excepcionalmente grandes (**Muona y Harju, 1989**); además de que sus generaciones se superponen (**Hairton et al., 1996**).

En el caso particular de las coníferas encontramos algunos patrones, que las hacen ser un buen modelo experimental para el estudio de características complejas, por ejemplo: su variación genética en general no ha sido reducida por procesos antropogénicos, y aquellos genes que por su frecuencia pudieran estar asociados con adaptación no están afectados significativamente por procesos demográficos derivados de la domesticación. (**Palaisa et al., 2003**). Al ser xenógamas, dispersadas por viento, y tener alta eficiencia en la dispersión de polen y grandes genomas, el desequilibrio de ligamiento decae rápidamente después de unos pocos cientos de pares de bases, y hace que sea posible identificar individualmente genes implicados en el desarrollo de características fenotípicas complejas. (**Neale y Savolainen, 2004**). Además estas plantas se encuentran generalmente en grandes poblaciones con poca estructuración genética. Esto, en conjunto con sus patrones de diversidad, las hace ideales para ser analizadas por métodos de identificación genética no dependientes de experimentos de cruce, tales como los mapas de asociación. Por lo tanto, al comparar poblaciones ubicadas en diferentes sitios dentro de una clina ambiental, es probable que aquellas áreas del genoma con mayor diferenciación tengan alguna relevancia adaptativa. (**Beaumont y Nichols, 1996**). Lo cual resulta ser especialmente atractivo ya que aunque no se tienen identificados mutantes bien definidos, al poseer una gran diversidad nucleotídica, y tasas de recombinación relativamente bajas, se ha propuesto que son plantas muy adaptables, dados sus altos valores de dN/dS y su gran variación de fenotipos en pruebas de jardín común (**Buschiazzo et al., 2012**).

1.1.1 Breve descripción del género *Abies*

El orden Coniferales (división Pinophyta) se compone por 6 familias, 71 géneros y 664 especies (**Gernandt, et al., 2011**), y constituye el grupo de gimnospermas más diverso y más ampliamente distribuido del mundo (**Farjon, 1998**). Suelen ser las especies dominantes de los bosques templados y abarcar variadas condiciones climáticas, creciendo desde bosques tropicales y hasta la línea última de vegetación arbórea en ambos extremos latitudinales (**Farjon, 2010**) Son polinizadas por viento y altamente heterócigas (**Hamrick, 1987**). Además de tener niveles de variación genética altos (**Vendramin et al., 1999; Delgado et al., 1999**).

Dentro de las coníferas, la familia con mayor diversidad de especies y la segunda con mayor distribución geográfica es la familia Pinaceae, que se compone de 4 subfamilias, que alojan a 11 géneros y más de 200 especies. Las pinaceas se distribuyen en el hemisferio norte, aunque han sido introducidas en el hemisferio sur en forma de plantaciones comerciales (p.e. *Pinus radiata* en Australia, Chile y Sudáfrica) (**Farjon, 2010**).

Los géneros *Abies*, *Cedrus*, *Pseudolarix*, *Keteleeria*, *Nothotsuga* y *Tsuga*, componen la subfamilia *Abietoideae*. El primero de ellos se compone de entre 48 y 55 especies y se distribuye en Norte y Centroamérica, Europa, Asia, y algunas poblaciones en el Magreb. Las especies de zonas tropicales y mediterráneas se encuentran en zonas montañosas por arriba de los 2000 msnm y con alta precipitación (**Xiang et al., 2006**). Haciendo uso de marcadores nucleares se ha , confirmado la monofilia de todos los géneros de *Abietoideae*, siendo el género *Cedrus* el que divergió mas tempranamente, mientras que *Abies* es ubicado como grupo hermano de *Keteleeria* (**Gernandt et al., 2008**).

En México existen aproximadamente 8 especies de Abetos (*Abies*) que crecen

principalmente en las montañas de las dos sierras Madres, en la Faja Volcánica Transmexicana y en la Sierra Madre del Sur, entre los 2000 y 4000 msnm, en lugares fríos o templados con alta precipitación **(Aguirre-Planter, et al., 2012)**. *Abies religiosa* (Humboldt, Bonpland y Kunth) Schlechtendal y Chamizo, también conocido como oyamel, es una especie de montaña, relativamente común a lo largo del Faja Volcánica Transmexicana. Se distribuye en poblaciones generalmente continuas entre los 2000 y 3500 msnm **(Martínez, 1948)**. Llega alcanzar alturas de 45 y hasta 60m con anchuras entre 1.5 a 2m. Los juveniles poseen una corteza blanca grisácea y lisa, que se oscurece y forma placas en individuos maduros. Un rasgo muy característico es la disposición cruciforme de sus ramas, posible razón de su nombre específico, aunque también es común escuchar que se debe a que las ramas de este árbol se han usado en ritos y para adornar los pisos de templos desde tiempos prehispánicos **(Eckenwalder, 2009)**. Los bosques de esta especie son el hábitat de invierno preferido por las mariposas Monarca (*Danaus plexippus*) **(Anderson y Browler, 1996)**.

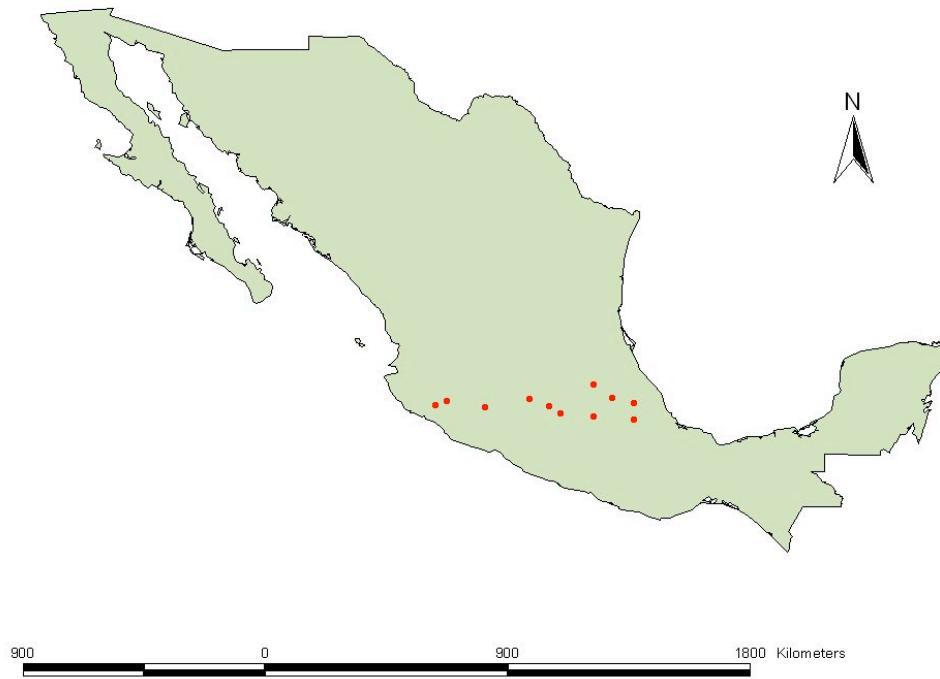


Figura 1a: Localización de 11 poblaciones de *Abies religiosa* en la República Mexicana, que señalan la eminente distribución de esta especie sobre la Faja Volcánica Transmexicana. (Mapa cortesía de Laura Figueroa Corona, Instituto de Ecología, UNAM)



Figura 1b: Fotografía de bosque de Oyameles (*Abies religiosa*) en las faldas del volcán Iztaccihuatl.

1.1.2 Particularidades de los genomas de coníferas.

El número cromosómico es muy estable en coníferas, ya que casi todos los géneros de la familia Pinaceae son diploides con $2n=24$ y sólo *Pseudotsuga* presenta $2n=26$. Además la colinearidad y sintenia de los mapas genéticos están conservadas, por lo cual es fácil encontrar regiones homologas al comparar la organización de los genomas de diferentes especies e incluso géneros. **(Shepherd y Williams, 2008).**

En plantas, el rango del tamaño del genoma va de entre 220Mb y 125,000Mb **(Bennett y Leitch, 2003)** y las coníferas se encuentran arriba de la mediana, con una extensión de entre 6500 Mb y 37000 Mb **(Leitch et al., 2005)**. Las variaciones interpoblacionales e interespecíficas del tamaño genómico parecen estar relacionadas con la posición dentro de gradientes geográficos y ambientales (latitud, humedad, precipitación anual) en los que están distribuidos los diferentes organismos. A su vez esto parece asociado con parámetros de crecimiento y de historia de vida como el tamaño de semilla, el tiempo generacional mínimo, los medios de dispersión, etc. **(Miksche y Hotta, 1993) (Grotkopp et al, 2004)**. Lo cual sugiere que la selección natural (y no solo la mutación) ha jugado un papel importante como elemento moldeador del tamaño del genoma de las coníferas **(Ahuja, 2005)**.

1.1.3 Mecanismos moldeadores de estructura y tamaño de genomas en coníferas.

Sin embargo lo anterior no explica cuales son los mecanismos que a nivel molecular y genético generan la variación para desplegar estas diferencias en el contenido de genético de estos organismos. A continuación se describen algunos mecanismos que afectan directamente el tamaño, estructura y composición del

genoma.

a) Duplicación completa del genoma (poliploidización)

La poliploidización puede ser vista como el aumento súbito y discreto del número de cromosomas de un organismo. Se sabe que en plantas y animales la duplicación genómica constituyó un medio rápido para la adquisición de nuevos genes y para la especiación, particularmente en la evolución temprana de estos linajes, ya que provee un vasto número de genes duplicados no sujetos a selección purificadora intensa en un lapso de una generación. **(Gibson y Spring, 2000)**. Es así que estos genes pueden funcionar como la materia prima que la mutación, la selección natural y los mecanismos de regulación y reparación del material genético procesará, produciendo con el paso del tiempo divergencia entre las secuencias de los cromosomas duplicados **(Adams y Wendel, 2005)**.

Las gimnospermas modernas no tienen una alta frecuencia de poliploidía real, presentándose solo en alrededor del 5% de las especies; en el caso particular de coníferas este número se reduce a menos de un 1.5%, pero también hay evidencia de que la poliploidía pudo haber jugado un papel importante en la evolución temprana de este taxón **(Ahuja y Neale, 2005)**.

b) Duplicación de génica

Los genes como entes individuales también pueden estar sujetos a la duplicación, de manera que los podemos encontrar por pares o en múltiples copias dentro del genoma, gracias a procesos como las duplicaciones en tandem, la recombinación cromosómica asimétrica, la conversión génica y la duplicación de segmentos cromosómicos.

El destino de los genes duplicados es variado, algunas secuencias divergen y desarrollan funciones distintas a las originales (neofuncionalización). En otros casos, la duplicación permite que exista la plasticidad suficiente para crear

variaciones del gen que desempeñen actividades similares, pero con sutiles diferencias, ya sea en los patrones de expresión o en su actividad catalítica en caso de enzimas. Es decir, las funciones del gen ancestral quedan repartidas entre las copias (subfuncionalización) **(Adams y Wendel, 2005)**.

Un tercer destino muy común es la pérdida de función de alguna de las copias génicas, lo que lleva a la acumulación mutaciones, hasta llegar al extremo de perder algún elemento importante para su expresión, como el promotor, o bien al desplazamiento del marco de lectura original, con lo cual se convertiría en un elemento no codificante conocido como pseudo-gen.

En cualquiera de los casos, las secuencias divergentes pueden terminar fijándose en la población, ya sea por deriva génica o por las posibles ventajas adaptativas de esta nueva copia **(Adams y Wendel, 2005)**. Por último es posible que aunque la secuencia de genes duplicados diverjan entre sí, se mantenga la misma función entre parálogos, a esto se le conoce como redundancia funcional **(Adams y Wendel, 2005)**.

Sea cual fuese el mecanismo productor de nuevas copias funcionales de un gen, la creación de nuevas entidades a partir de un gen ancestral resulta indispensable para entender la formación y presencia de familias génicas en los organismos, tema central del presente trabajo. Aunque se desconoce el estimado preciso de duplicaciones génicas en coníferas, se sabe de la existencia de un gran número de familias génicas en varias especies de los géneros *Pinus*, *Pseudotsuga*, *Picea*, y *Sequoia* **(Devey et al., 1994; Albouyeh y Ritland, 2011)**. Actualmente se dispone incluso de transcriptomas y catálogos de genes presentes en los genomas de *Picea*, *Pinus* y *Cryptomeria* **(Parchman et al., 2010; Futamura et al., 2006)** así como un conteo de las clases de proteínas codificadas por estos genes. Curiosamente, uno de estos esfuerzos reveló que existen diferencias en las abundancias relativas de tipos protéicos al comparar el genoma de *Picea* con los de algunas angiospermas **(Rigault et al., 2011)**.

c) Modulación de secuencias de DNA repetitivo y elementos móviles

El genoma de coníferas contiene una proporción relativamente grande de DNA repetitivo, llegando a ser cerca del 80% en *Picea glauca*, *Pinus strobus* (**Kriebel, 1993**) y *Pinus taeda* (**Kovach et al., 2010**); la mayoría de estas secuencias son entidades no codificantes como los microsatélites o elementos transponibles, pero existen también casos de genes repetitivos, como los rDNA (**Voytas et al., 1992**) Un ejemplo de elementos transponibles común y muy abundante en este grupo son los llamados LTR (*long terminal repeat*), de los que pueden llegar a hallarse hasta 10000 copias dispersos por todo el genoma (**Kossack y Kinlaw, 1999**).

Hay indicios de que la variación del tamaño de los genomas entre poblaciones pudiera estar asociado con la actividad diferencial de los elementos transponibles del genoma en función de diferentes condiciones ambientales, y puede haber una posible conexión entre la adaptación y la evolución del tamaño del genoma por la acción de estas entidades (**Kalendar et al, 2000**). Por otra parte, este tipo de entidades también son productores de variación genética, ya que crean muchos sitios polimórficos por inserción-delección (indel) de elementos transponibles de diferentes clases (**Morgante, 2006**).

Este tipo de variación se encuentra principalmente en regiones intergénicas no codificantes, aunque no es exclusiva de ellas, y es probable que algunos genes funcionales también contengan inserciones de elementos móviles. Por ejemplo, los transposones pueden conferir nuevas funciones al transportar exones foráneos a una proteína preexistente (*exon shuffling*) (**Juretic et al., 2005**) o pueden contribuir a la variación fenotípica, proveyendo diversos elementos *cis* de regulación en regiones flanqueantes a los genes (**Morgante, 2006**).

1.2 Familias génicas y factores de transcripción en plantas

Una familia génica es por definición un grupo de secuencias dentro del genoma que comparten características debido a su ancestría común (**Demuth et al., 2006**). Esta noción ya existía desde los inicios de la biología molecular, dada la evidencia de existencia de grupos de proteínas relacionadas estrechamente y diferenciables, dentro de una misma célula y entre organismos (**Pauling et al., 1949**).

Los genes contenidos en los organismos, pueden ser entendidos como productos de procesos evolutivos similares a los que dan origen a las especies; si bien los genes adquieren historias evolutivas independientes al divergir los linajes a partir de un ancestro común, en algunas ocasiones el origen de una nueva entidad molecular no implica la cladogénesis organísmica, y se debe a procesos que ocurren a nivel molecular como la poliploidización o la duplicación génica; pero al igual que las especies los genes pueden ser afectados por fuerzas evolutivas como la selección, la deriva génica, la mutación, siempre que sean transmitidos y conservados filogenéticamente (**Eisen, 1998**).

Con el auge de la genómica, el conocimiento disponible acerca de las familias génicas ha crecido exponencialmente en los últimos años (**Thornton y DeSalle, 2000**). La secuenciación de genomas y transcriptomas completos, así como el desarrollo de herramientas bioinformáticas que permiten realizar búsquedas más poderosas y minuciosas dentro de grandes bases de datos, han hecho posible caracterizar y mapear secuencias que guardan homología con secuencias ya conocidas. Esto a su vez ha permitido identificar regiones específicas dentro de los genes, que son motivos estructurales particularmente bien conservados entre moléculas y entre taxa, a tal grado que en algunos casos dicha región puede ser utilizada como diagnóstico de pertenencia a una familia de genes. (**Xuan et al, 2002**).

En plantas existe una grandísima variedad de familias génicas identificadas, para citar solamente un ejemplo, la base de datos para *Arabidopsis thaliana* (contenida en el sitio www.arabidopsis.org/browse/genefamily/index.jsp) hace referencia a casi un millar de familias génicas compuestas por un total de 8,331 genes. Mientras que hay familias génicas que contienen apenas un par de genes, hay otras que se componen por cientos de miembros.

Existen familias génicas de surgimiento y diversificación temprana que están presentes en una gran parte de los organismos y cuya estructura esta muy conservada debido a la naturaleza de los procesos que controlan. Como por ejemplo, las proteínas de la maquinaria central de replicación (*core DNA replication machinery*) (**Schultz et al., 2007**). Por el contrario, también se ha identificado la existencia familias génicas específicas de plantas, como lo son los genes *APETALA2*, *NAC*, *WRKY* y el factor de respuesta a auxinas (*ARF*). También existen casos de algunas familias génicas que aunque también se les ha encontrado en otros reinos, han diversificado significativamente en plantas, como por ejemplo los genes *bZIP*, los *MADS* y los *MYB* (**Rubin et al., 2000**).

La importancia del estudio de las familias génicas radica no solamente en la información con la que cada gen contribuye en relación a su función dentro de las células, sino también en la información que aportan sobre el funcionamiento del organismo, así como la relevancia que han tenido algunas moléculas en la evolución de algún linaje en particular. Esto último se deriva del hecho de que la diversificación de algunas familias génicas en plantas ha sido correlacionada con momentos importantes en la evolución de algunos grupos, como la aparición de un linaje. Por ejemplo, las peroxidas de clase III son una sinapomorfía de las plantas terrestres, por lo cual se piensa que fueron importantes para la colonización de a ambientes fuera del agua por embriofitas (**Duroux y Welinder, 2003**). Otras familias están relacionadas con la aparición de características importantes (novedades evolutivas) en la biología de algún taxón, como es el caso

de la familia de proteínas de pared celular DUF642, que se encuentran únicamente en espermatofitas. Esta evidencia filogenética, así como la interacción que tienen con polisacáridos de la pared celular, puede estar señalando que han sido un componente clave en la evolución de las plantas con semilla (**Vázquez-Lobo et al., 2012**). Cabe mencionar que es un hecho conocido que los patrones de diversificación de una familia génica pueden ser en parte el reflejo de la importancia que ha adquirido el proceso o procesos en los que intervienen estas moléculas en la fisiología y/o ecología del grupo, como por ejemplo los genes de auto incompatibilidad de algunas angiospermas (**Charlesworth et al., 2003**).

En el estudio de la biología funcional de plantas, la investigación de familias génicas también tiene mucho que aportar, ya que los patrones de expresión de los diferentes miembros pueden señalar la regulación específica de procesos metabólicos, fisiológicos y del desarrollo. Como por ejemplo, los genes del grupo *Apelata-2 like*, que participan en la formación de estructuras reproductivas de angiospermas (**Kim et al., 2006**). Además las familias génicas son grupos de genes que de manera natural forman redes de interacción, ya que al provenir de un mismo ancestro con una función definida, las copias divergentes pueden tener funciones redundantes, o subfuncionalizar entre sí. Los patrones encontrados en el destino de las nuevas copias son tema de interés para la biología molecular evolutiva, ya que reconstruyen procesos de organización funcional del genoma. Por ejemplo, algunos parálogos pueden actuar diferencialmente dependiendo del tejido o del momento del desarrollo (**Dubos et al., 2010**), otros pueden subfuncionalizar y complementar tareas (p.e. los genes *WOX3* en el desarrollo foliar) (**Nardmann et al., 2007**). Por último, las familias génicas son un excelente modelo para llevar a cabo la caracterización funcional a través del proceso conocido como genética invertida (*reverse genetics*), es decir, inferir y probar la función de genes, cuya secuencia y localización ya habían sido obtenidas previamente, haciendo uso de herramientas genómicas (**Riechmann y Ratcliffe, 2000**).

En cuanto a familias génicas de factores de transcripción, se sabe que el 45% de este tipo de moléculas que encontramos en *Arabidopsis thaliana* pertenecen a familias exclusivas de plantas y son en general las que conforman las familias génicas más grandes en estos organismos. Además, por su papel regulatorio, ofrecen interesantes sistemas de estudio genómico. La redundancia de funciones de los factores de transcripción representa un reto para la caracterización funcional, ya que es difícil descubrir y describir las funciones específicas de cada gen sin considerar la actividad de otros. Por ejemplo, al practicarse experimentos de mutación dirigida para “noquear” (eliminar) la expresión de un gen, generalmente no se encuentran mutantes fenotípicos bien definidos, ya que otros genes de la familia pueden estar supliendo las funciones perdidas, por lo cual solamente con la perspectiva genómica se pueden desentrañar estos procesos **(Riechmann y Ratcliffe, 2000)**.

