



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

Facultad de Estudios Superiores Iztacala

**Genómica de poblaciones de *Juniperus blancoi*
y *J. monticola*: diversidad genética y variación
altitudinal.**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGA

P R E S E N T A:

VERÓNICA REYES GALINDO

DIRECTORA DE TESIS:

DRA. ALICIA MASTRETTA YANES



Los Reyes Iztacala, Tlalnepanitla, Estado de México, 2016.



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.

Esta tesis fue realizada en el Laboratorio de Genética y Ecología del Departamento de Genética y Evolución en el Instituto de Ecología de la UNAM, bajo la dirección de la Dra. Alicia Mastretta Yanes.

Este proyecto tuvo el apoyo de una beca por parte del Programa de Ayudantes de Investigador Nacional Nivel III, CONACYT. Mediante el Dr. Daniel Ignacio Piñero Dalmau.

Referirse al escrito para más detalles del Proyecto.

*Cada vida le ofrece su cosecha
y se extingue después.
Cada alma va hacia él como una flecha,
y en su gran alma chispa nueva es.*

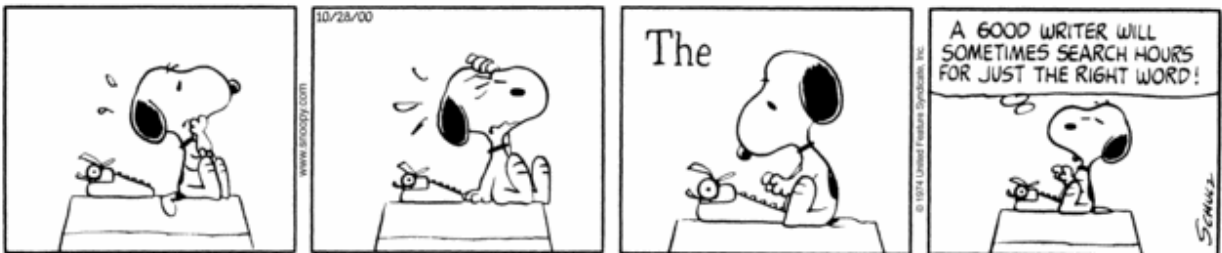
*Cada tránsito en él es enseñanza;
cada humana aflicción,
un ala nueva para su esperanza
de perfección.*

*Él la clave posee de tu estado;
él ha pesado
cada desliz;
él comprende por qué eres desgraciado,
¡por qué fuiste feliz!*

*Es el dueño y señor por quien laboras;
es tu conciencia; mas con vastedad
vertiginosa, él sabe cuanto ignoras,
¡y lleva en sí tu eternidad!*

*Él vela cuando duermes, y en tu mente
es un genial relámpago, un tropel
de rimas, trémulo y resplandeciente.
Tú pasas sí: mas él es permanente;
tú mudas, sí: mas él es siempre fiel.
Sólo vives para tu subconsciente,
y mueres sólo para él.*

El subconsciente, Amado Nervo



Dedicatoria

A mi familia genética:

Laura, Roberto y Dhëni

Por ser un ejemplo en cada momento de mi vida y a quien debo gran parte de lo que soy. Su esfuerzo, pasión comprensión y libertad brindada siempre me acompaña.

Somos la progenie de la historia, y debemos establecer nuestros propios caminos en el más diverso e interesante de los universos concebibles: un universo indiferente a nuestro sufrimiento y que, por lo tanto, nos ofrece la máxima libertad para prosperar, o para fracasar, de la manera que nosotros mismos elijamos.

Stephen J. Gould

AGRADECIMIENTOS

*Nosotros, que no más somos fantasmas,
queremos perdurar en la memoria
de otro fantasma: la posteridad,
que ha de surgir mañana de la sombra
en que nosotros nos desvanecemos,
y se irá sin remedio algunas horas
después, arrebatada por el viento...
¡Y a perdurar así llamamos gloria!*

Fragmento Inmortalidad, Amado Neruo.

Este instante de mi vida llamado “Tesis de Licenciatura” ha sido de los más contrastantes, lleno de alegría, éxito, tedio y a veces tristeza. Mediante todas las travesías emocionales en este tiempo, pude reafirmar mi camino científico. Este camino me permitió fortalecer lazos con personas que guardan una impresionante luz que los guía y a quienes agradezco me brindaran algo que no a cualquiera está dispuesto a dar: tiempo y cariño.

Primeramente quisiera agradecer a mi mamá Laura, quien siempre ha sido el soporte de la familia, enseñándome a mantener la fuerza y convicciones en todo momento. Te agradezco la fe y confianza depositada en cada meta que me propuse. A mi hermano Beto, alma misteriosa que desde los primeros días de mi existencia me motiva a “Intentar lo imposible” y a quien debo la sensibilidad que sólo un artista puede mostrar. A Dhëni, florecilla de esperanza quien siempre me recuerda la felicidad, asombro y magia que guardan los pequeños detalles que nos rodean.