Otros casos de familias de factores de transcripción que exigen una comprensión genómica son aquellos que en conjunto regulan procesos de desarrollo como genes homeóticos. Tal es el caso de los genes *AP1*, *AP3*, *PISTILLATA (PI)*, y *AG*, *SEP1*, *SEP2* y *SEP3* pertenecientes a la familia *MADS-box*, que por una expresión conjunta establecen el patrón de desarrollo de sépalos, pétalos, estambres y gineceo **(Pelaz et al., 2000)**.

Vale la pena mencionar que los sistemas de control transcripcional formados por las familias génicas de FT tienen una naturaleza compleja, ya que suelen “trazar” redes regulatorias diversas con un número limitado de factores, formando complejos multi-proteicos, donde diferentes FT funcionan como activadores o represores y cooperativamente regulan múltiples genes blanco. **(Singh, 1998)**.

1.2.1 Métodos filogenéticos y análisis evolutivo de secuencias.

Actualmente contamos con las herramientas filogenéticas y genómicas para empezar a abordar predicciones funcionales de familias génicas; esto mediante el discernimiento de verdaderas relaciones de homología entre genes. Empezando por los algoritmos de alineación de secuencias más sofisticados, ya que recordemos que las alineaciones de secuencias llevan implícita una primera hipótesis de homología (**Wheeler y Giribet, 2009**). Posteriormente, a partir de los sitios informativos de estas secuencias, se pueden establecer relaciones jerarquizadas entre secuencias haciendo uso de métodos de reconstrucción filogenética como Máxima Parsimonia (MP), distancia génica, Máxima Verosimilitud (ML) o análisis bayesianos. Con estas referencias, así como las relaciones de parentesco entre especies, es posible definir los genes en los árboles como homólogos o parálogos, y mapear patrones importantes que pueden estar relacionados con la función específica de genes o grupos de genes (**Eisen, 1998**) (ver sección 3.7 en Materiales y métodos).

1.3 Factores de transcripción MYB en plantas.

1.3.1 Factores de transcripción en el contexto genómico

En *Arabidopsis thaliana* se estima que más del 6% de los genes codificados son factores de transcripción (**Riechmann et al., 2000**). Sin embargo, estos se encuentran poco caracterizados funcionalmente, lo cual limita el entendimiento de las redes de regulación de la expresión génica. Se sabe que los factores de transcripción han tenido una radiación importante en plantas; así lo sugieren las diferentes familias que han diversificado de manera reciente, independiente y paralela en varios grupos (**Riechmann y Ratcliffe, 2000**). Por lo que puede tratarse de un proceso muy importante en relación con el funcionamiento y la evolución particular de estos organismos, que pudiera ser explicado por el hecho de que tienen una forma de vida sésil, una gran diversidad de sus procesos

bioquímicos, distribución en prácticamente todos los ambientes y una fisiología compleja, lo cual requiere un control fino de la expresión génica del desarrollo, el metabolismo y la respuesta a estímulos positivos o factores de estrés (**Wilkins et al., 2009; Stracke et al., 2007**).

Se han identificado estructuralmente numerosas familias génicas de factores de transcripción en plantas, para las cuales se han asociado dominios funcionales diagnósticos, Estas regiones presentan estructuras secundarias y terciarias muy particulares que determinan como se asocian con otras moléculas, principalmente con el DNA durante el proceso de transcripción. En la tabla 1 se describen la estructura de las principales familias de factores de transcripción (**Liu et al, 1999**).

Tabla 1: Principales dominios funcionales determinantes de las familias génicas de factores de transcripción en plantas (Liu *et al*, 1999).

Tipo de dominio	Estructura
Dedo de Zinc (Zinc Finger)	<ul style="list-style-type: none"> Motivo(s) de “dedo”, cada uno mantenido por residuos de cisteína o histidina organizados alrededor de un ión de Zinc.
bZIP	<ul style="list-style-type: none"> Región básica y un motivo de cierre (zipper) rico en leucina .
MYB related	<ul style="list-style-type: none"> Región básica que contiene de una a tres repeticiones imperfectas, cada una de las cuales forma un estructura hélice-vuelta-hélice.
Trihelix	<ul style="list-style-type: none"> Motivo de regiones básica, rica en prolina/glutamina que forma un dominio de trihélice de unión a DNA.
Homeodominio	<ul style="list-style-type: none"> Aproximadamente 60 residuos de aminoácidos que producen ya sea 3 o 4 α-hélices y un brazo Amino-terminal.
Myc b/HLH	<ul style="list-style-type: none"> Cúmulo de residuos de aminoácidos básicos adyacentes a un motivo de hélice-vuelta-hélice.
MADS	<ul style="list-style-type: none"> Aproximadamente 57 residuos aminoácidos que conforman una α-hélice y dos láminas β.
AT-hook motif	<ul style="list-style-type: none"> Región consenso nuclear R(G/P)RGRP que hace contacto con el surco menor de una región de DNA rica en A/T.
Caja HMG	<ul style="list-style-type: none"> Dominio con forma de L, que consiste en 3 α-hélices con un ángulo de 80° entre sus brazos.
AP2/EREBP	<ul style="list-style-type: none"> Región de 68 residuos aminoácidos con un dominio conservado que forma una putativa α-hélice anfipática.
B3	<ul style="list-style-type: none"> Secuencia de 120 aminoácidos conservados al final del Carboxilo terminal de VP1 y AB13
ARF	<ul style="list-style-type: none"> Región de 350 aminoácidos similar a la secuencia de B3

La caracterización estructural y funcional de los factores de transcripción en los diferentes grupos de plantas parece ser una tarea necesaria para poder transitar hacia un mejor entendimiento de la evolución y la ecología de grupos específicos de organismos, toda vez que varias de estas familias parecen estar involucradas en respuestas adaptativas al medio (**Prunier *et al.*, 2011; Namroud *et al.*, 2008**).

1.3.2 Factores de transcripción MYB (origen, estructura, evolución y funciones)

El primer gen MYB identificado fue el oncogen *v-Myb* derivado del virus de la mieloblastosis aviar. Se cree que este gen se originó a partir de otro, presente en vertebrados, que fue incorporado y mutó dentro del genoma viral, ya que existen al menos 3 genes de vertebrados claramente emparentados (*c-Myb*, *A-Myb* y *B-Myb*) (Klempnauer *et al.*, 1982). Posteriormente, se encontraron genes MYB en plantas, animales y hongos, y se calcula que esta familia génica tiene al menos mil millones de años de antigüedad (Lipsick, 1996).

El dominio de unión al DNA de las proteínas codificadas por estos genes es una secuencia altamente conservada entre los diversos linajes y determina la característica esencial de los genes de la familia génica (dominio MYB) . Este dominio actúa de manera específica reconociendo secuencias específicas y regulando la frecuencia de iniciación de la transcripción de un gen o grupo de genes ejerciendo influencia en el promotor. Además pueden presentar otro dominio que funciona como un elemento activador sensible a señales moleculares en momentos fisiológicos específicos en la célula (Lipsick, 1996).

La estructura del dominio MYB consiste en 1 a 4 repeticiones imperfectas (R) de 52 aminoácidos, cada una de las cuales forma 3 estructuras secundarias de α -hélice. La tercera hélice es la responsable de reconocer al DNA intercalándose en su surco mayor. La unión de otro motivo proteico similar, proveniente de otra repetición, es posible durante el contacto con el DNA, de manera que las dos hélices de reconocimiento actúan cooperativamente en la unión con la de secuencia específica de DNA (Du *et al.*, 2009; Ogata *et al.*, 1996).

Las proteínas se clasifican en diferentes clases en función del número de repeticiones adyacentes que las conformen y se les conoce como 1R, 2R, 3R o

4R. Proteínas pertenecientes a todas estas clases se encuentran en plantas, siendo las más abundantes las proteínas 2R y las más raras las 4R, que suelen tener un representante único en el genoma. **(Stracke et al., 2001)** En la figura 2 extraída de **Dubos et al., 2010** se ilustran las estructuras de las diferentes clases de factores de transcripción MYB.

El grupo de genes MYBs, y de factores de transcripción en general más abundante (que incumben a este trabajo) son la subfamilia de 2 repeticiones: R2R3 MYB. La estructura de estos factores es modular y contienen el dominio MYB de unión al DNA altamente conservado en el extremo amino terminal, así como un dominio de activación o represión en el extremo carboxilo terminal **(Dubos et al., 2010)**.

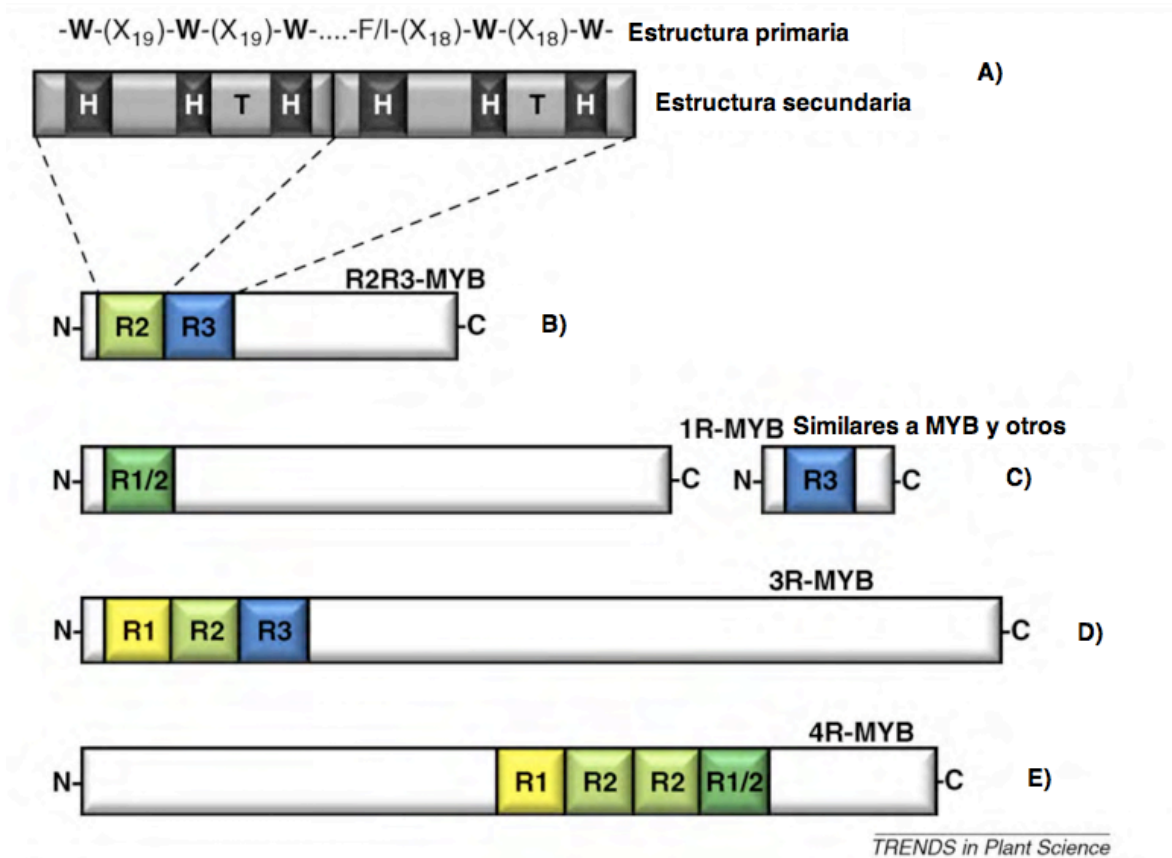


Figura 2: Estructura del dominio MYB, y de las diferentes clases de proteínas MYB.

A) Estructura primaria (repeticiones imperfectas de aminoácidos con residuos triptófanos (W) regularmente presentes que forman las 3 hélices del dominio MYB, B) Molécula del tipo R2R3-MYB (2R), C) estructura 1R-MYB, D) Estructura 3R-MYB, estructura 4R-MYB. Figura extraída de (Dubos *et al.*, 2010).

Las proteínas MYB-R2R3 han sido divididas para su clasificación en subgrupos con base en la conservación de su dominio MYB y en los motivos de aminoácidos de la región carboxilo terminal. La mayoría de estos subgrupos fueron definidos en *Arabidopsis sp.* (Stracke *et al.*, 2001). Estudios filogenéticos comparativos con otras especies han revelado tanto la existencia de genes ortólogos, como nuevos miembros que no tienen relación de ortología con otros genes en *Arabidopsis*. Esto sugiere que la evolución de los MYBs R2R3 puede estar influenciada por procesos particulares a cada linaje, contribuyendo al origen de

innovaciones evolutivas específicas a ciertos linajes de plantas (**Wilkins *et al.*, 2009**).

Por su gran número, así como por la diversidad de funciones biológicas que desempeñan, contrastan con las otras clases de MYB y no se les pueden enmarcar funcionalmente con un solo tipo de procesos. En la siguiente tabla se describen algunas de las funciones en las que se ha caracterizado la participación de genes R2R3-MYB.

Tabla 2: Funciones de genes R2R3-MYB en diferentes especies de plantas.

Nombre del gen	Tipo de proceso	Función específica	Especie en la que se describió	Referencia
Varios R2R3-MYb: <i>MYB5a</i> , <i>MYB5b</i> , <i>MYBPA1</i> y otros.	Regulación metabólica	Biosíntesis de flavonoides, antocianinas y taninos.	<i>Vitis vinifera</i>	Matus <i>et al.</i>, 2008
<i>AtMYB46</i> , <i>AtMYB52</i> , <i>AtMYB54</i> , <i>AtMYB69</i> , <i>AtMYB109</i> <i>AtMYB61</i> .	Regulación metabólica	Síntesis de componentes de pared celular: monolignoles, celulosa y mucílago.	<i>Arabidopsis thaliana</i>	Dubos <i>et al.</i>, 2010
<i>AtMYB2</i>	Regulación metabólica y respuesta a señales	Respuesta al ácido abscísico.	<i>Arabidopsis thaliana</i>	Abe <i>et al.</i>, 2003
<i>MYB62</i>	Regulación metabólica y respuesta a señales	Respuesta a giberelinas.	<i>Arabidopsis thaliana</i>	Devaiah <i>et al.</i>, 2009

<i>MYB23</i>	Desarrollo	Establecimiento del patrón de desarrollo en células epidérmicas.	<i>Arabidopsis thaliana</i>	Kirik, et al., 2005
MYB5	Desarrollo	Diferenciación de la cubierta seminal externa.	<i>Arabidopsis thaliana</i>	Li et al., 2009
<i>PhMYB1, AmMYBML2, AtMYB16</i>	Desarrollo	Diferenciación del aparato estomático y desarrollo temprano de la inflorescencia.	<i>Antirrhinum majus, Petunia hybrida y Arabidopsis thaliana</i>	Baumann et al., 2007
<i>MYB33; MYB65</i>	Desarrollo	Control de diferenciación en estructuras reproductivas.	<i>Arabidopsis thaliana</i>	Millat y Gluber, 2005
<i>MYB37</i>	Desarrollo	Formación del meristemo axilar.	<i>Arabidopsis thaliana</i>	Keller et al., 2006
<i>AS1, PHANTASTICA, ROUGH SHEAT2</i>	Desarrollo	Determinación de polaridad dorso-ventral en hojas.	<i>Arabidopsis thaliana, Antirrhinum majus, Zea mays</i>	Byrne et al., 2000
Varios R2R3-MYB	Desarrollo	Varios procesos ontogenéticos	Varias	Consultar: Stracke, 2001; Shen et al., 2011; Wilkins et al., 2009
<i>AtMYB30</i>	Respuesta a estrés	Activador de muerte celular programada en respuesta a ataque de	<i>Arabidopsis thaliana</i>	Raffaella et al., 2008

		patógenos		
<i>AtMYB60</i> , <i>AtMYB96</i> , <i>AtMYB15</i>	Respuesta a estrés	Regulación del movimiento estomático y tolerancia al frío	<i>Arabidopsis thaliana</i>	Agarwal et al., 2006
<i>AtMYB2</i>	Respuesta a estrés	Respuesta a deshidratación y salinidad elevada.	<i>Arabidopsis thaliana</i>	Yoo et al., 2005
<i>AtMYB62</i>	Respuesta a estrés	Resistencia frente a condiciones de deprivación de fósforo.	<i>Arabidopsis thaliana</i>	Hammond y White, 2011
<i>AtMYB102/AtM4</i> y <i>AtMYB41</i>	Respuesta a estrés	Defensa contra herbivoría, resistencia a deshidratación por daño mecánico y al estrés osmótico.	<i>Arabidopsis thaliana</i>	De Vos et al., 2006; Cominelli et al., 2008

Se ha visto que los genes parálogos pueden actuar en diferentes puntos de las vías que regulan y que tienen diferencias en el patrón de expresión tejido-específica además al analizar los patrones funcionales de MYB R2R3 en el contexto filogenético, resaltan algunos procesos (**Stracke, 2001**), por ejemplo que existe un alto grado de redundancia de funciones entre los miembros de subclados. Esto sugiere que la subfuncionalización ha sido un proceso recurrente en la evolución de la subfamilia y, en algunos casos, dos genes MYB pueden actuar como elementos reguladores antagónicos en la regulación de la expresión de uno o varios genes, uno como activador y el otro como represor, de manera que su expresión será una función de la concentración relativa y proporción de ambos MYB R2R3 (**Moyano et al., 1996**). Por otro lado, aunque los parálogos cercanos filogenéticamente suelen compartir funciones, el patrón de expresión de estos puede diferir de manera tejido-específica (**Stracke, 2001**).

1.3.3 Avances actuales en la caracterización de R2R3-MYB en coníferas.

En coníferas se ha demostrado que estos genes presentan asociaciones con la producción de lignina (**Goicoechea et al., 2005**) y en procesos fisiológicos como el crecimiento secundario y la formación de madera y la resistencia a estrés biótico y abiótico. Razón por la cual resultaron muy atractivos para aquellos interesados en “mejorar” la producción de madera con enfoque extractivo y en la ecología molecular de los bosques.

Algunos estudios hicieron uso del transcriptoma completo secuenciado para diferentes especies de *Piceas*, con el fin de detectar SNPs con variación genética atípica (outliers) en regiones codificantes por medio de barridos genómicos. Dicha variación genética fue contrastada con la variación en gradientes de factores ambientales (principalmente precipitación y temperatura) para diferentes poblaciones de coníferas en Canadá, encontrando tanto en *Picea mariana* (**Prunier et al., 2011**) como en *Picea glauca* (**Namroud et al., 2008**) genes candidatos a estar involucrados con la adaptación climática, varios de estos resultaron ser genes R2R3-MYB.

Luego de conocer la existencia de los genes PtMYB1 y PtMYB4, involucrados en *Pinus taeda* con la regulación de lignificación, **Bedon et al. (2007)** describieron la estructura y las funciones de los R2R3 en otras coníferas. Para ello aislaron y caracterizaron, a partir de cDNA, 18 secuencias de MYB R2R3 en coníferas (13 secuencias de *Picea glauca* y 5 en *Pinus taeda*), y construyeron un árbol filogenético utilizando estas secuencias (figura 3), y algunas otras reportadas.

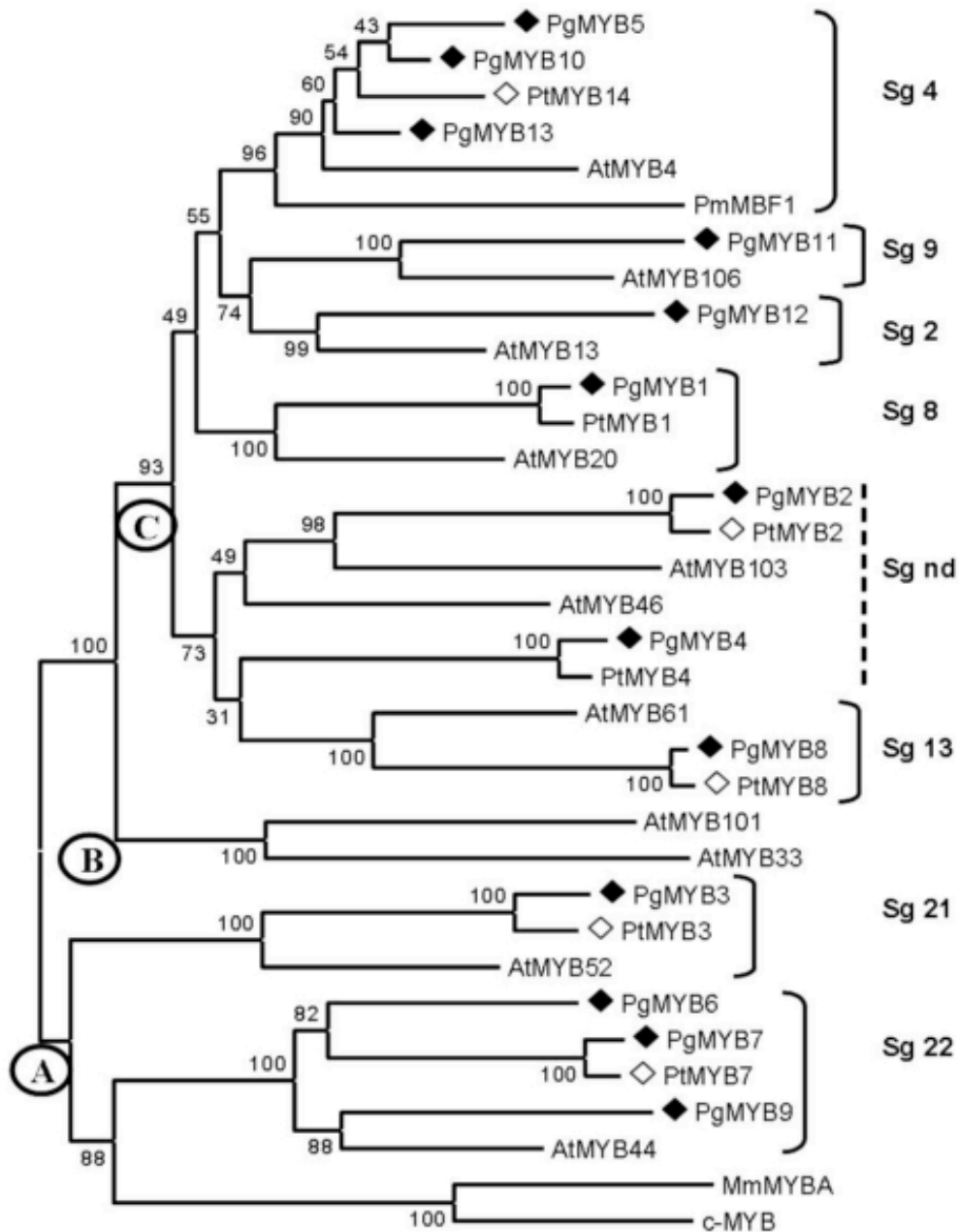


Figura 3: Árbol filogenético de proteínas R2R3-MYB de angiospermas y gimnospermas. Este árbol fue realizado con métodos de distancia (neighbour-joining; 1000 bootstraps) a partir del alineamiento de las secuencias codificantes completas de 13 proteínas MYB de piceas y 5 de pinos identificadas por Bedon *et al.*, 2007. Se escogieron algunas proteínas de *Arabidopsis* como marcadores para identificar los 3 principales grupos (círculos A, B, C) y subgrupos (Sg marcados a un lado del corchete; nd=no determinado) se incluyen también secuencias de Humano (c-MYB) así como de ratón (MmMYBA). Figura extraída de Bedon *et al.*, 2007.

Ellos también plantearon una relación potencial de funciones de los genes R2R3-MYB de *Picea* con la lignificación, luego de analizar la expresión de genes MYB como PgMYB2, 4 y 8 en los perfiles de transcripción de diferentes tejidos vasculares de individuos adultos, y de PgMYB 9,11 y 13 en tejidos con crecimiento primario y formación de madera de compresión (**Bedon et al., 2007**).

Entre los genes R2R3-MYB de estas coníferas fue identificado un subgrupo (sg4) que presenta una relación funcional en procesos de respuestas al estrés como la producción de componentes del parénquima (**Bedon et al., 2010**). Esto se correlaciona a su vez con la presencia de un motivo represor conocido como EAR en la región C-terminal variable de estos MYBs. Este EAR ya había sido implicado por otros en respuestas a estrés en *Arabidopsis* (**Ohta et al., 2001; Weigel et al., 2005**). El patrón filogenético observado por **Bedon, et al. (2010)** indica que el clado de secuencias estrechamente relacionadas con PtMYB14 (Sg4) se encuentra mucho más diversificado en las coníferas que en otras plantas, lo cual podría estar señalando una expansión adaptativa de estos genes dentro del genoma debida a sucesivas duplicaciones y mantenidas por selección natural.

2 Justificación y Objetivos

2.1 Justificación

La presente tesis aporta información acerca de la subfamilia de factores de transcripción R2R3-MYB en una importante especie forestal mexicana: *Abies religiosa*. Los genes R2R3-MYB son de gran relevancia tanto para la ciencia básica como para campos aplicados, ya que son buenos candidatos a estar involucrados en procesos adaptativos locales (**Prunier et al., 2011; Namroud et al., 2008**) y por lo tanto pueden impactar en el entendimiento de la ecología de estos árboles, y en la formulación de programas de conservación de los bosques templados de montaña, este es un ecosistema muy importante en la mitigación al cambio climático y en la provisión de servicios ambientales (incluyendo hidrológicos) pero especialmente sensible a este proceso global, (**FAO, 2008**). Además las coníferas son árboles dominantes de estos ecosistemas y tienen relevancia en términos económicos y de producción de materias primas, ya que son el grupo de plantas más importante a nivel mundial para ser procesados en productos maderables, fibras y papel (**Bills, 2003**).

Por otro lado, este tipo de estudios en *Abies religiosa* es pionero en un campo inexplorado, ya que hasta ahora sólo se conoce información acerca de genes R2R3-MYB en coníferas boreales y filogenéticamente apartadas del género *Abies*, que ha evolucionado independiente desde hace más de 300 millones de años de aquel otro taxón (Subfamilias Abietoidea y Pinoidea) (**Gernandt et al., 2008**).

Hasta ahora los estudios para caracterizar factores de transcripción se han realizado en organismos modelo (p.e. *Arabidopsis thaliana*), principalmente angiospermas, de manera que existe un sesgo importante en el estudio de la diversidad de las familias génicas con respecto a otras divisiones de plantas. Por

otro lado son pocas las familias génicas de regulación que han sido caracterizadas sistemáticamente en coníferas, y ninguno de los casos en especies mexicanas, que crecen bajo condiciones ambientales muy distintas de las que ocurren en otras latitudes, por lo que pueden propiciar procesos ecológicos y evolutivos muy diferentes. Dado lo anterior y la posible naturaleza adaptativa de los marcadores genéticos utilizados en este trabajo, se espera encontrar patrones divergentes de relevancia para los organismos mexicanos.

El producto concreto que genera esta tesis son genes de secuencia única que pueden estar asociados a procesos adaptativos, y a partir de los cuales se pudiera explorar con estudios posteriores la diversidad nucleotídica, alélica, entre poblaciones y especies. Además, este tipo de estudios podrían aportar nuevas metodologías para el descubrimiento de redes de regulación y de genes candidatos a ser blancos de la selección natural direccional. O bien para predecir qué interacciones y/o factores ambientales han sido determinantes en la historia evolutiva de las especies.

2.2 Objetivos

2.2.1 Objetivo General

-Caracterizar genes pertenecientes a la subfamilia génica R2R3-MYB en *Abies religiosa* (Pinaceae), e identificar algunos de los procesos a los que ha estado sujeta esta familia génica en la evolución de este linaje.

2.2.2 Objetivos Particulares

- a) Reconstruir las relaciones filogenéticas entre los genes R2R3-MYB presentes en el genoma en *Abies religiosa* (Pinaceae) y en otras coníferas: *Picea glauca* y *Pinus taeda* (Pinaceae).

- b) Identificar procesos que hayan incidido en la evolución de las secuencias de estos genes, como la adquisición de intrones, la recombinación no homóloga entre parálogos, la modulación de elementos repetitivos y/o la presencia de elementos móviles como transposones en las regiones no homólogas.

3 Materiales y Métodos

3.1 Obtención de tejido y extracción de DNA

Parte fundamental de este trabajo fue la amplificación de distintos genes pertenecientes a una misma familia génica, para lo cual resulta cómodo y accesible trabajar con un tejido haploide, que tiene representada una sola copia de cada locus. Para ello se extrajo el DNA de megagametofitos de las semillas de abetos. Se germinaron semillas de 10 individuos provenientes de 4 localidades en cajas de petri estériles con algodón húmedo. Sin embargo, solo germinaron individuos de una sola de las poblaciones, localizada en el Desierto de los Leones al poniente del Distrito Federal; 9 semillas de esta población fueron seleccionadas para diseccionar el gametofito y extraer de este el DNA con el kit DNeasy plant mini kit de QIAGEN (Hilden, Alemania) siguiendo las especificaciones del fabricante. Este se resuspendió en 50 μ l de agua y se rediluyó posteriormente a una concentración del 10%.

3.2 Amplificación

Cada reacción de amplificación se llevó a cabo en un volumen total de 25 μ l, con las concentraciones mostradas en la Tabla 3.

Tabla 3: Volúmenes utilizados para realizar la PCR de los genes del subgrupo 4 de R2R3MYBs en *Abies religiosa*

Reactivo	Volumen	Concentración
Agua	14 μ l	-
Amortiguador	2.5 μ l	5X
MgCl ₂	1 μ l	25mM
dNTP's	1 μ l	10 μ M
Primer (F + R)	1 μ l + 1 μ l	10 μ M
Taq Polimerasa	0.5 μ l	5u/ μ l

Se utilizaron los primers reportados por **Bedon et al. (2010)**. El primer forward estaba parcialmente degenerado para facilitar la amplificación de todos los miembros del subgrupo 4 de la familia génica de interés, cuya secuencia era 5'-TGGCGYTCSCTTCCCAAGGC-3'; el primer reverse tenía la secuencia 5'-CCAGTAGTTCTTTATCTCGTTGTC-3'. Las condiciones de amplificación consistieron en una desnaturalización inicial a 94°C por 5 minutos, seguida de 35 ciclos de una desnaturalización a 94°C por 1 minuto, una alineación a 54°C por 1 minuto, y una extensión a 72°C por 1 minuto con 15 segundos, concluyendo con una extensión final a 72°C por 10 minutos.

Para confirmar la amplificación de los fragmentos se corrió el producto de PCR en un gel de agarosa (2% en TAE) teñido con bromuro de etidio (7.0 μ l/100 ml buffer) aplicando 100 V, por un tiempo aproximado de 25 minutos. El resultado fue visualizado con luz ultravioleta y fotografiado (Kodak EDAS 240). En el gel se detectó una banda consistente de entre 450 y 500 pb en varias pruebas, la cual

era ligeramente de mayor tamaño pero factible en relación con la ventana de amplificación esperada según **Bedon et al., (2010)**.

Los productos de amplificación fueron enviados a secuenciar al servicio High-Throughput Genomics Unit (HTGU), del Departamento de Ciencias Genómicas de la Universidad de Washington, Seattle con resultados infructuosos potencialmente producidos por la presencia de más de un producto de amplificación (cada uno proveniente de un loci diferente) en la muestra enviada.

3.3 Clonación

Los productos de amplificación de 5 individuos diferentes fueron clonados en un vector pCR 2.1 del kit TA Cloning (Invitrogen™) según las especificaciones del fabricante. Con el vector ligado al producto, se transformaron células competentes de *E. coli* TOP10 quimio-competentes utilizando el protocolo del fabricante para una transformación mediada por choque térmico. Cada vial de bacterias competentes fue transformado por 1 solo producto de PCR, y a partir de cada uno de estos fueron cultivadas por duplicado de células en medio sólido LB con kanamicina (concentración final de 100µg/ml), a 37°C por 24 horas. Posteriormente se eligieron 30 colonias de color blanco, las que con mayor probabilidad fueron transformadas por el vector, de cada de las clonaciones duplicadas (una caja se uso para trabajar, y la otra como respaldo). Cada colonia fue puesta a crecer individualmente en 1ml de medio liquido LB con ampicilina y 7% de glicerol a 100µg/ml a 37°C, con agitación por 24 horas.

3.4 Extracción del plásmido

Se extrajeron los plásmidos para las tres clonaciones (30 colonias cada una) y una parcial (seis colonias). Para ello se utilizó el kit Montage Plasmid Miniprep_{HTS} (Millipore™), basado en la lisis de células y aplicación de vacío a través de placas

con filtros y columnas de afinidad diversa. Para obtener el vector se siguió el protocolo de lisado total del procedimiento Plasmid DNA Miniprep siguiendo las instrucciones del fabricante. Finalmente se diluyó y preservó en 50µl de la solución de elución proveída por el fabricante.

Para confirmar la extracción exitosa del plásmido, se corrieron todos los productos de la extracción en un gel de agarosa (al 1% en TAE) con bromuro de etidio aplicando 100V por aproximadamente 20 min. Aquellos pozos con una banda fluorescente de alto peso molecular (mayor que la franja más pesada del marcador, 2000pb) fueron considerados como extracciones exitosas (29 colonias).

Para aislar el producto clonado del vector que lo contenía, se digirieron las 29 extracciones exitosas con la enzima EcoR1 a una concentración final de 0.6u/µl utilizando los reactivos y el protocolo de digestión provisto por el kit Rapid DNA Digestion (Promega™, Madison, WI). Un total de 46 muestras (ya que cuando era posible se mandaban por duplicado) fueron mandadas al servicio HTGU del Departamento de Ciencias Genómicas de la Universidad de Washington, Seattle, obteniéndose resultados satisfactorios (ver más abajo).

3.5 Recopilación de secuencias

Los electroferogramas de las secuencias obtenidas se revisaron visualmente y se descartaron aquellos de baja calidad. Posteriormente fueron editadas utilizando los programas SeaView 4 (**Gouy et al., 2010**) y Se-AL v2.0a11 Carbon (**Rambaut, 1996-2002**) para tener únicamente la región de la secuencia correspondiente a la ventana de amplificación y corregir algunas falsas asignaciones de nucleótidos, así como artefactos de secuenciación como dobles picos.

Al final de este proceso fueron elegidas las secuencias, que cumplían los siguientes requisitos:

-Presentar cromatogramas de buena calidad, es decir con poco ruido de fondo.

-Una vez editados los errores y artefactos de la secuenciación, ser idéntica a la secuencia de la muestra duplicada, en caso de que esta existiese.

-Tener una longitud mayor a 400 pb (longitud aproximada de la ventana de amplificación esperada)

-Al ser descartadas (al azar) alguna de las dos secuencias de las muestras duplicadas, en cada secuencia restante debía existir al menos un polimorfismo real que las distinguiera de cualquier otra. Es decir, todas las secuencias debían de ser diferentes entre sí.

Después de escoger dichas secuencias se procedió a fijar una nomenclatura con la cual se trabajó el resto del análisis. A todos los nombres les fue antepuesto las letras ArMYB (para MYB de *Abies religiosa*), seguidas por un número romano del I al XV. El uso de números romanos fue implementado para distinguir estas secuencias de aquellas de otras especies incluidas en la reconstrucción de la hipótesis filogenética, y así evitar confusiones por relaciones de ortología inexistentes. Por ejemplo, PgMYB1 y PtMYB1 son ortólogos verdaderos, pero la secuencia ArMYB1 puede no tener ninguna relación estrecha con las dos anteriores.

Para reconstruir las relaciones filogenéticas se realizó una búsqueda a partir de genes y descarga de datos de genes relacionados reportados previamente (**Bedon et al., 2007**) en GenBank. Un total de 33 secuencias fueron seleccionados (17 de *Picea glauca*, 8 de *Pinus taeda*, 5 *Arabidopsis thaliana*, 3 de

Picea sitchensis). Las secuencias descargadas de GenBank fueron elegidas, con base en filogenias de la familia MYB publicadas anteriormente (**Bedon et al., 2007; Bedon et al., 2010; Dubos et al., 2010**), con el objetivo de tener muestreados todos los miembros del subgrupo 4 que han sido anotados como tales, así como algunos miembros de los subgrupos filogenéticamente cercanos, *i.e.* Sg2, Sg8 y Sg9 (**Dubos et al., 2010**).

Por otro lado, 6 secuencias adicionales no anotadas y provenientes de *Picea sitchensis* y *Picea glauca* fueron obtenidas al someter las secuencias de *Abies religiosa* al programa, para ello se ingreso a la opción blastn en el sitio <http://blast.ncbi.nlm.nih.gov/Blast.cgi>, y se ingresaron como “*query*” una por una las secuencias ArMYB dentro de la base de datos *Nucleotide Collection* (nr/nt) y dejando todos los parámetros por defecto.

Esta búsqueda se llevó a cabo con el objeto de descartar la posibilidad de que los patrones encontrados sean artefactos generados por un muestreo incompleto de secuencias (**Barracough y Nee, 2001**) es decir, para descartar la posible existencia de otros miembros de la familia MYB aún no identificados como tales y que presentasen ortología con las secuencias de *A. religiosa*. Esto también permitió confirmar que los genes ArMYB forman un grupo interno del clado de secuencias del subgrupo 4, tal y como se esperaría a partir del diseño experimental. En la tabla 1 del apéndice se muestra información adicional sobre las 48 muestras utilizadas en el análisis filogenético.

3.6 Alineación.

Como primer paso para formular una hipótesis filogenética se alinearon las 48 secuencias, a partir de la secuencia traducida de aminoácidos, se descartaron los intrones y se concatenaron los exones. Se alinearon las secuencias visualmente usando el marco de lectura de secuencias codificantes publicadas (**Bedon et al.,**

2007) utilizando el editor de alineaciones Se-AL v2.0a11 Carbon (**Rambaut, 1996-2002**). Luego, se cortaron los extremos que no eran cubiertos por la ventana de amplificación de las secuencias obtenidas para *Abies*. Así se llegó a una matriz de datos de 74 aminoácidos y 48 "taxa", esta matriz corresponde a aproximadamente 5/6 del dominio MYB-R2R3 compartido por todas las secuencias utilizadas.

3.7 Reconstrucción filogenética.

Para la matriz de nucleótidos se buscó el modelo evolutivo que mejor se ajustaba utilizando el criterio AIC en el programa jmodeltest (**Posada, 2008**) con los parámetros por defecto, a excepción del número de esquemas de sustitución probados, que se fijó en siete, ya que la opción por defecto utiliza un total de 11, siendo los últimos cuatro modelos evolutivos muy específicos para genomas virales o de vertebrado. Para los aminoácidos se utilizó la versión online del programa ProtTest http://darwin.uvigo.es/software/prottest_server.html (**Abascal, et al., 2005**). Todos los parámetros se dejaron por defecto, con la excepción de que se marcó la opción de optimización de la topología del árbol.

3.7.1 Reconstrucción utilizando ML

Para realizar las filogenias tanto de los aminoácidos como de los nucleótidos se utilizó el servicio online de RAxML BlackBox en el sitio; <http://phylobench.vital-it.ch/raxml-bb/> (**Stamatakis et al., 2008**), dejando todos los parámetros por defecto, con la excepción de que se marco la opción "Gamma model of rate heterogeneity", dados los modelos de evolución obtenidos con jmodeltest y ProtTest indicaban que la variación en tasas de sustitución tiene una distribución Gamma (diferente para cada posición). También se seleccionó la opción de realizar la búsqueda de máxima verosimilitud (casilla Maximum likelihood search). Obteniendo así la topología de mayor verosimilitud y los valores de bootstrap para

cada nodo a partir de 1000 réplicas (ver resultados)

3.7.2 Reconstrucción con métodos Bayesianos.

Se ingresaron las matrices de secuencias al programa Mrbayes v3.2.0 (**Ronquist y Huelsenbeck, 2003**) ajustando los parámetros del análisis con base en los modelos de evolución obtenidos con Modeltest y ProtTest (seis tasas de sustitución diferentes y distribución Gamma). Se corrieron dos muestreos cada uno con cuatro cadenas muestreando un árbol cada 500 generaciones por 1,000,000 de generaciones, a una temperatura de 0.2000 y empezando en un árbol aleatorio. Después de cerca de 400,000 generaciones, se alcanzó estabilidad en la diferencia promedio de la desviación estándar entre las corridas. El valor de esta diferencia oscilaba mínimamente en valores cercanos, pero por arriba de 0.05, por lo cual se le permitió continuar hasta el final (1,000,000 generaciones). Basado en la distribución de la probabilidad posterior obtenida al final del análisis, se descartaron (*burn-in*) los primeros 400 árboles muestreados, para luego obtener árboles consenso con el resto de los árboles y calcular las probabilidades posteriores calculadas para cada nodo. (ver Resultados).

En todas las reconstrucciones filogenéticas realizadas (bayesianas o ML, y aminoácidos o nucleótidos) se enraizaron los árboles resultantes utilizando como grupo externo la secuencia más divergente de la colección: AtMYB13 *Arabidopsis thaliana* catalogado dentro del subgrupo 2 de ArMYB y el cual aparece como grupo externo a los demás subgrupos en las reconstrucciones previas (**Bedon et al., 2010**) y se editaron utilizando el programa FigTree v1.3.1 (**Rambaut, 2006-2009**).