Durante este camino gran parte lo pase con amigos, sin embargo podría decir que los amigos son la familia que uno escoge tener al lado sin importar las fluctuaciones emocionales y las diferencias genéticas. Memow, Juanito, Olivino, Ferchis e Iris, ustedes se han convertido en amigos y cómplices con los que he madurado a lo largo de la carrera. Amigos con los que compruebo día a día que las diferencias pueden ser formas de complementarnos para salir adelante a pesar del tiempo y la distancia. Deseo que sigan pasando muchos años más de amistad y éxitos compartidos.

A través de este camino es muy fácil perderse. Llegó un momento de indecisión, donde dormía soñando encontrar algún sentido. Un día de Agosto desperté y Alicia junto al Dr. Daniel estaban ahí. A ellos les agradezco permitirme conocer organismos con vidas casi eternas, los juníperos, organismos que se rumora son veladores de sueños que rondan eternamente entre los volcanes. Me alegra poder haberme encontrado con grandes maestros académicos y de vida como ustedes. Deuda inmortal

podría llamarle a intentar pagarles su paciencia y compromiso durante mi formación científica y personal.

En mi nuevo hogar (Instituto de Ecología) tuve la oportunidad de conocer a un gran equipo de personas que piensan que la ciencia y el conocimiento guardan una gran belleza. Nancy, Myriam, Idalia, Christian, Laura, Ruth, Azalea, Alfredo, Gus y Beto. Además de todos los integrantes del “Mal del Puerco”, “La mafia colombiana” e integrantes del Departamento de Genética y Evolución, con quienes compartí dudas, recomendaciones académicas, comidas, salidas a campo y confesiones personales.

A pesar de tener un nuevo hogar, jamás olvide de donde venía: La Facultad de Estudios Superiores Iztacala o “Iztaharvard”. La FES-I fue el lugar donde realice la gran parte de mis estudios y de donde provienen diversas fuentes inagotables de consejos y experiencias: Chabela, Marco, Chicarcas, Toyo, Parra, Jo, Luisma, Miriam, Josimarts y Tere. Sin olvidar a mi amiguito Toño, socio museográfico quien me acompañó en momentos espirituales y gracias a quien llegue al mundo de la divulgación científica.

El mundo de la divulgación me permitió conocer a los físico-matemáticos que más quiero y respeto: Julio, Edgardo, Juan π , Alex, Leonel e Isaac grandes maestros y amigos que con paciencia me ayudaron a conseguir una meta más en este camino científico. Y junto a quienes descubrí la interdisciplina que guardan la Biología Matemática y la Biofísica.

También agradezco a personas de tan distintas carreras y pensamientos quienes me otorgaron sus manos amigas: Armando, Laux, Sandrita, Myriam, Marisol, Josué, Luisa y Mario. Al igual que a mis cuatitas Sandy, Anthony y Monche. Personas que deseo continúen en mi vida.

Gracias a cada uno de los integrantes de equipos nombrados más de una vez en mis recuerdos: “Grillos”, “Chapulínes”, “Levaduras”, “Bacterias”, “Coleópteros”, “Acociles”, “Cáncer(L-11)”, “Matemáticas”, “Matacochinos”, “Hermana República de Conciencia de las Matemáticas”, “Aleph zero” y los “Cuates”.

Finalmente agradezco a todos aquellos que tuvieron la paciencia de guiar, aconsejar y soportar el estrés de un tesista. Aprecio esas reuniones, fiestas, platicas, comidas, salidas y casas compartidas. Juntas, son experiencias que me permitieron ser actualmente quien soy.

Todo tiene su momento, ni antes ni después. Y este momento lo quise dedicar para usar esa palabra que viene desde lo más profundo de mi alma. Mis más sinceras y eternas GRACIAS !!!

ÍNDICE

INTRODUCCIÓN.....	1
MARCO TEÓRICO.....	3
Genética y genómica de poblaciones.....	3
Selección natural.....	4
Flujo genético.....	5
Deriva génica.....	6
Adaptación local.....	6
Convergencia evolutiva.....	7
Secuenciación de genomas reducidos.....	7
SISTEMA DE ESTUDIO.....	10
<i>Juniperus blancoi</i> y <i>Juniperus monticola</i>.....	10
Faja volcánica transmexicana y sierra madre Occidental.....	13
HIPÓTESIS.....	14
OBJETIVOS.....	15
Objetivo general.....	15
Metas particulares.....	15
MATERIAL Y MÉTODOS.....	16
Métodos moleculares.....	16
Muestreo y obtención de ADN.....	16
Diseño experimental de las librerías ddRAD.....	16
Ensamblaje de novo y filtrado de datos RAD.....	21
Demultiplexeo.....	21
Optimización del ensamblaje de novo.....	21
Ensamblaje de novo y análisis de calidad de todas las muestras.....	25
Filtrado por SNPs ligados.....	26
Genómica de poblaciones y análisis filogenético.....	26
Análisis de estructura poblacional por categorías MAF.....	26
Estructura genética de la distribución de la variación.....	27