3.8 Pruebas de recombinación:

Para probar posibles eventos de recombinación, se analizaron las secuencias

obtenidas en este trabajo con el programa Recombination Detection Program v.3.44 (**Martin et al., 2010**) que implementa diferentes algoritmos basados en métodos filogenéticos y estadísticos para determinar si una secuencia proviene de otras secuencias parentales que recombinaron. Para aumentar la resolución de estos análisis fue necesario incluir los intrones de las secuencias. Por ello se suministraron las secuencias originales completas en la versión online del programa MAFFT, disponible en <http://mafft.cbrc.jp/alignment/server/> (**Katoh et al., 2009**), utilizando todos los parámetros definidos por defecto, excepto que se escogió la estrategia E-INS (recomendada para alineaciones de menos de 200 secuencias y con múltiples dominios conservados y gaps largos). Se inspeccionó visualmente la alineación resultante en Se-AL v2.0a11 Carbon (**Rambaut, 1996-2002**) y en RDP, y se seleccionó la opción de enmascaramiento de secuencias muy similares; ya que si alguna de ellas se encuentra sobremuestreada, puede dificultar la detección de eventos únicos de recombinación.

3.9 Análisis de adquisición de intrones.

Con el objetivo de comprobar si los intrones de los diferentes genes encontrados en *A. religiosa* han evolucionado junto con el resto de la secuencia, o si por el contrario son adquisiciones recientes y móviles entre secuencias, se generó un árbol de máxima verosimilitud utilizando únicamente las secuencias de los 2 intrones que se encuentran en el dominio R2R3-MYB. Esta reconstrucción se realizó con el mismo procedimiento y ajuste de parámetros utilizados previamente, y fue comparada con la topología de los análisis hechos para las secuencias codificantes.

3.10 Búsqueda de secuencias repetitivas y/o móviles:

Por último, se indago acerca de la posibilidad de que en la evolución de los genes R2R3-MYB de *Abies religiosa* se han acumulado secuencias repetitivas o

elementos móviles transponibles; ya que este tipo de secuencias pueden estar relacionadas con la divergencia posterior a la duplicación de genes, así como a sitios promotores de *exon shuffling* y recombinación. Para ello se analizaron las secuencias en la versión en línea de diferentes programas reportados por **(Lerat, 2009)**, incluyendo RepeatMasker **(Smit et al., 1996-2004)**, LTR_STRUC **(McCarthy y McDonald, 2003)**, LTR_FINDER **(Xu y Wang, 2007)**, así como en los programas Tandem Repeat Finder **(Benson, 1999)** y Perfect Microsatellite Repeat Finder **(Leach, 2000)**. En todos los casos se utilizaron los parámetros por defecto.

4 Resultados

4.1 Obtención de secuencias

De un total de 28 secuencias de *Abies religiosa* de buena calidad se conservaron únicamente 15, después de descartar las duplicadas, que provenían de un mismo individuo. Estas secuencias provienen de 2 individuos únicamente, 12 secuencias (ArMYBI a ARMYBXII) de la clonación 1 y 3 adicionales (ArMYBXIII a ARMYBXV) de la clonación 2. La secuencia más corta tuvo una longitud de 463pb y la más larga de 1015pb (secuencia atípica, duplicada en tandem, atribuible a un error del secuenciador), siendo la longitud promedio 547 pb.

En cuanto a su estructura, todas las secuencias obtenidas contenían 3 exones y 2 intrones intercalados, a excepción de la secuencia ArMYBIII que contenía un segundo intrón excepcionalmente largo, por lo cual la ventana de amplificación no alcanzó a incluir el tercer exón. En cuanto a las secuencias de otras especies obtenidas de GenBank, estas se trataron en su mayoría de RNA mensajeros, por lo cual no contenían intrones.

Después editar las secuencias obtenidas para descartar los intrones y concatenar los exones, se llegó a una matriz de 74aa x 48 "taxa" o su contraparte de 222pb x 48 "taxa", cubriendo aproximadamente 5/6 partes del dominio MYB-R2R3.

Al ser analizados los modelos de evolución molecular que mejor ajustan para estas matrices se encontró que dada la variación observada estos fueron GTR+G los nucleótidos y JTT+G para los aminoácidos.

4.2 Relaciones filogenéticas de R2R3-MYB

En función del tipo de secuencia utilizada (nucleótidos o aminoácidos), así como del método de reconstrucción (máxima verosimilitud o bayesianos) con la región codificante de los genes MYB (tres exones concatenados), se obtuvieron distintas filogenias de genes. Sin embargo, hubo numerosas coincidencias entre las topologías de los diferentes análisis, las cuales se muestran en una sola representación gráfica (figura 4). Los otros tres árboles complementarios se incluyen en los apéndices de la tesis.

Se utilizó la topología obtenida por el análisis de máxima verosimilitud (ML) con secuencias de nucleótidos (222pb), ya que presentaba la mayor cantidad de nodos resueltos y concordaba con varias de las relaciones propuestas anteriormente para los MYB (**Bedon *et al.*, 2007; Bedon *et al.*, 2010**). En esta topología fueron marcados los nodos coincidentes con al menos otro de los análisis llevados a cabo (ML con aminoácidos, Bayesianos con nucleótidos y/o Bayesianos con aminoácidos). Los genes fueron etiquetados en función de su origen y/o clasificación dentro de la subfamilia R2R3-MYB.

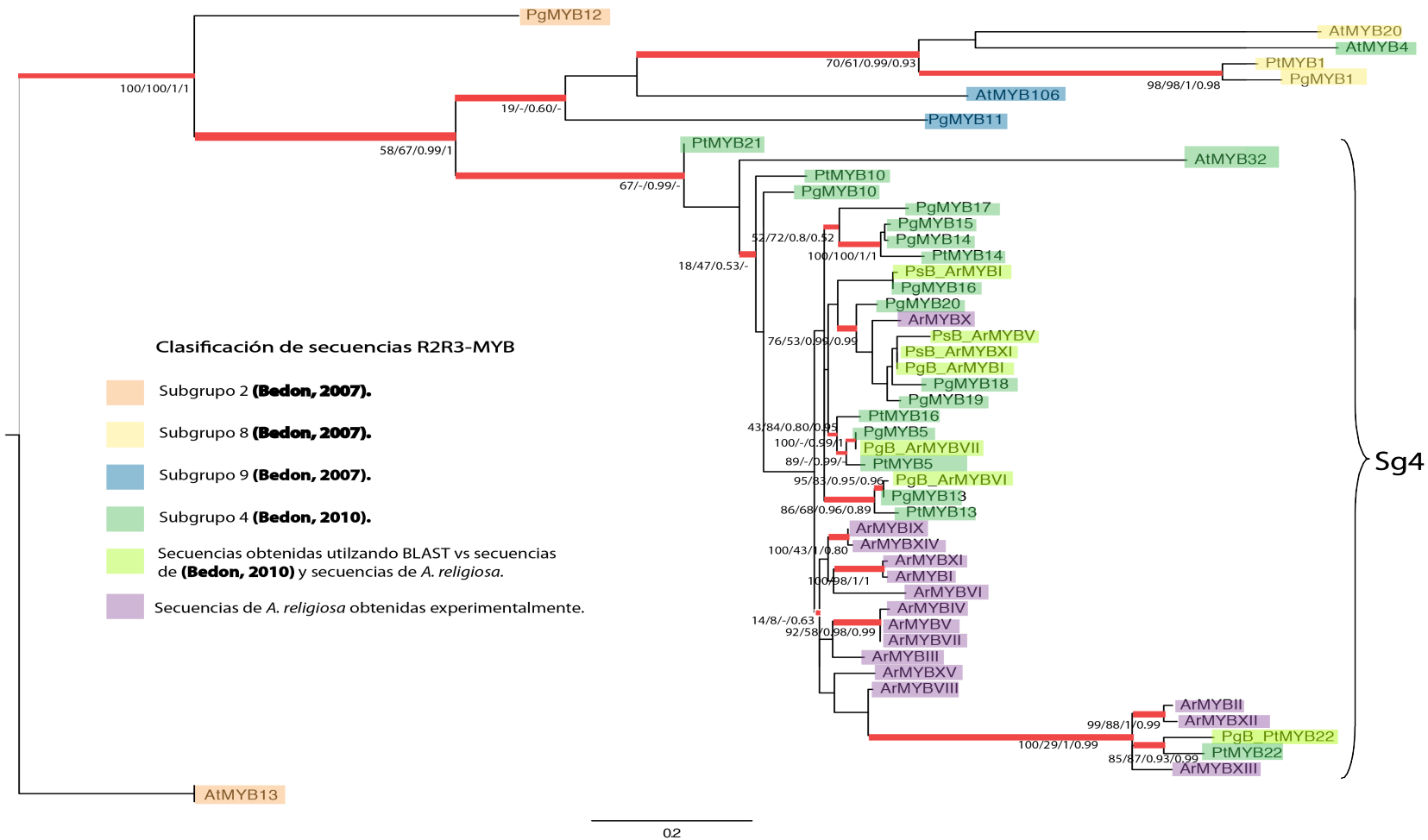


Figura 4: Reconstrucción filogenética de ML para secuencias R2R3-MYB, utilizando los 3 exones (222 pb o 74 aa). Las ramas engrosadas en rojo señalan los nodos coincidentes en 2 o más análisis, a un lado de estas ramas se reportan los valores de soporte estadístico en el siguiente orden Bootstrap (BS) de ML para alineación de nucleótidos/ BS de ML para aminoácidos/ Probabilidad Posterior de Bayesianos para nucleótidos/ PP de bayesianos para aminoácidos. At=*Arabidopsis thaliana*,Pg=*Picea glauca*, Ps=*Picea sitchensis*,Pt=*Pinus taeda*, Ar=*Abies religiosa*. Los colores señalan el origen o clasificación de las secuencias. La llave Sg4 señala secuencias de distinto origen que se piensa se agrupan funcionalmente en el subgrupo 4.

El árbol mostrado tiene 37 nodos resueltos, de los cuales 22 coinciden en 2 o más de los análisis filogenéticos, y están soportados en su mayoría por bootstraps y/o probabilidades posteriores cercanas a 100 y 1, respectivamente. La excepción más interesante es el nodo que agrupa a casi todas las secuencias ArMYB, con PgB_PtMYB22 y PtMYB22, que si bien tiene bajo soporte está presente en 3 de 4 de los análisis filogenéticos, siempre formando subgrupos compuestos por estas ArMYBs y los MYB22 de *Picea glauca* y *Pinus taeda*.

Como ya se ha dicho, todas las secuencias obtenidas experimentalmente para *A. religiosa* se agrupan dentro del gran clado descrito como Sg4, junto con las secuencias de *Pinus* y *Picea*. Sin embargo al interior de este puede distinguirse que la mayoría de ArMYBs forman subclado junto con los genes PtMYB22 y PgB_PtMYB22 que se puede separar a su vez en tres grupos monofiléticos (ArMYB XIV, IX, XI, I, VI); (ArMYB IV, V, VII, III, XV, VIII) y (ArMYB II, XII, XIII, PtMYB22 y PgB_PtMYB22). Solo la secuencia ArMYBX no pertenece a este grupo, y se agrupa cercanamente, dentro del subclado que contiene a casi todas las secuencias de *Pinus* y *Picea*. Todas las secuencias de coníferas anotadas como pertenecientes al subgrupo 4, así como las secuencias no anotadas que fueron descargadas de GenBank y las secuencias ArMYB obtenidas experimentalmente, forman un grupo monofilético, tal como lo señala la llave Sg4 en la figura 5. Ninguna secuencia de otro subgrupo de R2R3-MYB se encontró dentro de este clado en ninguno de los análisis filogenéticos.

4.3 Señales de recombinación:

Los 5 métodos de detección utilizados identificaron 32 eventos de recombinación de los cuales 16 fueron significativos ($P < 0.05$) y soportados por al menos 2 estadísticos diferentes (**Tabla 4; Figura 5**). Las supuestas secuencias recombinantes, secuencias parentales, los posibles puntos de quiebre de la recombinación, la probabilidad posterior asociada, se muestran en la tabla 4.

Tabla 3: Posibles eventos de recombinación detectados

Metodo de detección	Secuencia recombinante	Secuencia parental 1	Secuencia parental	Región recombinante	Valor de P	Soportado por otros métodos
GeneConve	X	VIII	UNK(VI)	60-127	4.10 x 10e-2	chim,siscan, 3seq
MAXCHI	IX	VIII	X(64.2%)	360-510	1.94 x 10e-2	chim, 3seq
	X(INV IX)	VIII	XIV (67%)	360-525	8.71 x 10e-2	chim, 3seq
	VIII(INV IX)	XIV (88.9%)	UNK(X)	380-525	4.10x 10e-2	chim, 3seq
	XIV	VIII	X(67%)	380-525	1.94 x 10e-2	chim, 3seq
	X(INV XIV)	UNK(VIII)	XIV (67%)	380-525	8.741 x 10e-3	chim, 3seq
	VIII(INV XIV)	XIV (88.9%)	UNK(X)	380-525	4.10 x 10e-2	chim, 3seq
	VI	XV(84.8%)	UNK(II)	410-530	6.798 x 10 e-3	chim, siscan
	II (VI)	UNK (VI)	XV (42.9%)	410-530	1.062 x10-2	chimaera, siscan
CHIMAERA	I*	X	XIV(84.7%)	60-370	1.103 x 10 e-2	chim,siscan
	X(INV I)	XI(83.7%)	UNK(XIV)	65-370	1.559 x 10e-2	chim,siscan
	XI	X	XIV(84.7%)	60-370	1.103 x 10 e-2	chim, siscan
	XIV	VIII(90.2%)	UNK(X)	360-560	2.440 s 10e-2	maxchi
SisCan	X	XI(83.9%)	Unk(XIV)	25-200	4.466 x 10e-3	3seq
	XI	X(83.9%)	XIV	25-200	4.477 x 10e-3	3seq
3seq	X	Unk(XIV)	XI(83.4%)	1-60, 370-611	4.235 x 10e-2	Chimaera

Tabla 4: Posibles eventos de recombinación detectados entre las secuencias de la subfamilia R2R3-MYB en *Abies religiosa*. Las secuencias recombinantes marcadas con (INV) son las que originalmente fueron detectadas como parentales de la secuencia marcada a un lado, y fueron invertidas y analizadas como recombinantes. El porcentaje entre paréntesis en las columnas 3 y 4 hacen referencia al porcentaje de similitud de dicha secuencia con respecto a la recombinante. La marca UNK en estas mismas columnas, señala aquellas secuencias que no tienen suficiente soporte estadístico para ser señaladas con certeza como secuencias parentales.

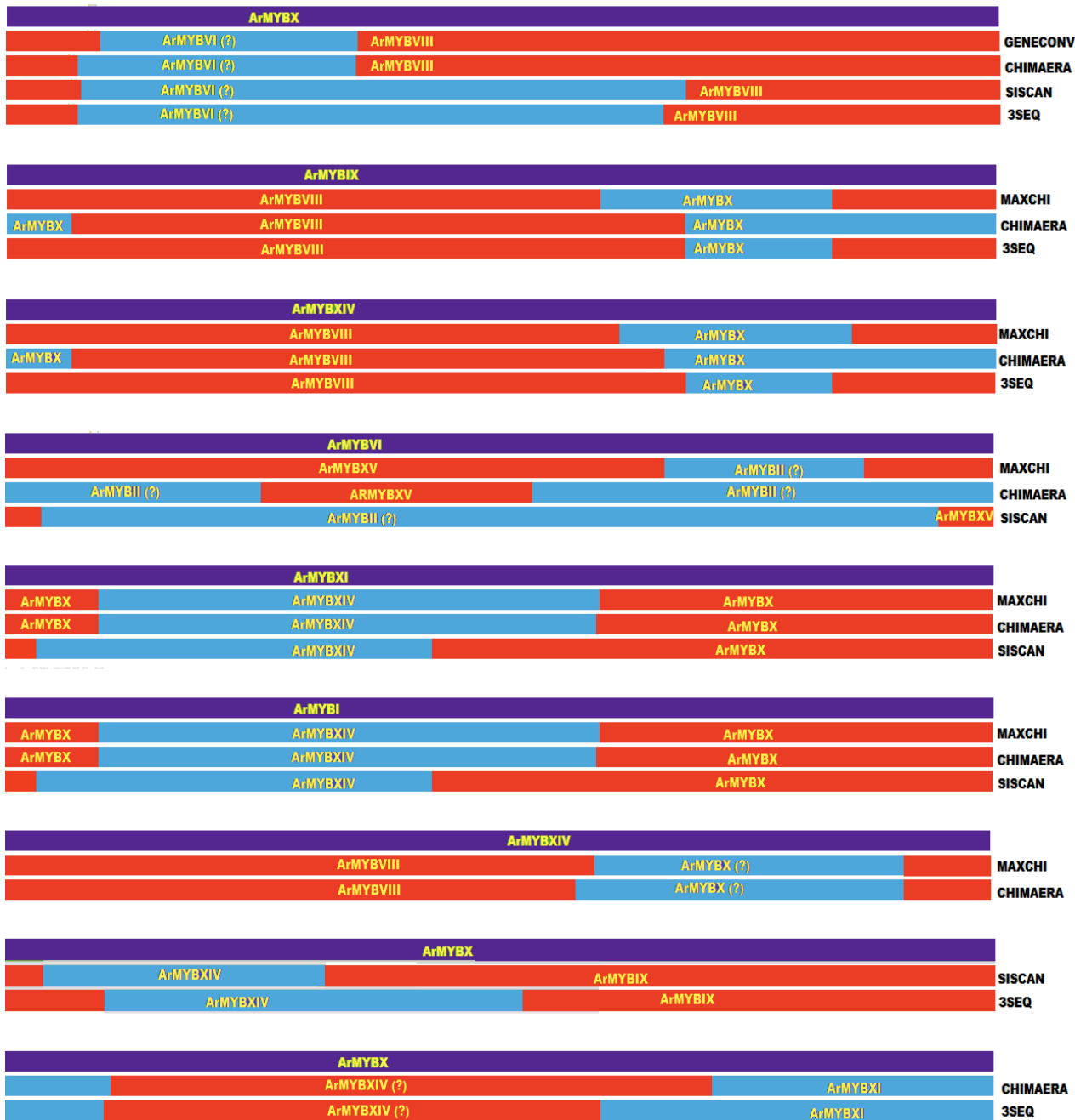


Figura 5: Representación de eventos únicos de recombinación (no se incluyen las inversiones reportadas en la tabla 4). Las barras moradas esquematizan a las secuencias detectadas como recombinantes. Debajo de cada una de éstas, se muestra la posible composición a partir de dos secuencias parentales representadas con azul y rojo, a un lado de las barras bicolor se reporta el método de detección que predice esta conformación. El punto de quiebre (contacto entre barra azul y barra roja) está localizado de manera proporcional a la longitud total de la secuencia; su ubicación exacta se reporta en la tabla 3.

Las pruebas de recombinación revelaron que los puntos de quiebre se dan generalmente en el primer y último tercio de las secuencias, principalmente cerca de las posiciones 60 y 400, y siempre en alguno de los dos intrones de la molécula. En algunos casos la región recombinante se puede extender hasta el siguiente exón. Por ejemplo la secuencia ArMYBX formada a partir de ArMBVI y ArMYBVIII. Por otro lado, se identificaron algunos genes que aparecen recurrentemente en las pruebas como lo son ArMYBX, ArMYBVIII, ArMYBXI y ArMYBXIV, que son en general secuencias que se encuentran separadas de las demás, formando ramas basales dentro de sus subclados, y son independientes entre sí, como lo muestra la reconstrucción en la figura 4.

4.5 Adquisición de intrones:

Las topologías del árbol obtenido por ML de los dos intrones concatenados (figura 5) y de la región codificante (figura 3) resultaron muy similares en cuanto a los subgrupos y distancias evolutivas de las secuencias. Los intrones de ArMYBX estuvieron más diferenciados y presentaron una distancia evolutiva mucho mayor con respecto a todas las demás. Al igual que en el árbol de la figura 3, el nodo de menor soporte estadístico fue el que agrupa a todos los ArMYB exceptuando a ArMYBX. La diferencia más notable entre las topologías de la figuras 3 y 5, consistió en que las secuencias ArMYB I, VIII, IX y XV, se encontraron resueltas en el árbol de intrones como secuencias más basales, aunque esto pudo deberse a la falta de enraizamiento de este árbol, que no contó con un grupo externo.