Identificación de SNPs asociados a categorías altitudinales.....	27
Relaciones filogenéticas.....	28
RESULTADOS.....	29
Ensamblaje de novo y filtrado de datos RAD.....	29
Demultiplexeo.....	29
Optimización del ensamblaje de novo.....	31
Filtrado por SNPs ligados.....	37
Genómica de poblaciones y análisis filogenético.....	38
Análisis de estructura poblacional por categorías MAF.....	38
Estructura genética de la distribución de la variación.....	41
Identificación de SNPs asociados a categorías altitudinales.....	44
Relaciones filogenéticas.....	45
DISCUSIÓN.....	49
Ensamblaje de novo y filtrado de datos RAD.....	49
Demultiplexeo.....	49
Optimización del ensamblado de novo.....	49
Filtrado por SNPs ligados.....	51
Genómica de poblaciones y análisis filogenético.....	52
Poblaciones ancestrales de <i>J. blancoi</i> y <i>J. monticola</i>	52
Estructura genética de dos especies de <i>Juniperus</i> asociada a	
contrastantes historias evolutivas.....	52
SNPs asociados a categorías altitudinales.....	55
CONCLUSIONES.....	56
REFERENCIAS.....	57
ANEXO I.....	66
ANEXO II.....	67

INTRODUCCIÓN

Las especies y los organismos que las componen son producto de su historia evolutiva, es decir, de ancestría-descendencia a partir de modificaciones de un ancestro común (Castillo 2007). Dichas variaciones son generadas continuamente por mutación, pero su persistencia en el genoma está determinada por diferentes factores históricos, ambientales y genómicos (Wright 2001), como la selección natural que contribuye a dar forma a la variación fenotípica encontrada entre los individuos de una especie. La selección natural es la fuerza que favorece las características más aptas para prevalecer durante determinadas circunstancias ambientales (Darwin 1859). Sin embargo, un estudio de Dobzhansky & Pavlovsky (1957) basados en Wright (1949) y trabajando con moscas de la fruta vislumbró que los cambios evolutivos involucran interacciones entre la selección natural y la deriva génica.

La variación genética puede estar presente en ambientes heterogéneos según la distribución de la especie y es susceptible de sufrir presiones selectivas distintas según el ambiente. Esto conduce a la adaptación local de las poblaciones.

En México es posible encontrar claros ejemplos de especies que tienen poblaciones en condiciones ambientales diversas, pues es un país megadiverso con una gran heterogeneidad ambiental (Plascencia *et al.* 2011). Por ejemplo, los semi-desiertos mexicanos han fungido como centro de diversificación de los *Juniperus*, uno de los géneros de coníferas (Familia Cupressaceae) más diversos (Mao *et al.* 2010) y cuyas especies se distribuyen en condiciones desde semi-desérticas hasta de alta montaña (Adams 2014). En particular, dentro de la sección Sabina, las especies *J. blancoi* y *J. monticola* (Martínez 1946) tienen pocas poblaciones, todas restringidas a picos de montañas o arroyos. Cada especie cuenta con 3 variedades descritas en diferentes distribuciones altitudinales y ambientales que van de los 1,300 msnm hasta los 4,500 msnm (Mastretta-Yanes 2009; Adams 2014). Por ello, *J. blancoi* y *J. monticola* representan un sistema interesante para estudiar loci asociados a diferentes condiciones altitudinales.

Las zonas donde crecen *J. blancoi* y *J. monticola* presentan una compleja historia biogeográfica, climática y geológica que puede permitir tanto la persistencia

poblacional como eventos de divergencia y especiación (Mastretta-Yanes *et al.* 2015b). El acervo genético de estas poblaciones de coníferas debe estar bajo diferentes presiones de selección natural de acuerdo a la situación ambiental en la que se encuentran. Por lo que comparar la diferenciación genética bajo diferentes condiciones ambientales (alpinas y bosques de menor altitud) de poblaciones de *J. blancoi* y *J. monticola* representa un sistema interesante en el que evaluar si la selección natural está actuando sobre genotipos vinculados a adaptación a condiciones alpinas. Esto es interesante porque a diferencia de otras coníferas que habitan en rangos de temperatura limitados, *J. blancoi* y *J. monticola* tienen registradas poblaciones a alturas contrastantes asociadas a distintas condiciones ambientales.