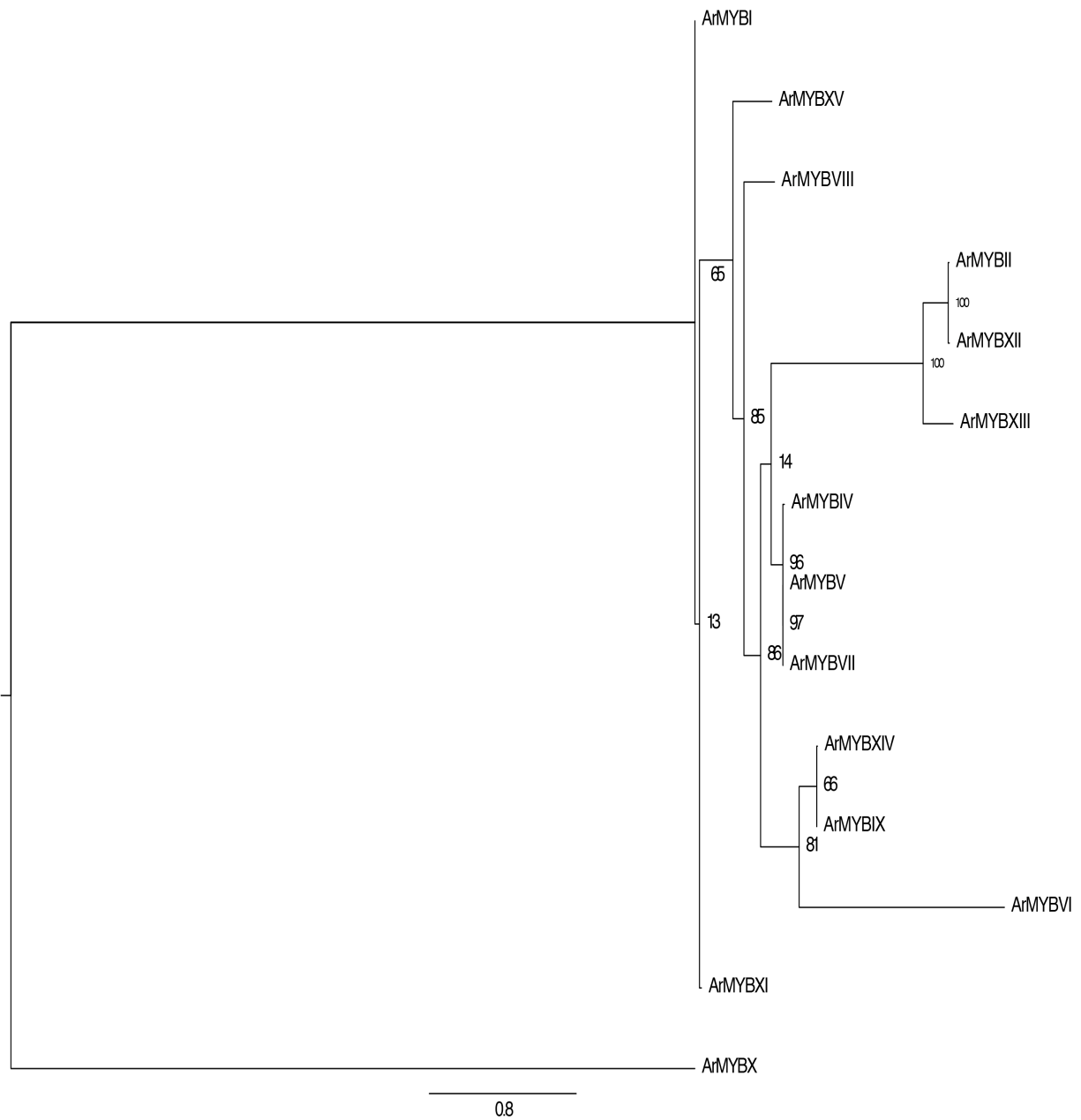


Figura 6: Árbol de máxima verosimilitud (ML) de los intrones concatenados de los genes de la subfamilia R2R3-MYB de *Abies religiosa*. El valor de bootstrap, a partir de 1000 replicas, se reporta en porcentaje para cada nodo del árbol. No se incluyó en el análisis a ArMYBIII, ya que el intrón 1 de este gen resultó demasiado largo e impidió que se pudiera secuenciar completamente el intron 2.

4.5 Secuencias repetitivas y/o móviles:

Ninguno de los programas utilizados detectó secuencias asociadas a transposones, retrotransposones o algún otro elemento móvil en ninguno de los MYBs de *A. religiosa*. Sin embargo, se hallaron varias secuencias repetitivas, con motivos entre 2 y 36 pares de bases; todas ellas dentro del primer intrón de las secuencias, estos datos se resumen en la Tabla 5. Cuatro de las secuencias encontradas pueden clasificarse como minisatélites, y otras 3 como microsatélites. En los genes ArMYBIV, ArMYBV y ArMYBVII, se presenta el mismo minisatélite de 22pb, estos tres genes son filogenéticamente cercanos y monofiléticos tanto en sus secuencias codificantes como intrónicas (figuras 4 y 6). Todas las demás secuencias repetitivas se presentan en solamente un gen.

Tabla 5: Secuencias repetitivas encontradas dentro de los genes de la subfamilia R2R3-MYB de *Abies religiosa*. Se incluyen el motivo estructural de repetición, y el número de veces que se repite (patrón), la región que ocupa dentro de las secuencias originales de nucleótidos (sitio), la longitud del motivo de repetición, la región del gen en la que se encuentra, y el programa que se utilizó para su detección.

Secuencia	Patrón	Sitio	Longitud	Región	Programa
ArMYBIII	TTCTTGATTTTCTGACGCAGCAAATTTTCTTGATG x 9.5	102-483	36pb	intron1	TRF
ArMYBIV	GATTATATATAAAGCTTTTCTT x 3	87-149	22pb	intron1	TRF
ArMYBV	GATTATATATAAAGCTTTTCTT x 3	67-129	22pb	intron1	TRF
ArMYBVII	GATTATATATAAAGCTTTTCTT x 3	88-150	22pb	intron1	TRF
ArMYBIX	AT x 4.5	57-65	2pb	intron1	PMRF
ArMYBX	TTTTTC x 4	52-71	20pb	Intron1	Repeatmasker
	TTGTAAA x 2.5	90-106	7pb	Intron1	PMRF

5 Discusión

5.1 Relaciones filogenéticas de R2R3-MYB

5.1.1 Agrupamiento y posible evolución de secuencias dentro del subgrupo 4 de R2R3-MYB

El primer patrón que emerge de las relaciones reconstruidas entre los genes R2R3-MYB analizados, es que todas las secuencias obtenidas de *Abies religiosa*, así como las secuencias anotadas y no anotadas de *Picea* y *Pinus*, se agrupan en un grupo monofilético (llave Sg4 en figura 3) anteriormente llamado subgrupo 4 (**Bedon et al., 2007; Bedon et al., 2010**). Esto sugiere que estos genes están y podrían llevar acabo roles similares en cada una de estas especies, como la regulación de la expresión de componentes del parénquima en situaciones de estrés (**Bedon et al., 2010**) (**Bomal et al., 2008**). Para corroborar esto, es necesario realizar experimentos de perfiles de transcripción y/o microarreglos frente a condiciones de stress inducido.

De interés aún mayor, resulta la topología al interior de este grupo. En particular el agrupamiento de casi todos los genes ArMYB en un grupo independiente de la mayoría de las secuencias de *Pinus taeda*, *Picea glauca* y *Picea sitchensis*. También resalta el hecho de que en ningún caso se agrupen secuencias de *A. religiosa* en parejas o tripletes con secuencias de *P. taeda* y/o *P. glauca*, como si ocurre con las otras coníferas: PtMYB1 y PgMYB1; PtMYB14 y PgMYB14; PtMYB5 y PgMYB5; y PtMYB13 con PgMYB13 (**Bedon et al., 2007**). Esto implica que no hay un reconocimiento claro de ortólogos individuales en los genes de ArMYB con contrapartes de *Pinus* o *Picea glauca*, y pudiera implicar que probablemente este subgrupo de R2R3-MYB evolucionó de manera independiente, luego de la separación de las subfamilias de pinaceas, *Pinoideae* y *Abiетоideae*. Es posible que las duplicaciones génicas que dieron origen a la diversificación de este subgrupo se dieron posteriormente a la separación de estas subfamilias, y que fueron

mantenidas y ampliadas durante los 300 millones de años posteriores a la divergencia, ya sea por el efecto filogenético o bien por las condiciones ambientales contrastantes a las que se encuentran sometidas las especies boreales (*Pinus taeda* y *Picea glauca*) y subtropicales (*Abies religiosa*). Llevar a cabo análisis de los genes homólogos de pinoideas tropicales y subtropicales, así como de abietoideas boreales podrían ser valiosos para completar el sistema, y ayudaría a resolver la interrogante acerca de la naturaleza de la divergencia, ¿se debe esta última únicamente al efecto filogenético de la separación de clados hace 300 millones de años, o bien existe un componente de adaptación ambiental involucrado? Esta hipótesis está, apoyada por la evidencia previa de que los genes MYB pueden estar implicados en respuestas adaptativas en coníferas analizadas en clinas ambientales **(Prunier et al., 2011; Namroud et al., 2008)**.

El contraste marcado de las familias génicas de pinoideas y abietoideas ya había sido detectada anteriormente. Por ejemplo, existe evidencia de patrones divergentes en los patrones de diversificación de los genes nucleares 4CL (4-coumarato: coenzima A sintasa, responsable de la vía biosintética de la lignina) , ya que puede ser identificada una duplicación génica para los géneros *Abies* y *Pseudotsuga* (Abietoideae) no presente en otros géneros de pináceas **(Wang, et al., 2000)**. Por otro lado, también hay información acerca de la naturaleza diferenciada en marcadores bioquímicos, como es el caso de los ácidos grasos de las semillas, moléculas que si bien se encuentran presentes en todas las pináceas, despliegan diferencias en cuanto a su abundancia relativa al comparar a las subfamilias Abietoideae y Pinoideae, lo cual a su vez podría ser reflejo de las diferencias en las familias génicas encargadas de la síntesis de dichos lípidos en estos taxa. **(Wolff et al., 2000)**.

5.1.2 Procesos que pueden afectar la reconstrucción filogenética

No obstante que las secuencias ArMYB se agruparon monofiléticamente, varios de los nodos al interior del grupo no se encontraron resueltos, ya que tuvieron un valor de soporte bajo y/o se desplegaron en una politomía. Esto puede deberse a que la variación dentro de la familia génica no posee la información suficiente para resolver las relaciones de parentesco, ya sea por el tamaño de la secuencia, por el proceso de sustitución nucleotídica propio de estos genes.

En la figura 3, son notorios dos casos de aparente incongruencia filogenética, pero que quizás estén aportando información acerca de la historia general de esta familia génica. En primer lugar tenemos a PtMYB22, que junto con una secuencia no anotada de *Picea glauca* (Pgb_PtMYB22), se encuentra dentro de un grupo formado por casi todas las secuencias de *Abies religiosa*. Esto puede ser explicado por tres escenarios. El primero es que PtMYB22 puede ser una secuencia muy divergente con respecto a las de otros genes del subgrupo 4 de *Pinus* y *Picea*, de manera que la posición de la rama que lo sustenta sea un simple artefacto de los métodos de reconstrucción filogenética (**Sanderson y Doyle, 1992; Baker et al., 2000**) sin embargo esto resulta poco probable, ya que se cuenta con un muestreo amplio de secuencias y una reconstrucción de máxima verosimilitud muy robusta. Por otro lado, la presencia de PtMYB22 y de la secuencia no anotada de *Picea glauca* en el clado formado por la mayoría de genes de *Abies* puede ser evidencia de que la duplicación génica ocurrió antes de la cladogénesis de las Subfamilias *Abetoideae* y *Pinoideae* y que en el último ancestro común de estas taxa existió un polimorfismo. Posteriormente, una de las formas de este polimorfismo diversificó en casi todos los miembros del subgrupo 4 en *Pinoideae*, mientras que la otra forma dio origen únicamente a PtMYB22 (y sus ortólogos). En *Abetoideae* se habría dado el proceso de diversificación opuesto, lo cual también explicaría la presencia de ArMYBX lejos de las otras secuencias de *Abies*.

Este último patrón puede estar apoyado por el hecho de que la retención y diversificación de polimorfismos ancestrales es común en coníferas (**Bouillé y Bousquet, 2005; Jaramillo-Correa et al., 2008**). Lo cual es atribuible a las bajas tasas de evolución de estas plantas y a los grandes tamaños efectivos ancestrales, que hacen que la tasa de pérdida o fijación de polimorfismos por deriva génica sea muy baja (**Hamrick et al., 1992; Perron et al., 2000**). Es interesante pensar que este patrón puede presentarse no sólo a nivel de genes, sino que también puede ser el origen de subfamilias o familias génicas completas.

Un tercer escenario que puede explicar los patrones de aparente incongruencia filogenética, es que el muestreo de secuencias llevado a cabo en *Abies* haya sido incompleto, faltando otras secuencias hipotéticas que se encuentren relacionadas más estrechamente con algunos miembros de *Picea glauca* y *Pinus taeda*, y que por lo tanto, la localización de ArMYBX en la filogenia estaría representando una relación de ortología verdadera, al igual que PtMYB22 y PgB_PtMYB22 con respecto a ArMYBII, ArMYBXII y ArMYBXIII. Lo que implicaría también que la subfamilia ArMYB es mucho más diversa. Sin embargo, lo anterior resulta poco probable, ya que todo apunta a que el muestreo de genes MYB dentro de los genomas de coníferas ha sido bastante completo. **Bedon et al. (2010)** contaban con el transcriptoma completo de *Picea glauca* cuando identificaron los 10 MYBs del subgrupo 4 en esta especie, la presente tesis complemento este conjunto con al menos otras 3 secuencias diferentes de *P. glauca* obtenidas por una búsqueda en Blast, a la vez que se obtuvieron 15 secuencias diferentes de *Abies religiosa*, lo cual ubica la extensión de la familia génica en el mismo orden de magnitud para ambas especies, y muy difícilmente se esperaría encontrar las secuencias suficientes como para identificar un ortólogo de *Picea* y *Pinus* para todas las secuencias de *Abies* y viceversa. Por lo cual es muy probable que se tengan todas o al menos la mayoría de genes R2R3-MYB del subgrupo 4 en este análisis, y que las relaciones filogenéticas (al menos las sustentadas por el

dominio analizado) estén reflejando la evolución independiente entre estas secuencias.

5.2 Recombinación no homóloga de genes R2R3-MYB en *A. religiosa*

El hallazgo de señales de posibles recombinaciones no homólogas en 9 de las 15 secuencias de genes MYB de abetos, nos hace pensar que la recombinación ha sido un agente que ha intervenido en su evolución molecular. Este proceso ya ha sido confirmado para algunas familias génicas en plantas, por ejemplo en un modelo experimental de genes construidos sintéticamente conocidos como RBCSB en *A. thaliana* (**Jelesko et al. 2003**) así como para varias familias génicas en maíz (**Emrich et al., 2007**), implicando que es de gran relevancia para la generación de diversidad

El hecho de que estos procesos ocurran tan a menudo como se observó aquí podría explicar la limitada resolución filogenética que se pudo alcanzar entre los genes ArMYB, ya que las relaciones de parentesco podrían ser reticuladas si consideramos la existencia de secuencias quiméricas originadas por la recombinación de dos genes no homólogos. Siendo por lo tanto imposible representarlas mediante un árbol filogenético. (**Posada et al., 2002; Gribaldo y Phillipe, 2002**)

En un sentido más amplio, estos resultados nos llevan a pensar que la recombinación no homóloga es posiblemente un mecanismo que ha influido en la evolución de esta subfamilia de genes. Este proceso generaría no solamente nuevas secuencias, con posibles nuevas funciones, de manera directa, sino también nueva variación sobre la cual otras fuerzas evolutivas pudieran actuar y diversificar a este grupo de genes. Por ejemplo, el gen *pericarp color1* que codifica un factor de transcripción R2R3-MYB en *Zea mays*, y que esta encargado de regular la vía biosintética de flavonoides en órganos florales, también ha sido objeto de eventos de recombinación con parálogos, generando con esto nuevas formas alélicas, que varían en sus

patrones de expresión entre tejidos. En vista de que las regiones recombinantes suelen ser elementos reguladores como *enhancers*, la recombinación en este gen tiene repercusiones fenotípicas en la coloración de la mazorca del maíz (Zhang y Peterson, 2005). Entonces, si procesos similares se están dando en los genes ArMYB, estos podrían estar originando patrones de expresión y de regulación génica más finos, y por lo tanto ser claves para la adaptabilidad de la especie.

5.3 Adquisición de intrones y secuencias repetitivas en los ArMYB.

La comparación entre los árboles filogenéticos de la figura 4 y 6 sugiere, por su similitud en las topologías, que no se han presentado eventos recientes de adquisición de intrones en los genes analizados, y que las regiones no codificantes y codificantes han evolucionado de manera conjunta durante la diversificación de la subfamilia génica en *A. religiosa*.

Sin embargo, parece ser que estos intrones no son entidades que han evolucionado “pasivamente” dentro de los genes ArMYB, ya que algunos patrones y procesos resaltan a este respecto. En primer lugar es notable que el punto de quiebre de los eventos de recombinación detectados ocurren únicamente en estas regiones de las secuencias, y que la formación de quimeras recombinantes fue detectada generalmente entre genes distantes en la filogenia, sin que esto quiera decir que no exista entre secuencias más cercanas, sin embargo nos enfrentamos en este caso a una limitación metodológica ya que quizás son más comunes, pero no pueden ser detectadas por el alto grado de similitud, todo esto que nos lleva a pensar que posiblemente las regiones intrónicas tienen una historia común ancestral debido a que se adquirieron como un conjunto en los genes ArMYB (quizás en algún momento cercano a la diversificación de la familia génica), razón por la cual los eventos recombinantes ocurren sistemáticamente en estos sitios, aunque las secuencias hayan divergido significativamente durante el resto de su historia evolutiva. De hecho la existe la posibilidad de que la presencia de

“hot spots” de recombinación no homóloga en intrones sea el motor inicial de divergencia evolutiva sean los entre genes (dada la alta incidencia con la que se encuentran puntos de quiebre en estas regiones).

La modulación de elementos repetitivos en tandem parece ser un mecanismo que en mayor o menor grado ha influido en la evolución de la subfamilia R2R3-MYB. Particularmente en el primer intrón, en donde aparecen repeticiones regiones de entre 9 y 385 pb en 6 secuencias. Este patrón resalta una vez mas la importancia de los intrones como “hot spots” mutacionales. Esto ya se ha visto antes en los intrones de otros genes de factores de transcripción de plantas, como es el caso de APETALA1 y APETALA3 en asteraceas de la tribu Madiinae (**Barrier et al., 2000**).

Algunos estudios han señalado que la modulación de microsatélites puede ocurrir al interior de genes, y que la distribución de estas secuencias repetitivas no es aleatoria. Incluso se ha propuesto que pueden propiciar la pérdida o ganancia de funciones por cambios en el marco de lectura de un gen, así como modificar sus patrones de transcripción cuando se encuentran en su región reguladora flanqueante, o promover efectos epigenéticos, como la compactación del DNA en heterocromatina (**Li et al., 2003**), siendo esto una primera razón de interés para buscar este tipo de secuencias dentro de familias génicas.

Por otro lado, existe una clara relación de la presencia de secuencias repetitivas con eventos de recombinación (**Majewski y Ott, 2000**) (**Brandström et al., 2008**), aunque todavía no es clara la cadena causal de este proceso. En esta tesis encontramos este patrón en una dinámica entre tres elementos: la evolución de los intrones, la modulación de secuencias repetitivas, y los eventos de recombinación que ocurren simultáneamente en una región. Siendo posible que la extensión y la abundancia de los microsatélites promueva un mayor número de eventos de recombinación, como es el caso de la secuencia ArMYBX.

Aunque en este estudio no se encontró evidencia de elementos transposones dentro de las secuencias, es recomendable ampliar el esfuerzo de muestreo para detectar si en algún gen MYB (o en las regiones del genoma que tienen influencia sobre este gen) se presenta algún elemento móvil, ya que también se encuentra reportado que estas secuencias están correlacionadas con la presencia de microsatélites y que pueden tener efectos funcionales sobre los genes (**Juretic *et al.*, 2005; Morgante *et al.*, 2006**).

6 Conclusiones y perspectivas

6.1 Conclusiones

El presente trabajo logró hacer una aproximación inicial a la descripción de los genes R2R3-MYB de *Abies religiosa*. Lo anterior representa un avance importante tanto en términos de ecología molecular como de biología evolutiva. Recordemos que en este campo no existía nada de información para coníferas no pinoideas, ni para especies de pináceas subtropicales. Esto es relevante dada la naturaleza de los procesos regulados por el grupo de genes investigados, así como por las condiciones ambientales en las que se distribuye la especie estudiada.

El primer aporte y el producto más importante de este trabajo son las 15 secuencias nuevas de genes de copia única hallada para *Abies religiosa*. Con las que en el futuro se podrá empezar a buscar variación y realizar estudios evolutivos desde la óptica de la genética de poblaciones

Con la caracterización de la secuencia de algunos de sus miembros, se confirmó la presencia de la subfamilia R2R3-MYB en abetos. Asimismo se comprobó que todas las secuencias obtenidas se agrupan cercana y monofiléticamente con secuencias que fueron clasificadas con anterioridad dentro del subgrupo 4 en otras coníferas. Se observó que los ArMYBs son un grupo parcialmente independiente de las secuencias homólogas de otras especies analizadas, lo que puede señalar divergencias funcionales debidas a procesos adaptativos en condiciones ambientales diferentes y/o a la diversificación independiente a partir de diferentes genes ancestrales .

Se detectó también que la recombinación no homóloga entre parálogos ha generado nuevas secuencias, que combinan partes de genes parentales distintos, lo que impacta directa y positivamente a la cantidad de variación de

esta familia

Asimismo se encontró que los intrones de los genes ArMYB han jugado un papel importante en la evolución del grupo, ya que es precisamente en estos sitios donde se concentra una gran parte de los puntos de quiebre de recombinación y las secuencias repetitivas encontradas.

6.2 Perspectivas

Queda mucho por hacer dentro de la caracterización de genes R2R3-MYB en abetos. En primer lugar se propone que trabajos futuros exploren con mayor detalle la variación de estas secuencias dentro de este género.

En este momento están siendo diseñados primers para la amplificación y secuenciación específica de cada uno de los genes ArMYB reportados en este estudio, para así poder explorar las variantes alélicas de cada uno de los genes y con esto, se podrían empezar estudios de genética de poblaciones para confirmar o descartar el rol adaptativo de estos genes en *Abies religiosa*, algo que está fuera de los objetivos de la presente tesis. Para esto se pueden realizar barridos genómicos y mapas de asociación con climas ambientales con el objetivo de buscar variables ambientales que puedan estar correlacionadas con la variación de esta familia génica.

Por otro lado, estos genes pueden ser usados como referentes para extender el muestreo de esta familia génica y hallar la secuencia completa de todos los parálogos ArMYB haciendo uso del transcriptoma de la especie, con lo cual se podrían terminar de resolver las consideraciones filogenéticas que quedaron abiertas en la discusión de este trabajo.

Por último y a largo plazo, es indispensable caracterizar funcionalmente estos genes mediante análisis de microarreglos (*microarrays*), no sólo para conocer la regulación génica de *A. religiosa*, sino también para contrastar estos

patrones con aquellos que ya son conocidos en coníferas boreales y así discernir que tan comunes o divergentes han sido sus procesos evolutivos.

Se reitera una vez más que este trabajo plantea un primer vistazo para entender la evolución de un grupo de genes desde el enfoque molecular, pero todo lo dicho adquirirá trascendencia solamente cuando trabajos futuros lo proyecten hacia su implicación en procesos funcionales, ecológicos y evolutivos más amplios.

¿Cómo un gen o grupo de genes, puede ayudar a explicar la adaptación, funcionamiento e interacción de un organismo en su contexto fisiológico y ambiental? Esto es el corazón mismo de la biología molecular evolutiva.

Referencias

Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics*: **21(9)**:2104-2105.

Abe H et al. (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* **15**, 63–78.

Adams KL and Wendel JF (2005) Polyploidy and genome evolution in plants, *Current Opinion in Plant Biology*, **8**, 135-141.

Agarwal M, Hao Y, Kapoor A, Dong CH, Fujii H, Zheng X, and Zhu JK (2006) A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. *The Journal of biological chemistry*, **281**, 37636-45.

Aguirre-Planter E, Jaramillo-Correa JP, Gómez-Acevedo S, Khasa DP, Bousquet J, Eguiarte LE (2012). Phylogeny, diversification rates and species boundaries of Mesoamerican firs (*Abies*, Pinaceae) in a genus-wide context. *Molecular Phylogenetics and evolution*, **62**, 263-274.

Ahuja MR and Neale DB (2005) Evolution of Genome Size in Conifers. *Silvae Genetica*, **54** **3**, 126-137.

Albouyeh R, Ritland K.(2011). Heritability and species divergence for gene expression of spruce terpenoids are highly correlated, indicating adaptive divergence of a key gene family involved in insect defense. Submitted to *Journal of Heredity* (June 2011).

Anderson JB and Browler, LP (1996) Freeze-protection of overwintering monarch butterflies in México: critical role of the forest as a blanket and an umbrella. *Ecological Entomology* **21**, 107–116.

Baker WJ, Hedderson TA, Dransfield J (2000) Molecular Phylogenetics of Subfamily Calamoideae (Palmae) Based on nrDNA ITS and cpDNA rps 16 Intron Sequence Data. *Molecular Phylogenetics and Evolution*. **14**, 195-207

Barrier M, Friar E, Robichaux R, Purugganan M (2000) Interspecific evolution in plant microsatellite structure. *North*, **241**, 101-105.

Baumann K. et al. (2007) Control of cell and petal morphogenesis by R2R3 MYB transcription factors. *Development* **134**, 1691–1701

Beaumont MA and Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. London Ser. B. Biol. Sci.* **263**, 1619–1626.

Bedon F, Bomal C, Caron S, Levasseur C, Boyle B, Mansfield SD, Schmidt A, Gershenzon J, Grima-Pettenati J, Séguin A, and MacKay J (2010) Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid-oriented responses. *Journal of experimental botany*, **61**, 3847-64.

Bedon F, Grima-Pettenati J, and Mackay J, (2007) Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*), *BMC plant biology*, **7**, 17.

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**, 573-580

Bennett MD and Leitch IJ (2003) Angiosperm DNA C-values database. <http://www.rbgekew.org.uk/cval/homepage.html>

Bills DJ (2003). The development of a global trade. *Acta Hort.* **615** 339-347

Bomal C, Bedon F, Caron S, Mansfield SD, Levasseur C, Cooke JEK, Blais

S, Tremblay L, Morency MJ, Pavy N, Grima-Pettenati J, Séguin A, Mackay J (2008) Involvement of *Pinus taeda* MYB8 and MYB1, in Phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative in planta analysis. *Journal of Experimental Botany*, **59**, 3925-3939.

Bouillé M and Bousquet J (2011) Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintenance of genetic diversity in trees. *American Journal of Botany* **92(1)**, 63-73.

Brandström M, Bagshaw AT, Gemmell NJ, Ellegren H (2008) The relationship between microsatellite polymorphism and recombination hot spots in the human genome. *Mol Biol Evol*, **25(12)**, 2579-2587.

Buschiazzo E, Ritland C, Bohlmann J, Ritland K (2012) Show but not low: genomic comparisons reveals slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC Evolutionary Biology*. **12:8**

Byrne ME et al. (2000) Asymmetric leaves1 mediates leaf patterning and stem cell function in Arabidopsis. *Nature* **408**, 967-971

Charlesworth D, Mable BK, Schierup MH, Bartolomé C, Awadalla P (2003) Diversity and Linkage of genes in the Self-Incompatibility Gene Family in Arabidopsis Lyrata. *Genetics*. **164**, 1519-1535.

Cominelli E et al. (2008) Over-expression of the Arabidopsis AtMYB41 gene alters cell expansion and leaf surface permeability. *Plant Journal*, **53**, 53-64.

Connor KF and Lanner RM (1990) Effects of tree age on secondary xylem and phloem anatomy in stems of Great Basin bristlecone pine (*Pinus longaeva*). *American Journal of Botany*, **77**, 1070-1077.

De Vos M et al (2006) The *Arabidopsis thaliana* transcription factor AtMYB102 functions in defense against the insect herbivore *Pieris rapae*. *Plant Signal Behav.* **1**, 305–311.

Delgado P, Piñero D, Chaos A, Pérez-Nasser N, Alvarez-Buylla ER (1999) High population differentiation and genetic variation in the endangered Mexican pine *Pinus Rzedowskii* (Pinaceae). *American Journal of Botany*, **86**, 669-676.

Demuth JP, Bie TD, Stajich JE, Cristianini N, Hahn MW (2006) The Evolution of Mammalian Gene Families. *PLoS ONE* **1**(1): e85

Devaiah BN et al. (2009) Phosphate starvation responses and gibberellic acid biosynthesis are regulated by the MYB62 transcription factor in *Arabidopsis*. *Mol Plant* **2**, 43–58

Devey ME, Fidler TA, Liu BH, Knapp SJ and Neale DB (1994) An RFLP linkage map for loblolly pine based on three generation outbred pedigree. *Theor. Appl. Genet.* **88**, 273–278.

Du H et al. (2009) Biochemical and molecular characterization of plant MYB transcription factor family. *Biochemistry (Mosc)*, **74**, 1–11.

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, and Lepiniec L, (2010) MYB transcription factors in *Arabidopsis*, *Trends in Plant Science*, **Jul. 2010**, 1-9.

Duroux L, Welinder KG (2003) The peroxidase gene family in plants: A Phylogenetic Overview. *Journal of Molecular Evolution.* **57**, 397-407

Eckenwalder JE (2009). Conifers of the world: the complete reference. Edición ilustrada de Timber Press, 720 pp

Emrich SJ, Li L, Wen TJ, Yandeu-Nelson MD, Fu Y, Guo Ling, Chou HH, Aluru S, Ashloch D, Schnable PS (2007) Nearly identical Paralogs: Implications for Maiz (*Zea mays L.*) Genome evolution. *Genetics*. **175(1)**, 429-439.

Eisen JA (1998) Phylogenomics: Improving Functional Predictions for Uncharacterized Genes by Evolutionary Analysis. *Genome Res*. **8**, 163-167

Farjon A (1998) World Checklist and Bibliography of Conifers, The Royal Botanic Garden, Kew.

Farjon A (2010) A Handbook of the World's Conifers. Brill Academic Publishers, Leiden.

Food and Agriculture Organization (FAO)(2008) Managing forest for climate change, publicación pública de la organización. www.fao.org/forestry/tof

Futamura N, Ujino-ihara T, Nishiguchi M, Kanamori H, Yoshimura K, Sakaguchi M, Shinohara K, Anderson TES (2006) Analysis of expressed sequence tags from *Cryptomeria japonica* pollen reveals novel pollen-specific transcripts. *Tree Physiology* **26**. 1517-1528.

Gernandt DS, Willyard A, Syring JV, Liston A., Chapter 1 en Plomio C, Bousquet J, Kole C (ed) (2011) Genetics, Genomics, and Breeding of Conifers. *CRC Press*,. NY.

Gernandt DS, Magallón S, Geada G, Zerón O, Willyard A, Liston A (2008) Use of simultaneous analices to guide fossil-based calibrations of Pinaceae phylogeny. *Int. J. Plant Sci*. **169(8)**, 1086-1099.

Gibson TJ and Spring J (2000) Evidence in favor of ancient octoploidy in the vertebrate genome. *Biochem Soc. Tans*. **28**, 259-264.

Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, Lapierre C, Pollet B, Verhaegen D, ChaubetGigot N, and Grima-Pettenati, J (2005) EgMYB2, a new transcriptional activator from Eucalyptus xylem, regulates secondary cellwall formation and lignin biosynthesis. *Plant J.* **43** 553–567.

Gouy M, Guindon S & Gascuel O (2010) SeaView version 4 : a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, **27(2)**, 221-224.

Gribaldo S, Phillipe H (2002) Ancient Phylogenetic Relationships. *Theoretical Population Biology.* **61**, 391-408.

Grotkopp M, Rejmánek MJ, Sanderson, and Rost TL (2004) Evolution of genome size in pines (Pinus) and its life-history correlates: supertree analyses. *Evolution; international journal of organic evolution.* **58**, 1705-29.

Guasconi V, Yahi H, Ait-Si-Ali S . (2003) Transcription factors. *Atlas Genet Cytogenet Oncol Haematol.*

Hairton N, Stephen S, Kearns C. (1996) “Overlapping generetions: the storage effect and the maintenance of biotic diversity” en **Rhodes O, Chesser R, Smith M.** Populations Dynamics in Ecological Space and Time The University of Chicago Press.

Hammond JP and White PJ (2011) Sugar signaling in root responses to low phosphorus availability. *Plant physiology*, **156**, 1033-40.

Hamrick JL (1987) Gene flow and distribution of genetic variation in plant populations. En **Urbanska K** Differentiation patterns in higher plants, 53–67, Academic Press, New York, New York, USA

Hamrick JL, Godt MJW, Sherman-Broyles SL. (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests* **6**, 95–124

Harrison SCA (1991) Structural taxonomy of DNA-binding domains. *Nature* **353** 715-719.

Ito M (2005) Conservation and diversification of three-repeat Myb transcription factors in plants. *Journal of Plant Research* **118**, 61–69.

Jaffé FW, Tattersall A, and Glover BJ (2007). A truncated MYB transcription factor from *Antirrhinum majus* regulates epidermal cell outgrowth. *Journal of experimental botany*, **58(6)**, 1515-24.

Jaramillo-Correa JP, Aguirre-Planter E, Khasa DP, Eguiarte LE, Piñero D, Furnier GR, Bousquet (2008) Ancestry and divergence of subtropical montane forest isolates: molecular biogeography of the genus *Abies* (Pinaceae) in southern México and Guatemala, **17**, 2476-2490.

Jelesko JG, Carter K, Thompson W, Kinoshita Y, Grissem W. (2003). Meiotic Recombination Between Paralogous RBCSB Genes on Sister Chromatids of *Arabidopsis thaliana*. *Genetics*, **166**, 947-957.

Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, et al. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. *EMBO J*, **19**, 6150–61.

Juretic N, Hoen DR, Huynh ML, Harrison PM, Bureau TE (2005) The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. *Genome Res*, **15**, 1292-1297.

Kalendar R, Tankskanen J, immone S, Nevo E and Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA*, **97**, 6603–660.

Katoh, Asimenos, Toh (2009) Multiple Alignment of DNA Sequences with MAFFT. *Methods in Molecular Biology* **537**, 39-64.

Keller T et al. (2006) Arabidopsis REGULATOR OF AXILLARY MERISTEM S1 controls a leaf axil stem cell niche and modulates vegetative development. *Plant Cell*. **18**, 598–611

Kim S, Soltis PS, Wall K, Douglas ES (2006) Phylogeny and Domain Evolution in the *APETALA2*-like Gene Family. *Mol Biol Evol*, **23(1)**, 107-120

Kirik V et al. (2005) Functional diversification of MYB23 and GL1 genes in trichome morphogenesis and initiation. *Development* **132**, 1477–1485

Klempnauer KH, Gonda TJ, Bishop JM (1982) Nucleotide sequence of the retroviral leukemia gene *v-myb* and its cellular progenitor *c-myb*: the architecture of a transduced oncogene. *Cell*, **31**, 453-463.

Kossack DS and Kinlaw CS (1999) IFG a gypsy-like retrotransposon in *Pinus* (Pinaceae) has an extensive history in pines. *Plant Mol. Biol.*, **39**, 417–426.

Kovach A, Wegrzyn JL, Parra G, Holt C, Bruening GE, Loopstra CA, Hartigan J, Yandell M, Langley CH, Korf I, Neale BD (2010) The *Pinus taeda* genome is characterized by diverse and highly diverged repetitive sequences. *BMC Genomics*, **11**: 420.

Kriebel HB (1993) Molecular structure of forest trees en: **AHUJA MR and LIBBY WJ**, Clonal Forestry I. Genetics and Biotechnology.. (Eds) Springer Verlag, Berlin,. 224–240

Legay S, Sivadon P, Blervacq AS, Pavy N, Baghdady A, Tremblay L, Levasseur C, Ladouce N, Lapierre C, Séguin A, Hawkins S, Mackay J and

Grima-Pettenati J (2010) *EgMYB1*, an R2R3 MYB transcription factor from eucalyptus negatively regulates secondary cell wall formation in *Arabidopsis* and poplar. *New Phytologist*, **188**, 774–786.

Leach R (2000) Program Tandyman. *Los Alamos Nacional Laboratory*

Leitch IJ, Soltis DE, Soltis PS and Bennett MD (2005) Evolution of DNA amounts across land plants (Embryophyta). *Ann. Bot.*, **95**, 207–217.

Lerat E (2009) Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity*, **104**, 520-533.

Li SF et al. (2009) The *Arabidopsis* MYB5 transcription factor regulates mucilage synthesis seed coat development and trichome morphogenesis. *Plant Cell* **21**, 72–89

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol*, **21(6)**, 991-1007.

Lipsick JS(1996) One billion years of Myb. *Oncogene*, **13**, 223-235.

Liu L, White MJ, and Macrae TH (1999), Functional domains, evolution and regulation. **257**,

Liu T S (1971). A Monograph of the Genus *Abies*. Taipei.

Majewski J, Ott J (2000) GT repeats are associated with recombination on human chromosome 22. *Genome Res* **10**, 1108-1114

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, **26**, 2462-2463

Martínez M (1948). Los Abies mexicanos. *Anales del Instituto de Biología de México* **19**, 11–104.

Matus JM, Aquea F, and Arce-Johnson P (2008) Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes, *BMC plant biology*, **8**, 83.

McCarthy EM, McDonald JF (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* **19**, 362-367.

Miksche JP and Hotta Y (1993) DNA base composition and repetitious DNA in several conifers. *Cromosoma*, **41**, 29–36.

Millar AA, Gubler F (2005) The *Arabidopsis* GAMYB-like genes MYB33 and MYB65 are micro RNA-regulated genes that redundantly facilitate anther development. *Plant Cell* **17**, 705–721

Morgante M (2006), Plant genome organisation and diversity: the year of the junk!, *Current opinion in biotechnology*, **17**,168-73.

Moyano E, Martinez-Garcia JF, Martin C (1996) Apparent redundancy in myb gene function provides gearing for the control of flavonoid bio- synthesis in *Antirrhinum* flowers. *Plant Cell*, **8**, 1519–1532.

Muona O. and Harju A (1989) Effective Population Sizes, Genetic Variability, and Mating System in Natural Stands and Seed Orchards of *Pinus sylvestris*. *Silvae Genetica*, **38**, 221-228.

Namroud J, Beaulieu N, Juge J, Laroche and Bousquet J (2008), Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular ecology*, **17**, 3599-613.

Nardmann J, Zimmerman R, Durantini D, Kranz E, Werr W (2007) Wox gene phylogeny in Poaceae: A Comparative approach addressing leaf and embryo

development. *Mol. Biol. Evol.* **24**, 2474-2484.

Neale DB and Savolainen O (2004). Association genetics of complex traits in conifers. *Trends in plant science*, **9(7)**, 325-30.

Ogata K et al. (1996) The cavity in the hydrophobic core of Myb DNA- binding domain is reserved for DNA recognition and trans-activation. *Nat. Struct. Biol.* **3**, 178-187.

Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-Takagi M. (2001). Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *The Plant Cell* **13**, 1959-1968.

Palaisa KA, et al. (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell*, **15**, 1795-1806.

Parchman TL, Geist KS, Grahnen JA, Benkman CK, Buerkle CA (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics*, 11:180

Patzlaff A, McInnis S, Courtenay A, Surman C, Newman LJ, Smith C, Bevan MW, Mansfield S, Whetten RW, Sederoff RR, Campbell MM (2003) Characterisation of a pine MYB that regulates lignification. *Plant J*, **36(6)** 743-754.

Pauling L, Itano H, Singer SJ, Wells IC (1949) Sickle Cell Anemia, a Molecular Disease. *Science* **110** 543-548.

Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* 2000, **405**, 200-203.

Perron M, Perry DJ, Andalo C, Bousquet J (2000) Evidence from sequence-tagged site markers of recent progenitor derivative species pair in conifers. *PNAS*, **97**, 11331-11336.

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution and Systematics* **37**: 187-214.

Posada D, Crandall KA, Holmes EC (2002) Recombination in evolutionary Genomics. *Ann. Rev. Genetics*, **36**, 75-97

Posada D (2008). jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*, **25**,1253-1256.

Prunier J, Laroche J, Beaulieu J, and Bousquet J(2011) Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce, *Molecular ecology*, **20**, 1702-16.

Raffaele S et al. (2008) A MYB transcription factor regulates very- long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in Arabidopsis. *Plant Cell* **20**, 752–767.

Rambaut A (2006-2009) Tree figure Drawing Tool. “FigTree v.1.3.1” Institute of Evolutionary Biology, University of Edinburgh

Riechmann JL and Ratcliffe OJ (2000) A genomic perspective on plant transcription factors. *Current opinion in plant biology*, **3**, 423-34.

Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, et al (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* **290** 2105–2110.

Rigault P, Boyle B, Lepage P, Cooke JEK, Bousquet J, Mackay JJ (2011) A

white spruce gene catalog for conifer genome analyses. *Plant Physiology*, **157(1)**

Ronquist F, and Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**,1572-1574.

Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W et al.(2000) Comparative genomics of the eukaryotes. *Science*, **287**, 2204-2215.

Rubio, V. et al. (2001) A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. *Genes & development* **15**, 2122–2133.

Sanderson MJ, Doyle JJ (1992) Reconstruction of Organismal and Gene Phylogenies from Data on Multigene Families: Concerted Evolution, Homoplasy, and Confidence. *Systematic Biology*. **41(1)**, 4-17

Schultz RW, Tatineni VM, Hanley B, Thompson WF (2007) Genome-wide analysis of the core DNA replication machinery in higher plants Arabidopsis and rice, *Plant physiology*, **144**, 1697-1714.

Shen H, He X, Poovaiah CR, Wuddineh W, Ma J, Mann DGJ, Wang L, Jackson, Tang Y, Neal Stewart C, Chen F, and Dixon R (2011) Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks, *New Phytologist*,

Shepherd M, Williams CG (2008) Comparative mapping among subsection Australes (genus *Pinus*, family Pinaceae) *Genome* **51(5)** 320-331.