La persistencia de las coníferas ante los actuales cambios climáticos es importante económicamente en la industria maderera, así como a nivel ecológico. Sin embargo, las coníferas han sido poco estudiadas a nivel genómico por que son organismos no modelo con tamaños de genomas muy extensos, los cuales llegan a más de 20 Gb (Zonneveld 2012). Para sobrepasar esta limitante en este trabajo utilizamos datos de *double digest Restriction Site Association DNA sequencing* (ddRAD-seq). La secuenciación ddRAD combinada con un adecuado manejo bioinformático permite reducir el error generado de un ensamblaje *de novo*.

Los datos obtenidos a partir de los loci ensamblados pueden asociarse a diferentes características ambientales. Estos datos permiten buscar loci asociados a condiciones alpinas, en este caso a partir de la comparación de poblaciones alpinas y de baja altitud en *J. blancoi* y *J. monticola*. Encontrar loci asociados a condiciones alpinas en coníferas es relevante porque nos puede informar sobre los mecanismos de adaptación local que les han permitido adquirir resistencia al frío o a la radiación UV, factores que varían de acuerdo a las diferencias altitudinales. Esto es útil para posteriormente emplear dichos loci como genes candidatos de resistencia a condiciones alpinas en juníferos y otras coníferas.

MARCO TEÓRICO

En la evolución biológica, los genes son considerados los “replicadores” y los individuos son los “vehículos” que utilizan los genes para su transmisión. Los genes con las modificaciones que son toleradas por la selección natural tienen continuidad evolutiva, pero los individuos son efímeros (Cabrero & Camacho 2002).

GENÉTICA Y GENÓMICA DE POBLACIONES

La genética de poblaciones es una disciplina interesada en el origen, cantidad y distribución de la variabilidad genética presente en las poblaciones naturales, así como el destino de esta variación a través del tiempo y el espacio. Esto constituye la base fundamental del cambio evolutivo, por lo que la evolución se considera como la transformación genética de las poblaciones en el tiempo y el espacio (Templeton 2006).

Para determinar los procesos históricos y demográficos de una población, los genetistas de poblaciones construyeron modelos matemáticos y propusieron métodos para probar e inferir los parámetros de los modelos a partir de los datos moleculares disponibles (Nielsen & Slatkin 2013).

La genética de poblaciones comenzó explorando en un solo gen el comportamiento de dos alelos que segregan de acuerdo a las leyes de Mendel, utilizando el equilibrio de Hardy-Weinberg como modelo nulo de lo que sucede si no opera ningún proceso evolutivo. Posteriormente se analizaron algunas decenas y centenas de genes que seguían sin ser comparables con el análisis brindado por el genoma completo (Eguiarte *et al.* 2013). Los recientes avances en la metodología molecular y estadística permiten reconstruir la historia evolutiva de las poblaciones con un mayor número de marcadores potenciando investigaciones ecológicas y de conservación a partir de la genómica de poblaciones.

Las variaciones a nivel de genes o genomas son conducidas ya sea por deriva génica, flujo génico y/o selección natural.

Selección natural

Considerada por muchos como “La arriesgada idea de Darwin” (Dennett 1996), la teoría de la selección natural es la pieza central del origen de las especies y de la teoría evolutiva (Futuyma 2013). La selección natural parte del supuesto de que los organismos se reproducen heredando sus características a la progenie, sin embargo las condiciones ambientales generalmente no serán las adecuadas para todos los miembros de una población, ya que usualmente hay variación en sus características. De forma que solo aquellos con las características que les confieran mayor adecuación sobrevivirán y podrán reproducirse con mayor probabilidad que los individuos con características asociadas a una menor adecuación (Darwin 1859).

La selección natural esta tradicionalmente dividida dentro de cuatro tipos: a) selección estabilizadora que se produce siempre que un individuo con un rasgo ha tenido éxito con respecto a la gran mayoría de los individuos que comprenden una población. La mayoría de las variaciones con rasgos extremos en la distribución son seleccionados en contra, a efecto de mantener el estatus quo con la mayoría de los individuos que poseen ese rasgo de éxito; b) selección direccional, que favorece individuos con rasgos de un extremo de la distribución y generan una mayor adecuación a los individuos que los portan; c) selección disruptiva, que suele abarcar un rango de variación y los fenotipos más aptos están en cada uno de los dos extremos de variación, siendo las formas intermedias seleccionadas en contra. El último tipo, d) el la selección balanceadora, en la cual dos o más variaciones de un mismo gen o rasgo son mantenidos en la población (Zeigler 2014) (Figura 1).

Aunque la selección natural disruptiva sea una fuerza motriz muy importante en la evolución, otras fuerzas, en particular, el flujo génico, son aspectos integrales del proceso de adaptación local (Kisdi 2002; Kawecki & Ebert 2004).