Simon M. et al. (2007) Distinct and overlapping roles of single-repeat MYB

genes in root epidennal patterning. *Dev. Biol.* **311**, 566–578.

Singh KB (1998) Transcriptional regulation in plants: the importance of combinatorial control. *Plant Physiol* 1998, **118**, 1111-1120

Smith HC, Sowden MP (1996) Base-modification mRNA editing through deamination--the good, the bad and the unregulated. *Trends Genet*,**12**,418-24

Stamatakis A, Hoover P, Rougemont J (2006) A rapid Booststrao Algorithm for the RAxML Web-Servers. *Systematic Biology*, **75(5)**, 758-771.

Stracke R, et al. (2007) Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant Journal*, **50**, 660–677.

Stracke R. et al. (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol*

Strauss BS (1971) Lipid Storage Diseases (edit. by Bernsohn, J., and Grossman, H. J.), **291** Academic, New Cork.

Thornton JW, DeSalle R (2000) Gene Family Evolution and Homology: Genomics Meets Phylogenetics. *Annu. Rev. Genomics. Hum. Genet.* **01**, 41-73.

Vázquez-Lobo A, Roujol D, Zuñiga-Sánchez E, Albenne C, Piñero D, Buen AG, FAMET E. (2012) The highly conserved spermatophyte cell wall DUF642 protein family: Phylogeny and first evidence of interaction with cell wall polysaccharides in vitro. *Mol Phylogenet Evol.* **63(2)**, 510-520.

Vendrami GG, Degen B, Petit RJ, Anzidei M, Madaghile A, Ziegenhagen B (1999) High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. *Molecular Ecology.* **8(7)**, 1117-1126.

Voytas DF, M. P. Cummings MP, Konieczny A, Asubel FM and Rodermeel S (1992): *Copia*-like retrotransposons are ubiquitous among plants. *Proc. Natl. Acad.Sci*, **89** 7124–7128.

Wang XQ, Tank DC, and Sang T (2000) Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular biology and evolution*, **17**, 773-81.

Waters, M.T. et al. (2009) GLK transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*. *Plant Cell* **21**, 1109–1128

Weigel RR, Pfitzner UM and Gatz C (2005) Interaction of NIMIN1 with NPR1 modulates PR gene expression in *Arabidopsis*. *Plant Cell* **17**: 1279–1291

Wheeler WC and Giribet G (2009). Phylogenetic hypotheses and the utility of multiple sequence alignment pp 95-104, in M. Rosenberg eds., *Sequence Alignment: Methods, Models, Concepts, and Strategies*. University of California Press: Berkeley, CA.

Wilkins O, Nahal H, Foong J, Provart NJ, and Campbell MM (2009), Expansion and Diversification of the *Populus* R2R3-MYB Family of Transcription Factors, *Society*, 981-993.

Wolff RL, Lavialle O, Pédrone F, Pasquier E, Destailats F, Marpeau AM, Angers P, Aitzetmüller K (2002) Abietoid seed fatty acid compositions a review of the genera *Abies*, *Cedrus*, *Hesperopeuce*, *Keteleeria*, *Pseudolarix*, and *Tsuga* and preliminary inferences on the taxonomy of Pinaceae. *Lipids*. Jan; **37(1)**:17-26.

Xiang X, Cao M, Zhou Z (2006) Fossil History and modern distribution of the genus *Abies* (Pinaceae). *Frontiers of forestry in China*, **2 (4)**, 335-365.

Xu Z, Wang H (2007) LTR_FINDER: an efficient tool for prediction of full-length

LTR retrotransposons. *Nucleic Acids Res.*

Xuan Z, McCombie WR, Zhang MQ (2002) GFScan: A Gene Family Search Tool at Genomic DNA level. *Genome Research*, **12**, 1142-1149

Yang SW et al. (2009) FAR-RED ELONGATED HYPOCOTYL1 and FHY1-LIKE associate with the Arabidopsis transcription factors LAF1 and HFR1 to transmit phytochrome A signals for inhibition of hypocotyl elongation. *Plant Cell* **21**, 1341–1359.

Yoo JH, Park CY, Kim J.C, Heo WD, Cheong MS, et al. (2005) Direct interaction of a divergent CaM isoform and the transcription factor, MYB2, enhances salt tolerance in Arabidopsis. *The Journal of biological chemistry*, **280**, 3697-3706.

Zhang F y Peterson T (2005) Comparisons of Maize pericarp color1 Alleles Reveal Paralogous Gene Recombination and an Organ-Specific Enhancer Region. *The Plant Cell*. **17**, 903-91

Apéndices

Apéndice1: Tabla de las secuencias utilizadas en la reconstrucción filogenética y otros análisis de esta Tesis.

	<i>Nombre</i>	<i>Especie</i>	<i>Origen</i>	<i>Tipo de secuencia</i>	<i>Longitud Original (pb)</i>	<i>Clasificación</i>	<i>ID GenBank</i>
1	ArMYBI	<i>Abies religiosa</i>	Amplificación y clonación	DNA	508	N/A	
2	ArMYBII	<i>Abies religiosa</i>	Amplificación y clonación	DNA	486	N/A	
3	ArMYBIII	<i>Abies religiosa</i>	Amplificación y clonación	DNA	768	N/A	
4	ArMYBIV	<i>Abies religiosa</i>	Amplificación y clonación	DNA	493	N/A	
5	ArMYBV	<i>Abies religiosa</i>	Amplificación y clonación	DNA	472	N/A	
6	ArMYBVI	<i>Abies religiosa</i>	Amplificación y clonación	DNA	504	N/A	
7	ArMYBVII	<i>Abies religiosa</i>	Amplificación y clonación	DNA	493	N/A	
8	ArMYBVIII	<i>Abies religiosa</i>	Amplificación y clonación	DNA	489	N/A	
9	ArMYBIX	<i>Abies religiosa</i>	Amplificación y clonación	DNA	504	N/A	
10	ArMYBX	<i>Abies religiosa</i>	Amplificación y clonación	DNA	541	N/A	
11	ArMYBXI	<i>Abies religiosa</i>	Amplificación y clonación	DNA	1015	N/A	*secuencia doble atribuida a error de sec.
12	ArMYBXII	<i>Abies religiosa</i>	Amplificación y clonación	DNA	486	N/A	
13	ArMYBXIII	<i>Abies religiosa</i>	Amplificación y clonación	DNA	463	N/A	
14	ArMYBXIV	<i>Abies religiosa</i>	Amplificación y clonación	DNA	504	N/A	
15	ArMYBXV	<i>Abies religiosa</i>	Amplificación y clonación	DNA	484	N/A	
16	AtMYB4	<i>Arabidopsis thaliana</i>	GenBank (busqueda directa)	DNA (anotado)	1287	Sg 4	NM_120023.2 GI: 3069
17	AtMYB13	<i>Arabidopsis thaliana</i>	GenBank (busqueda directa)	DNA (anotado)	912	Sg 2	NM_100499.1 GI: 1839
18	AtMYB20	<i>Arabidopsis thaliana</i>	GenBank (busqueda directa)	DNA (anotado)	1186	Sg 8	NM_105294.2 GI: 3069
19	AtMYB32	<i>Arabidopsis thaliana</i>	GenBank (busqueda directa)	DNA (anotado)	944	Sg 4	NM_119665.2 GI: 3069
20	AtMYB106	<i>Arabidopsis thaliana</i>	GenBank (busqueda directa)	DNA (anotado)	1543	Sg 9	NM_110979.2 GI: 7938
21	PsB_ArMYBI	<i>Picea sitchensis</i>	GenBank (Blastn vs secuencia obtenida previamente)	DNA (No anotado)	N/D	N/A	N/D
22	PsB_ArMYBV	<i>Picea sitchensis</i>	GenBank (Blastn vs secuencia obtenida previamente)	DNA (No anotado)	N/D	N/A	N/D
23	PsB_ArMYBXI	<i>Picea sitchensis</i>	GenBank (Blastn vs secuencia obtenida previamente)	DNA (No anotado)	N/D	N/A	N/D
24	PgMYB1	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds completo	1480	Sg 8	EF601064.1 GI: 147744
25	PgMYB5	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds completo	1227	Sg 4	EF601068.1 GI: 147744
26	PgMYB10	<i>Picea glauca</i>	GenBank (busqueda directa)	DNA (anotado)	757	Sg 4	DQ399064.2 GI: 14583
27	PgMYB11	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds completo	2307	Sg 9	EF601074.1 GI: 147744
28	PgMYB12	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds completo	2113	Sg 2	EF601075.1 GI: 147744

29	PgMYB13	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds completo	756	Sg 4	EF601076.1 GI: 14774
30	PgMYB14	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds parcial	895	Sg 4	FJ469917.1 GI: 223674
31	PgMYB15	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds parcial	870	Sg 4	FJ469918.1 GI: 223674
32	PgMYB16	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds parcial	1026	Sg 4	FJ469922.1 GI: 223674
33	PgMYB17	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds completo	918	Sg 4	FJ469923.1 GI: 223674
34	PgMYB18	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds parcial	1109	Sg 4	FJ469919.1 GI: 223674
35	PgMYB19	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds parcial	1057	Sg 4	FJ469920.1 GI: 223674
36	PgMYB20	<i>Picea glauca</i>	GenBank (busqueda directa)	mRNA, cds parcial	1407	Sg 4	FJ469921.1 GI: 223674
37	PgB_ArMYBI	<i>Picea glauca</i>	GenBank (Blastn vs secuencia obtenida previamente)	DNA (No anotado)	N/D	N/A	N/D
38	PgB_ArMYBVII	<i>Picea glauca</i>	GenBank (Blastn vs secuencia obtenida previamente)	DNA (No anotado)	N/D	N/A	N/D
39	PgB_ArMYBVI	<i>Picea glauca</i>	GenBank (Blastn vs secuencia obtenida previamente)	DNA (No anotado)	N/D	N/A	N/D
40	PgB_PtMYB22	<i>Picea glauca</i>	GenBank (Blastn vs secuencia obtenida previamente)	DNA (No anotado)	N/D	N/A	N/D
41	PtMYB1	<i>Pinus taeda</i>	GenBank (busqueda directa)	mRNA, cds completo	1023	Sg 8	AY356372.1 GI: 16882
42	PtMYB5	<i>Pinus taeda</i>	GenBank (busqueda directa)	mRNA, cds parcial	1029	Sg 4	FJ469924.1 GI: 223674
43	PtMYB10	<i>Pinus taeda</i>	GenBank (busqueda directa)	mRNA, cds parcial	1024	Sg 4	FJ469925.1 GI: 223674
44	PtMYB13	<i>Pinus taeda</i>	GenBank (busqueda directa)	mRNA, cds parcial	855	Sg 4	FJ469926.1 GI: 223674
45	PtMYB14	<i>Pinus taeda</i>	GenBank (busqueda directa)	mRNA, cds parcial	967	Sg 4	EU482901.1 GI:16882
46	PtMYB16	<i>Pinus taeda</i>	GenBank (busqueda directa)	mRNA, cds parcial	961	Sg 4	FJ469927.1 GI: 223674
47	PtMYB21	<i>Pinus taeda</i>	GenBank (busqueda directa)	mRNA, cds parcial	995	Sg 4	FJ469928.1 GI: 223674
48	PtMYB22	<i>Pinus taeda</i>	GenBank (busqueda directa)	mRNA, cds parcial	977	N/D	FJ469929.1 GI: 223674

Apéndice 2: Secuencias completas de Nucleótidos de *Abies religiosa* obtenidas experimentalmente.

ArMYBI

GGCTTCCAAGTAGTTCTTTATCTCGTTGTCGGTCCGCCCTGGCAATCTCCCTGCAATAAGAGACCACCTGCGCAGTTTCAAGGACTAATTAATATCA
GTTCCGGTCCATAATTGAAACATCAAATATGAGTTCGGCCCATAATTGAAACCCTCTCGGTCCATGACAGATTAGGGTTTAGTAAAAATATGAGTCCT
GATAAGTAAGGGAATTGATTGAAGCAATAGAGGCATTGAAATAAAAAAAGCATACTTGTTGCCGAGAAGGTAATGGAGTTTGATAATGAGATCATCT
TCTTCTTCAAAGAAGTCTCCACGCTTGAGATCGGGACGCAGGTAGTTCACCCATCGGAGCCTGCAGCTCTTCCGCATCTCAGCAGCCCTGCAGT
CGAAAAATCCAGAAAACGTTCAATATATTGTCATAAATCAATTTACAAGCAGCTATTGATTGAAATGAAAAAGCAGAGGGTTGGCATGAAACCTGCG
GCCTTGGGAAGCGAACGCCAAAGC

ArMYBII

GGCTTTGGCGCTCCCTTCCCAAGGCCGCGAGGTTAACTTAACTTAACTACAGTTATTAAGTTATTTCTATTCTGTTCCACTACATTATATGCACAGTT
TCCTTGCATTTTTTAGTTTGATTTTAACGATTTTCTTCGTCGTTGGATGCTAAACAAGTGGTTTCGATTTTATAGGTCTCCTTCGTTGTGGCAAAGTTG
CAGGCTGAGATGGATAAATTATCTTCGCCCTGATCTCAAACGAGGGAATTTTCGTGAAGACGAAGACGATCTTATTCTCAAAGTGCACGCCCTCCTT
GGAAATAGGTAGGTTTGAAGAATATAGTCTGCCTTGTGTTAGTCCCTTTCGGCTAATTTTTAATGTTAATACATTGAAATTGAAGGTAATTATTGTTTCTT
TTTGTCTTTCGATTTTCACTTTCAGGTGGTCTCTGATAGCTGGCCGTTTGCCAGGACGAACGGACAACGAGATAAAGAAGTACTGGAAGC

ArMYBIII

GGCTTTGGCGTTCGCTTCCCAAGGCCGCGAGGTTGCTTGCCTGAACTCCTCCGCCCTTCTAAATACATGCTTCTGAAATCGAGAGATTTTATATATAG
ACTTTTCTTGATATTCTGACGCAGCAAAAATTTTCTGGATGTTCTTGATTTTCTGACGCAGCAAAAATTTTCTTGATGTTCTTGATTTTCTGAGGCAGTAA
ATTTTTCTTGATGTTCTTGATTTTCTGACGCAGCAAAAATTTTCTCGATGTTCTTGATTTTCTGAGGCAGCAAAAATTTTCTTGATATTCTTGATTTTCTGA
GGCAGCAAAAATTTTCTTGATATTCTTGATTTTCTGAGGCAGTAAAATTTTCTTGATGTTCTTGATTTTCTGACGCAGCAAAAATTTTCTCGATGTTCTTG
TTTTCTGACGCTGTAAGATTCTTGATGTTCTTGATTTTCTGAGGCAGTAAAATTTTCTTGATGTTCTTGATTTTCTGACGCAGTACGCTTTTCTTG
TTTTCTGATGCAGGGTTGCTACGATGCGGGAAGAGCTGCAGGCTGCGATGGATAAACTACCTGCGCCCCGATCTCAAGCGTGGAAACTTCTCTGA
AGAAGAGGATGAGCTCATCATCAAACCTCCACTCCCTCCTCGGCAACAAGTATGTTATTTTATCGCCGCTATATATTGCTCCAATCAAATCCCTTACTC
ATCAGGCCTCATATTTATATTAAGTTCTATCTGTTAAGGGCCGACGGGTTTCAAATTTATTGAAATGACAAATCA

ArMYBIV

GGCTTCCAGTAGTTCTTTATCTCTTTTGTCCGTTCCGCCAGGCAATCTCCCTGCAATAAGCGACCACCTGCACAGTTGAACAGACTAATCCATAAAT
GAAACTTCAAATTTGAGTTTTTCTGATAGATCCAGGTTTAAATTTGGGGATTAGTAAGGGAATTTTGATTGGAGCCGCACAGATTCAATGAAATAAC
ATACTTGTGCGGAGGAGGGCGTGCAGTTTGTGATGAGATCGTCTTCTTTCAGAGAAAATTTCCCGCTTGAGATCTGGACGCAGGTAGTTTATC
CATCGGAGCTTGCAGCTCTTCCCGCATCGCTGCAACCCTGTGCTCCAAAAATAAAGAAAACTTTATTTAATCAAGAAAAGCTTTAAATATAATCAAGA

AAAGCTTTAAATATAATCTCTTGAATCCAGAAGCGGAAATTAATTATATTCAAGAGCAGAGCTCGGGCGAGAAACCTGCGGCCTTGGGAAGCGAGC
GCCAAAGC

ArMYBV

AGGCCGCAGGTTTCTTGCCCGAGCTCTGCTCTTGAATATAATTAATTTCCGCTTCTGAATTCAAGAGATTATATTTAAAGCTTTTCTTGATTATATTTAA
AGCTTTTCTTGATTAAATAAAGTTTTCTTTATTTTTGGAGCACAGGGTTGCAGCGATGCGGGAAGAGCTGCAGGCTCCGATGGATAAACTACCTGCG
TCCAGATCTCAAGCGGGGAAATTTCTCTGAAGAAGAAGACGATCTCATCATCAAAGTGCACGCCCTCCTCGGCAACAAGTATGTTATTTTCATTGAAT
CTGTGCGGCTCCAATCAAATTCCTTACTAATCCCCAAATATTAACCTGGATCTATCAAAAAACTCAATTTTGAAGTTTCATTTATGGATTAGTCT
GTTCAACTGTGCAGGTGGTCGCTTATTGCAGGGAGATTGCCTGGGCGAACGGACAACGAGATAAAGAAGTACTGGAAGC

ArMYBVI

GGCTCCAGTAGTTCTTTATCTCGTTGTCAGTGCGCCCTGCCAATCTCCCTGCAATAAGAGACCACCTGCACATTTTTACAGCTGATACATTAGATC
TCTGTATTCTTTTCATGACAGATCCCGTCATGAAACTGCAAAATTGAGATTCTTTTCATGACAGATCCCGTCCGGATTGAGTCCATAAGTAAAAGTAACTGCAA
AATTGAGATTCTTTTCATGACAGATCCCGTCCGCTAAAGCGAGGAAATTGCATACTTGTTACCAACGAAGGAATGGAGTTTGGATGATGAGTTCGTCTT
CTTCTTCTGAGAAATTCACGCTTTCAGATCGGGACGCAGGTAGTTTATCCATCGGAGCCTGCAGCTTTTTCCGCATCGCAACAGCCCTGCATCAG
AAAATCAAAAATAAGCAACGTTTAAACATAATCTCTTAAATTCAAAGTCACTGATATCGAAAACAGAGAGCGGGCAAGAAACCTGCGGCCTT
GGGAAGCGAACGCCAAAGC

ArMYBVII

GGCTTTGGCGTTCGCTTCCCAAGGCCGCAGGTTTCTTGCCCGAGCTCTGCTCTTGAATATAATTAATTTCCGCTTCTGAATTCAAGAGATTATATTTA
AAGCTTTTCTTGATTATATTTAAAGCTTTTCTTGATTAAATAAAGTTTTCTTTATTTTTGGAGCACAGGGTTGCAGCGATGCGGGAAGAGCTGCAGGCT
CCGATGGATAAACTACCTGCGTCCAGATCTCAAGCGGGGAAATTTCTCTGAAGAAGAAGACGATCTCATCATCAAAGTGCACGCCCTCCTCGGCA
ACAAGTATGTTATTTTCATTGAATCTGTGCGGCTCCAATCAAATTCCTTACTAATCCCCAAATATTAACCTGGATCTATCAGAAAAACTCAATTTTG
AAGTTTCATTTATGGATTAGTCTGTTCAACTGTGCAGGTGGTCGCTTATTGCAGGGAGATTGCCTGGGCGAACGGACAACGAGATAAAGAAGTACT
GGAAGC

ArMYBVIII

GGCTTTGGCGCTCCCTTCCCAAGGCCGCAGGTTTCTTGCCCTGAACTCTGCTCTTCTTCAATGGAATTAATAACTGATTGTGAATTGAAGAGATTATAT
TTAAACTTTTCTTGATTTTCTGATACAGGGCTGCTGCGATGCGGTAAGAGCTGCAGGCTACGATGGATAAACTACCTGCGCCCCGATCTTAAGCGC
GGGAATTTCTCAGAAGAAGAAGACGAGCTCATCATCAAAGTCCACTCCCTTCTCGGCAATAAGTATGTAATTTTCATCGCTTCGATTCCGTCAATCAA
TTCCCTTGCCCTATCAGGACTGATGTTTATATGAAAATGGATCCTTAGAGGTTTTTTAACTGCCAATCCCAATTCTGGGAAGAAGTCTTCAATAGTGAAC
CAAAGTATATTTATTTATCTGTGAAACTGCGTAGGTGGTCTTATTGCAGGGAGATTGCCCGGGCGGACGGACAACGAGATAAAGAAGTACTGGAAGC