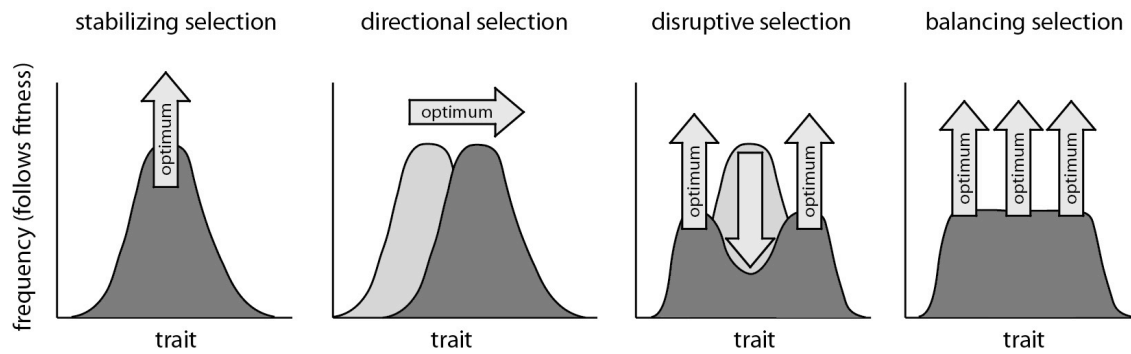


Figura 1. Tipos de selección. El área oscura indica la distribución de la frecuencia actual de un rasgo que se realiza por los individuos de una población. El área clara indica la misma distribución de una generación pasada. Las flechas indican el rasgo óptimo que fue seleccionado positivamente, los demás valores fueron seleccionados negativamente. (a) selección estabilizadora: mantiene a la población en un valor óptimo estable; (b) selección direccional: transforma el valor de un rasgo mediante el incremento de la frecuencia de los individuos cercanos a una distancia óptima; (c) selección disruptiva: incrementa la frecuencia de valores grandes y pequeños de un rasgo a costa de los valores intermedios; y (d) selección balanceadora: selección del arreglo óptimo entre varios rasgos. Tomado de Loewe (2008).

Flujo génico

Se refiere a todos los mecanismos que generan movimiento de genes de una población a otra, debido al movimiento de gametos, semillas, individuos juveniles o adultos que pueden actuar como una fuerza que mantiene integrada a la especie (Aguirre-Planter 2007).

El flujo génico permite conectar a un conjunto de demos (poblaciones) que presentan adaptación local, adaptación debida a la selección natural actual o muy reciente, que varía en relación con las diversas condiciones ambientales que experimenta cada población (Kawecki & Ebert 2004). El efecto del flujo génico tiende a oponerse a los efectos de selección local y así limitar su adaptación. Sin embargo también puede reponer la población local (e.g. contrarrestar efectos de la endogamia) y

la variación genética local que son pre-requisitos para la evolución por selección natural (Lenormand 2002).

Deriva genética

Definida por Wright (1949), la deriva génica se refiere al cambio aleatorio en la frecuencia de los genes. Tales variaciones son causadas por accidentes en el muestreo de genes en poblaciones de tamaño finito. La hipótesis del azar en la deriva génica no debe ser usada como escapatoria, ya que para ser aceptada requiere ser demostrada (Dobzhansky & Pavlovsky 1957).

Adaptación local

Es la adaptación a ambientes locales. Particularmente si la dispersión es limitada las poblaciones a menudo se adaptan a ambientes locales. Una manera de destruir el complejo de genes adaptados localmente puede ser la hibridación o flujo génico entre diferentes poblaciones (Templeton *et al.* 1986)

Para que una población pueda adaptarse localmente debe existir un ambiente heterogéneo que genere una presión selectiva local (Blanquart *et al.* 2013). La base genética de la adaptación local podría ser explicada por: (a) pleiotropía antagónica, en la cual hay un efecto contrario en la adecuación que genera los alelos de un locus del genoma en ambientes distintos o (b) neutralidad condicional, en el que un alelo muestra ventaja en su adecuación en determinado ambiente, pero es neutral en un ambiente distinto (Kawecki & Ebert 2004; Mitchell-Olds *et al.* 2007; Anderson *et al.* 2013).

La adaptación a las condiciones ambientales locales juega un papel fundamental en la generación y mantenimiento de la biodiversidad durante la contracción y expansión de la distribución geográfica de las especies y mediante la especiación, además de las dinámicas ecológicas y evolutivas de las interacciones de especies (Blanquart *et al.* 2013). La adaptación local también refleja una profunda complejidad al

ser el resultado de las interacciones entre la selección natural, el flujo de genes y otros procesos evolutivos (Kawecki & Ebert 2004).

Convergencia evolutiva

A menudo diferentes especies evolucionan de manera similar a los cambios ambientales. Un ejemplo común son el desarrollo de las alas en insectos, aves y murciélagos (Losos 2011), pero la convergencia también puede darse a nivel molecular, como la evolución de fenotipos parecidos a partir de los mismos genes. Un ejemplo de lo anterior recientemente se encontró en coníferas (Yeaman *et al.* 2016).

El estudio de genes responsables de adaptación incrementó el interés en determinar si episodios independientes de adaptación se han logrado a través de exactamente los mismos mecanismos moleculares o no (Tiffin & Ross-Ibarra 2014).

SECUENCIACIÓN DE GENOMAS Y OBTENCIÓN DE SNPS

En los últimos años se han desarrollado múltiples tecnologías para el análisis de genomas (Metzker 2010), que pueden ser clasificados en dos grandes grupos; marcadores bioquímicos y marcadores moleculares. Dentro de esta segunda clasificación encontramos a la secuenciación de ADN y a la identificación de SNPs (Single Nucleotide Polymorphism), que son mutaciones puntuales en secuencias de ADN (Glenn 2011; Eguiarte *et al.* 2013). Los SNPs son abundantes en los genomas, su variación entre los individuos suele estar representada por dos o más de los cuatro nucleótidos posibles y evolucionan en términos generales de acuerdo a un modelo sencillo de mutación de sitios infinitos (Schlötterer 2004). Los SNPs proporcionan una muestra representativa del genoma, reducen la varianza de muestreo entre loci y permiten identificar loci candidatos a la selección (Morin *et al.* 2004).

Actualmente, los SNPs pueden analizarse por diversos medios (Khlestkina & Salina 2006), uno común es el llamado RAD-seq (*Restriction Site Association DNA sequencing*, Davey *et al.* 2011). El método RAD consiste en el análisis del ADN asociado a sitios de restricción, por medio de secuenciación masiva que permite

submuestrear el genoma en lugares presuntamente homólogos en muchos individuos para identificar SNPs e indels en secuencias de ADN cortas (Baird *et al.* 2008).

El uso de RADseq está incrementando en un amplio espectro de análisis evolutivos, que van desde relaciones filogenéticas dentro de un género (Jones *et al.* 2013) hasta estudios de asociación en la escala del genoma para identificar regiones bajo selección (Parchman *et al.* 2012) y de estudios ecológicos y de conservación (Narum *et al.* 2013).

Desde el 2008 a la fecha se han desarrollado diferentes variaciones del RADseq, entre las que se incluyen la doble digestión RAD (ddRAD; Figura 2).

El método ddRAD ofrece un amplio grado de modificación, pues dependiendo de las enzimas elegidas y el tamaño de fragmentos seleccionados, se pueden obtener hasta decenas de miles de SNPs por individuo a muy bajo costo, además de permitir identificar y excluir fragmentos excesivamente frecuentes, cualidad muy valiosa para organismos con genomas grandes no secuenciados aún. A su vez esta técnica lidia con ciertas desventajas como ser susceptible a pérdida de alelos y requerir ADN genómico de más alta calidad en comparación de los demás métodos RAD (Puritz *et al.* 2014).

Como cualquier método, ddRAD también posee errores (Mastretta-Yanes *et al.* 2015a), los más frecuentes son los relacionados directamente con técnicas de laboratorio inadecuadas, la presencia de indels (Peterson *et al.* 2012); variación en la cobertura (Catchen *et al.* 2013); parálogos y regiones repetidas (Hohenlohe *et al.* 2012) y el mapeo sin usar genoma de referencia (Hohenlohe *et al.* 2012).

Además, la escala sobre la cual los datos serán ensamblados y filogenéticamente útiles esta inherentemente limitada por mutaciones que alteran los sitios de reconocimiento de restricción causando “locus dropout” que se ven manifestados como un factor limitante para identificar loci ortólogos (Eaton 2014). Los errores además pueden llevar a inferencias inexactas en cuanto homocigotos, equilibrio Hardy-Weinberg, endogamia, estructura de la población y expansión demográfica.

Sin embargo, el efecto de estos errores puede mitigarse al reportar las tasas de error o incorporarlas dentro del análisis de datos (Mastretta-Yanes *et al.* 2015a). Efectuando este análisis en las tasas de error, potenciamos la confiabilidad de la

secuenciación RAD en el alineamiento de los datos, por lo que pueden utilizarse para descubrir polimorfismos (SNP) entre especies que no tienen secuencia de referencia (Hohenlohe *et al.* 2012), característica presente en la mayoría de organismos que no son modelos de estudio.