ArMYBIX

GGCTTTGGCGCTCGCTTCCCAAGGCCGCAGGTTTCTTGCCCGAGCTCTGATCTTCCATATATAATTTCTGCTTCTGAATTCAAGAGATTATATTTA

AAGCTTTTCTTGATTTTAATTTCTGCTTCTGAATTCAAGAGATTATATTTAATTTCTTGATTTTCGGACTGCAGGGCTGCTGCGATGCGGGAAGAGCTG
CAGGCTGCGATGGATAAACTACCTGCGTCCCGATATCAAGCGTGGAACTTCTCAGAAGAAGAAGACGATCTCATCAAACCTCCATCCCTCCT
CGGCAACAAGTATGTTATTTCAATTTATTCTTTTCTTCAATGAATTCCTGACTTAACCCTACATACTAAGCCTAGATCTGTGCGTAAAACCTTGATCC
TGTGTTTTCGCTTAAGCTGATCTTAATTTGTCTGTTAACTGTGCAGGTGGTCCCTTATTGCAGGGAGATTGCCCGGGCGGACGGACAACGAGATA
AAGAACTACTGGAAGC

ArMYBX

GGCTTTGGCGCTCGCTTCCCAAGGCCGCGAGGTTTCTTGCCTGCGCTATGCTTTTTCTTTCTTTTCTTTTCACTGGAATTGATAGCTTCTTGTAAT
GTAAATTGATTGATTTCAAGAAATTGTGTTAACGTTTTCTTTGTTTTTTGTAAATTGATTGATTTCAAGAAATTGTGTTAAAGTTTTCTTGATTTTTCT
GCCCGAGGGCTGCTGAGATGCGGAAAGAGCTGCAGGCTCCGTTGGATAAACTACCTGCGTCCCGATCTTAAGCGTGGAAAGCTTACAGAAGAAG
AAGACGAGCTCATCAAACCTCACTCCGTCGTCGGCAACAAGTATATAATTATATCCCTCTGCAATGGATTCCCTTATTTCAACGGACATATTA
ACCTAGATCTGTCATGGAATTTTTTTTTATAAGTTTTAGTTATGGACCGAGCTGATATTAATCATTCTGTGAACTGTGCAGGTGGTCTCTTATTGC
AGGAAGATTGCCAGGGCGGACGGACAACGAGATAAAGAACTACTGGAAGC

ArMYBXI

GGCTTCCAGTAGTTCTTTATCTCGTTGTCCGTCCGCCCTGGCAATCTCCCTGCAATAAGAGACCACCTGCGCAGTTTCAAGGACTAATTAATATCAG
TTCGGTCCATAATTGAAACATCAAATATGAGTTCGGCCCATTAATTGAAACCCTCTCGGTCCATGACAGATTAGGGTTTAGTAAAAATATGAGTCCTGA
TAAGTAAGGGAATTGATTGAAGCAATAGAGGCATTGAAATAAAAAAAGCATACTTGTGCCGAGAAGGTAATGGAGTTTGATAATGAGATCATCTC
TTCTTCAGAGAAGTCTCCACGCTTGAGATCGGGACGCGAGGTAGTTCACCCATCGGAGCCTGCAGCTCTTCCGCATCTCAGCAGCCCTGCAGTCG
AAAAATCCAGAAAACGTTCAATATATTGTCATAAATCAATTTACAAGCAGCTACTGATTGAAATGAAAAAGCAGAGGGTTGGCATGAAACCTGCGGC
CTTGGGAAGGGAACGCCAAAG

ArMYBXII

GGCTTTGGCGTTCGCTTCCCAAGGCCGCGAGGTTAACTTAACTTAACTACAGTTATTAACTTATTTCTATTCTGTTCCACTACATTATATGCACAGTT
TCCTTGCATTTTTTAGTTTGATTTAACGATTTTCTTCGTCGTTGGATGCTAAACAAGTGGTTTCGATTTTATAGGTCTCCTTCGTTGTGGCAAAGGTTG
CAGGCTGAGATGGATAAATTATCTTCGCCCTGATCTCAAACGAGGGAATTTTTGTGAAGGCGAAGACGATCTTATTCTCAAACCTGCACGCCCTCCT
GGAAATAGGTAGGTTTGAAGAATATAGTCTGCCTGTTTAGTCCCTTTCGGCTAATTTTTAATGTTAATACATTGAAATTGAAGGTAATTATTGTTTCT
TTTGTCTTTCGATTTTCACTTTCAGGTGGTCTCTGATAGCTGGCCGTTTGCCAGGACGAACGGACAACGAGATAAAGAACTACTGGAAGC

ArMYBXIII

GGCTTCCAGTAGTTCTTTATCTCGTTGTCCGTTTCGACCTGGCAAACGACCAGCTATCAGAGACCACCTGCAAGTGAAAAAGAAACAATGATTTCGCT
TCATTTCCCATATATTAGCTCGAAAAGTTCACGGAAAGGGACTAAACAAGTGGGAATAATTCTTCAAACGTAATTTTCCAAGGAGAGCGTGCAGT
CTGAGAATAAGATCGTCTTCATCTTCAGAAAAATTCAGTCTTTGAGATGAGGACGAAGATAATTTGTCCATCTCAGCCTGCAACTTTTGTACGCC
GAAGCAGACCTGGTATTATCACATGAAAAACAGTTTCTTAGCATCCAACCTCTAGCAACATCCTTACTATTAATAAGAGATCGAAGGAAGCTGTC
GATATATGTAGTACGGTAGAGCAGAAATATAGAATAATGTGTCAAACCTGCGGCCTTGGGAAGCGAGCGCCAAAGC

ArMYBXIV

GGCTTCCAGTAGTTCTTTATCTCGTTGTCCGTCCGCCCGGGCAATCTCCCTGCAATAAGGGACCACCTGCACAGTTTAACAGACAAATTAAGATCA
GCTTAAGCGAAAACACAGGATCAAGGTTTTACGACAGATCTAGGCTTAGTATGTAGGGTTAAGTCGGGGAATTCATTGAAGGAAAAGAATAAATGA
AATAACATACTTGTTGCCGAGGAGGGAATGGAGTTTGATGACGAGATCGTCTTCTTCTTCTGAGAAGTTTCCACGCTTGATATCGGGACGCAGGTA
GTTTATCCATCGCAGCCTGCAGCTCTTCCCGCATCGCAGCAGCCCTGCAGTCCGAAAATCAAGAAATTAATATAATCTCTTGAATTCAGAAGCAG
AAATTAATCAAGAAAAGCTTTAAATATAATCTCTTGAATTCAGAAGCAGAAATTATATATATGGAAGATCAGAGCTCGGGCAAGAAACCTGCGGCC
TTGGGAAGCGAGCGCCAAAGC

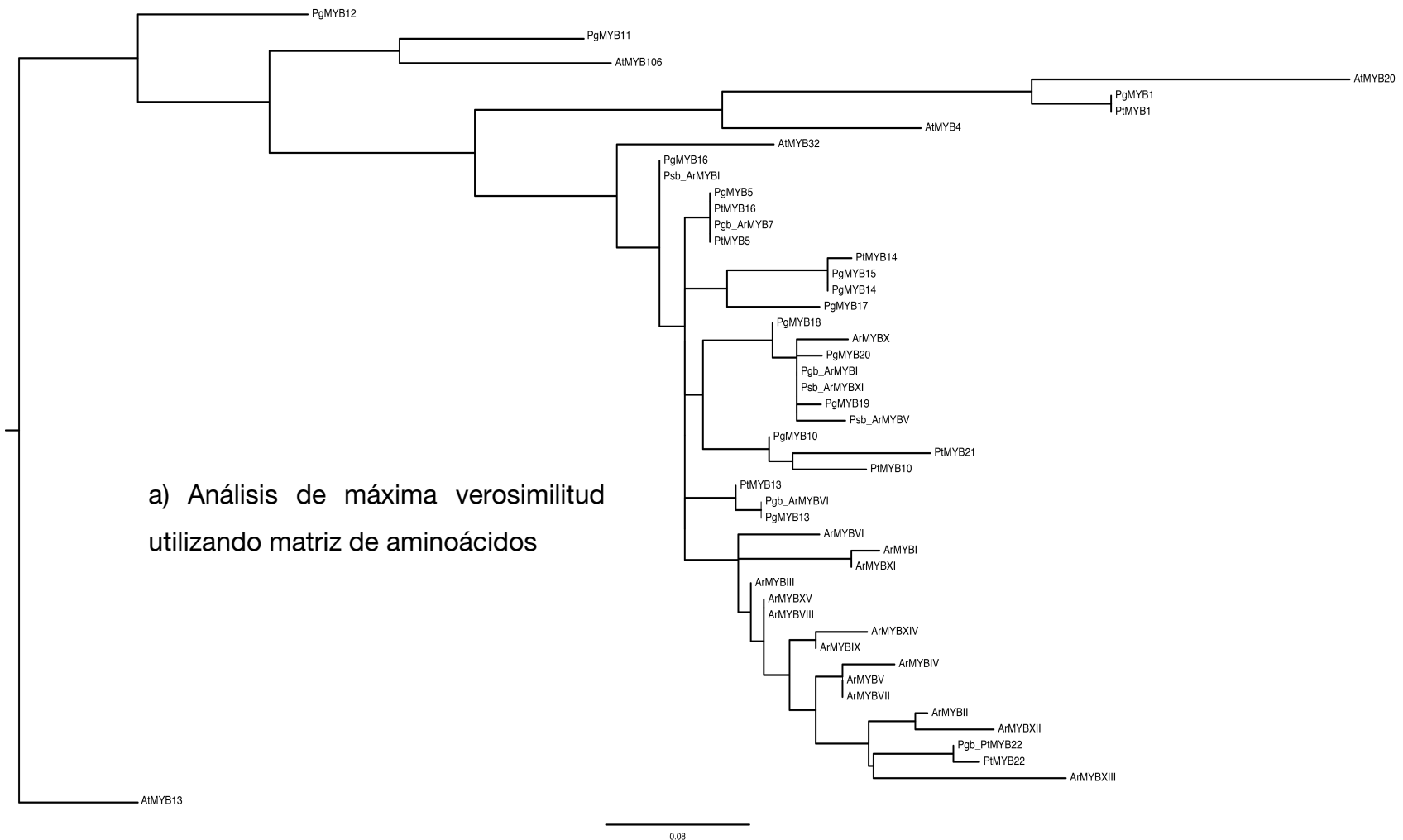
ArMYBXV

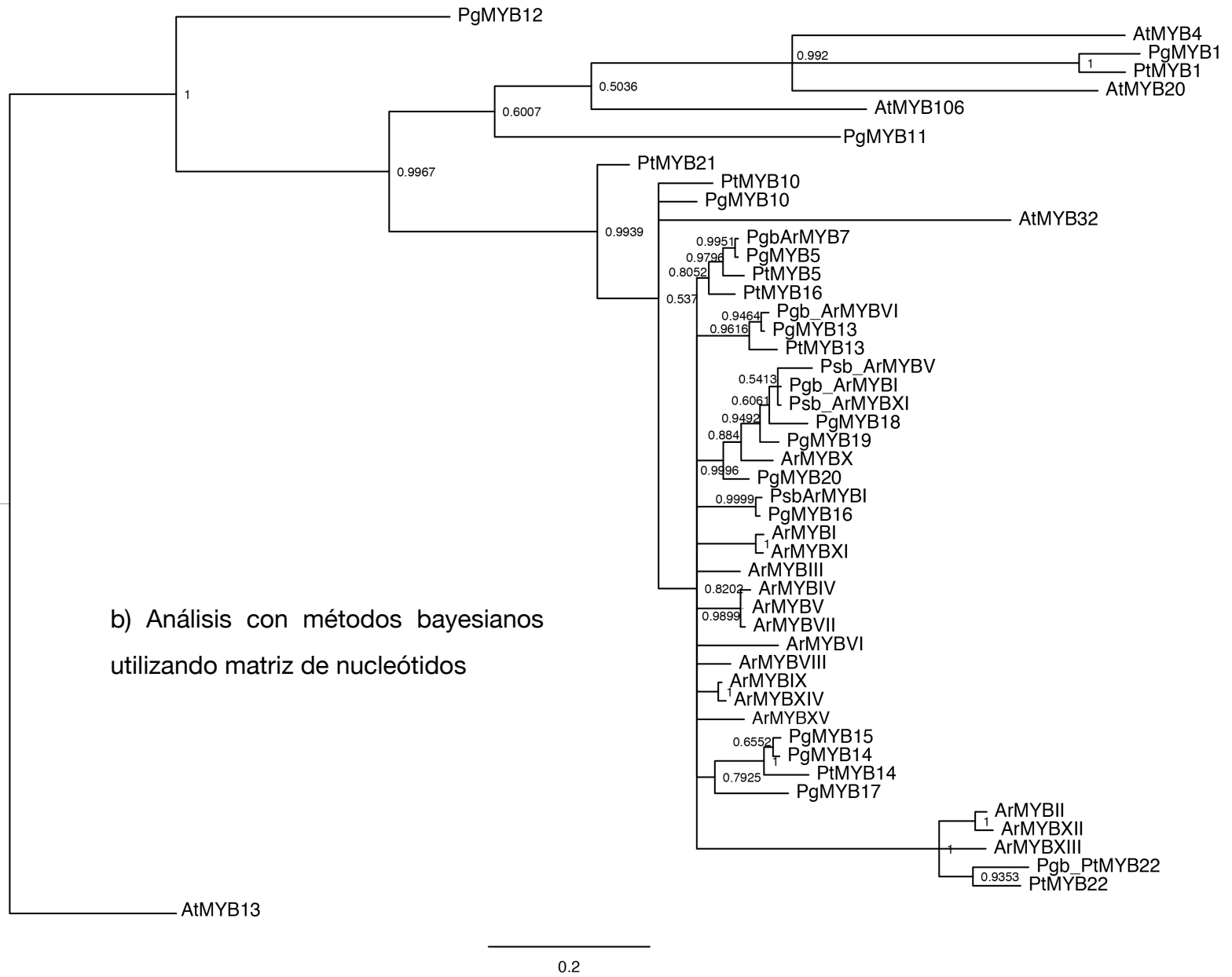
GGCTTTGGCGTTCGCTTCCCAAGGCCGCAGGTTTCTTGCTCGATCTTTGCTCTTCGATGAAATTAATAACTGTTTCTGAAGTAAAGAGATTAATATA
TAACGTTTGCTTGATTTTACGATCGCAGGGCTGCTGAGATGCGGAAAGAGCTGCAGACTCCGATGGATTAACCTGCGTCCAGATTTAAAGCGT
GGAAATTTCTCAGAAGAAGAAGACGAGCTCATCATCAAACCTCCACTCCCTTCTCGGCAATAAGTATGCAATTTTCATTGGCTTTATTGCTTAAATCAAT
TTCCTTACTGATATTTATATTAACCCTAACCTGTCACGGACCGAGAGGTTTCTTGTCTAAATCTTGAGCGGAACTCTCAATTGTGGATCGAACTGA
TATGAATTAATTACTGTGAAACTGTGCAGGTGGTCTTATTGCAGGGAGATTGCCCGGGCGAACGGACAACGAGATAAAGA ACTACTGGAAGC

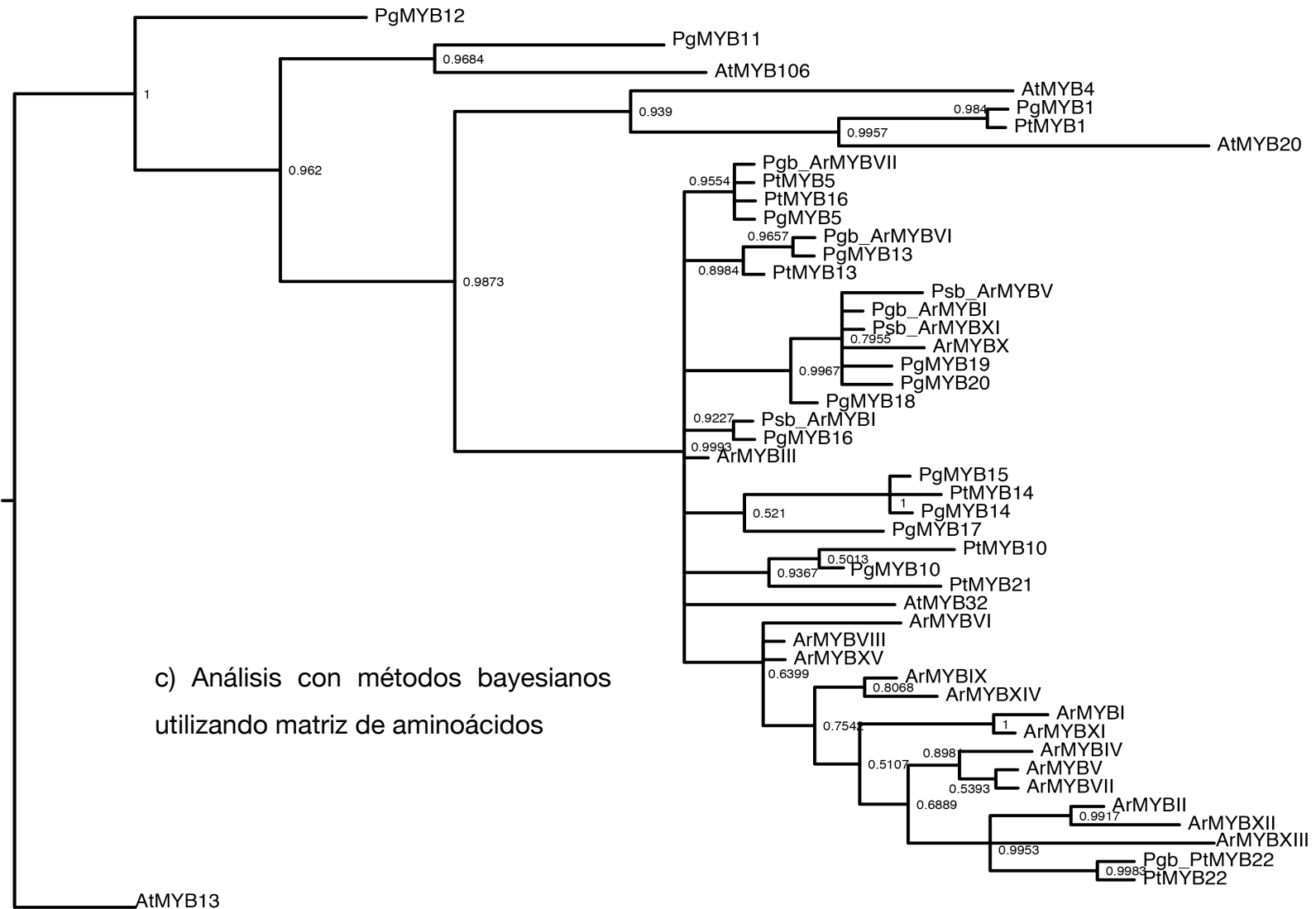
Apéndice 3: Tabla de localización de posiciones y regiones génicas (Intrones-Exones) en pb de secuencias de *Abies religiosa*. No se reporta Inicio de Exón 1, porque en todos los casos empieza con el nucleótido 1, ni tampoco el final de exón 3, ya que este llega hasta el final de cada secuencia (Con excepción de ArMYBIII). Tampoco se reporta la longitud de estas regiones ya que pueden abarcar mas allá de lo secuenciado.

	<i>inicio Intrón 1</i>	<i>Inicio Exón 2</i>	<i>Inicio Intrón 2</i>	<i>Inicio Exon 3</i>	<i>longitud I.1</i>	<i>Longitud E.2</i>	<i>Longitud I.2</i>
ArMYBI	30	128	256	440	98	132	184
ArMYBII	30	179	306	438	149	132	132
ArMYBIII	30	511	641	N/D	481	132	¿?
ArMYBIV	30	166	297	429	136	132	132
ArMYBV	12	150	280	412	138	132	132
ArMYBVI	30	125	257	443	95	132	186
ArMYBVII	30	165	298	430	135	132	132
ArMYBVIII	30	127	258	426	97	132	168
ArMYBIX	30	172	303	478	142	132	175
ArMYBX	30	207	339	478	177	132	139
ArMYBXI	30	127	262	443	97	132	181
ArMYBXII	30	174	306	423	144	132	117
ArMYBXIII	30	164	296	400	134	132	104
ArMYBXIV	30	171	302	439	141	132	137
ArMYBXV	30	127	258	421	97	132	163

Apéndice 5: Reconstrucciones filogenéticas adicionales de genes R2R3-MYB, estas topologías complementan y aportan información al árbol expuesto en la figura 3 en la sección de resultados. Clave: At=*Arabidopsis thaliana*, Pt=*Pinus taeda*, Pg=*Picea glauca*, Ps=*Picea sitchensis*, Ar=*Abies religiosa*, Psg y Pgb, genes MYB de *Picea sitchensis* y *Picea glauca* respectivamente encontrados en una búsqueda con BLAST,







c) Análisis con métodos bayesianos
utilizando matriz de aminoácidos

0.2