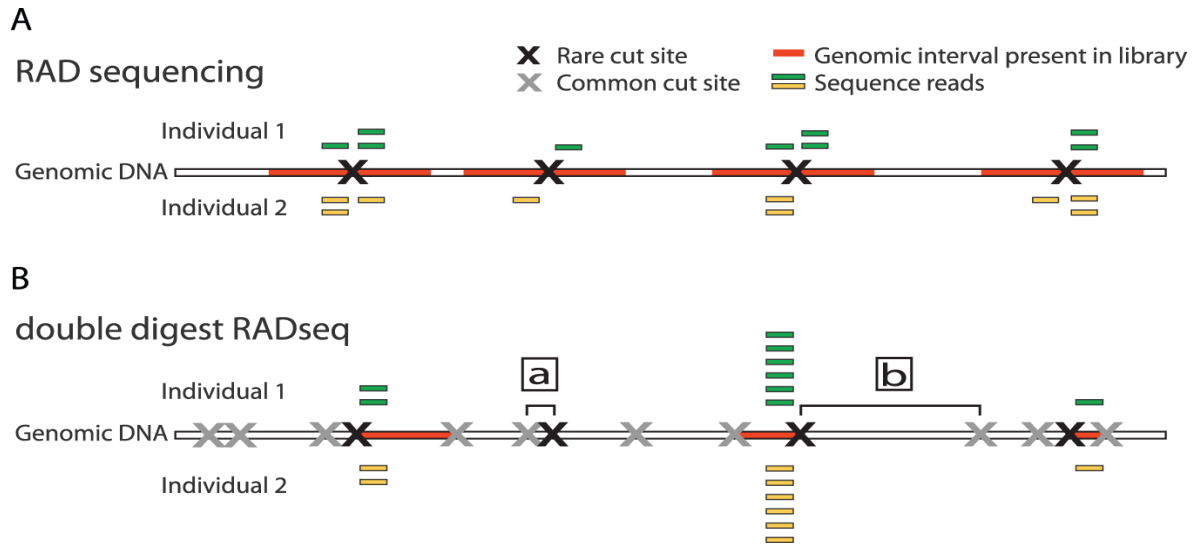


Figura 2. Secuenciación RAD comparada con ddRAD. (A) RAD-seq utiliza una enzima de restricción, con tamaño de selección amplio genera librerías representadas de manera reducida. (B) ddRAD-seq con dos enzimas de restricción, excluye regiones flanqueadas por [a] sitios cercanos o [b] distantes del sitio de restricción. Recuperación de fragmentos representados en la librería son señalados en rojo. Correlación de regiones entre individuos representada con barras verdes y amarillas. Tomado de Peterson *et al.* (2012).

El ensamblado de secuencias RAD puede realizarse con diversos programas especializados, como *Stacks* (Catchen *et al.* 2013), *Rainbow* (Chong *et al.* 2012) o *pyRAD* (Eaton 2014). Dentro de estos, *pyRAD* intenta maximizar la información filogenética entre muestras dispares de secuenciaciones RAD. La mayor diferencia entre *pyRAD* y *Stacks*, es la manera en que los loci son agrupados. *Stacks* reúne las lecturas (*reads*) agrupándolos con respecto al menor número de diferencias (N) entre ellas, pero no contempla indels. En cambio *pyRAD* usa secuencias similares basado en un alineamiento global de las secuencias a través del programa Vsearch, permitiendo un mejor conjunto de datos que incluye la relación de muestras alejadas (Eaton 2014).

Por estos motivos, *pyRAD* se recomienda para análisis filogenéticos o que involucren a varias especies.

SISTEMA DE ESTUDIO

JUNIPERUS BLANCOI Y JUNIPERUS MONTICOLA

Los bosques de encinos y coníferas son tipos de vegetación muy característicos que comienzan a hacerse presentes en el Centro y Sur de México (Arroyo-Cabrales *et al.* 2008). Uno de los grupos más diversos dentro de las coníferas que habitan estos ecosistemas es *Juniperus*, que comprende arboles y arbustos siempre verdes, generalmente dioicos rara vez monoicos y es uno de los géneros de más reciente diversificación dentro de las coníferas (Rai *et al.* 2008; Adams 2014).

El tamaño del genoma de los *Juniperus* es ~12 Gb distribuidos en 11 cromosomas (Nystedt *et al.* 2013), característica que lo coloca junto a los genomas más pequeños dentro de las coníferas, las cuales llegan a más de 20 Gb (Zonneveld 2012; Zimin *et al.* 2014). Aún así los juníperos tienen los genomas más grandes en comparación a las Angiospermas, que por lo general son menores de 5 Gb (Dodsworth *et al.* 2015).

Actualmente el género *Juniperus* se distribuye a lo largo del hemisferio norte (Figura 3). *Juniperus* tiene aproximadamente 75 especies distribuidas en tres secciones (Figura 4): *Caryocedrus* (sección más ancestral del género, con una sola especie en el Mediterráneo), *Juniperus* (con 14 especies) y *Sabina* (sección más reciente, que presenta la mayor diversidad, aproximadamente 60 especies), con amplia distribución en México (Adams 2014).

El origen de los *Juniperus* fue probablemente en Eurasia. Su diversificación comenzó durante el Eoceno, se ralentizó en el Oligoceno y finalmente aceleró en el Mioceno-Plioceno. Los juníperos de la sección *Sabina* llegaron del Norte de América por un proceso divergente ocurrido en Europa hace ~38 Ma. Colonizaron hábitats cálidos-semiáridos hasta ocupar climas fríos y húmedos abundantes durante el Mioceno (Mao *et al.* 2010). Interesantemente, parece que las especies de *Juniperus*

adaptadas a altas montañas son las de más reciente divergencia dentro del género (Mao *et al.* 2010). Entre estas especies de relativa reciente diversificación que ocupan hábitats alpinos se encuentran *J. blancoi* y *J. monticola*.

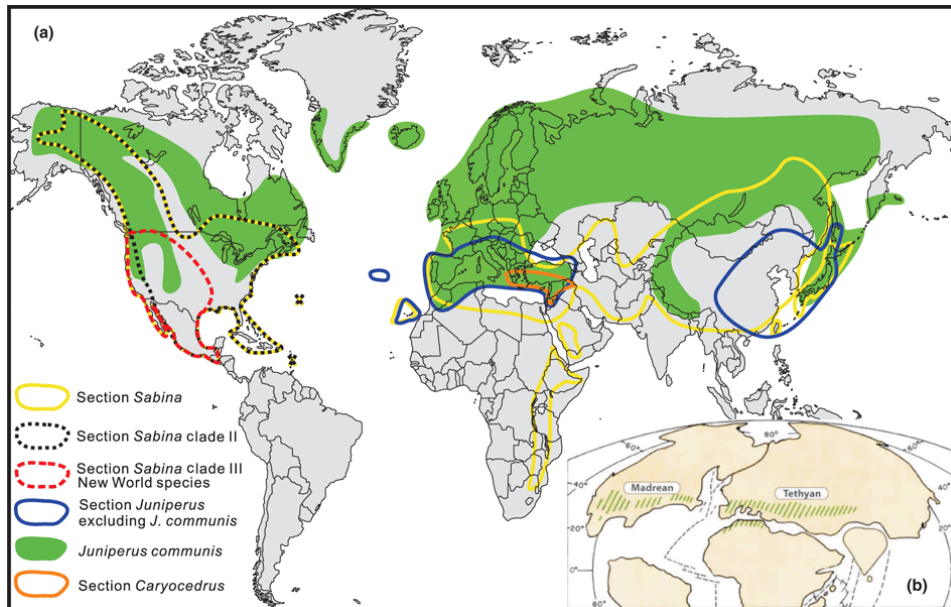


Figura 3. Distribución del género *Juniperus*. Tomado de Mao *et al.* (2010).

Juniperus blancoi es una especie rara con nueve poblaciones conocidas, pequeñas y separadas en alturas de 1,350 a 3,250 msnm (Mastretta-Yanes 2009). Cambios en el clima, formación de montañas y heterogeneidad ambiental favorecieron su diversificación hace 9 Ma (Moreno-Letelier *et al.* 2014). Cuenta con tres variedades: (1) *huehuetensis* representada por arbustos en los cerros más altos de la SMOc (Mohinora y Huehuento), (2) *mucronata* con árboles que habitan a lo largo de las orillas de arroyos en Sonora (Yécora) y (3) *blancoi* con árboles y arbustos que habitan a lo largo de arroyos permanentes y planicies de inundación distribuidas al fondo de los arroyos, en bosques de pinos-encinos-juníferos (El Salto, Rio Patitos, La Preciosita San José Rincón y Presa Brokman) (Mastretta-Yanes *et al.* 2012).

Juniperus monticola tiene más y mayores poblaciones que *J. blancoi* pero también puede considerarse una especie rara, ya que sus poblaciones son relativamente pequeñas y restringidas a picos aislados de la FVTM y SMO. Su diversificación fue a partir de un linaje de Norte América hace ~5 Ma. Cuenta con tres

variedades: (1) *orizabensis*, arbustos que habitan en el límite de la vegetación arbórea y por encima de este. Se distribuyen desde los 2,850 m hasta los 4,500 m en grietas rocosas. (2) *monticola*, árboles o arbustos de hábitats rocosos en bosques subalpinos de encino-juníferos, pinos-oyameles, con distribución desde los 2,400 m hasta los 4,300 m, por encima del límite de la vegetación arbórea. (3) *compacta*, arbustos que habitan desde los 3,000 m hasta 4,500 m de elevación en el borde del límite de la vegetación arbórea de bosques de pino o por encima de este en montañas con pastizales alpinos, usualmente en grietas rocosas o sobre afloramientos volcánicos rocosos (Adams 2014).

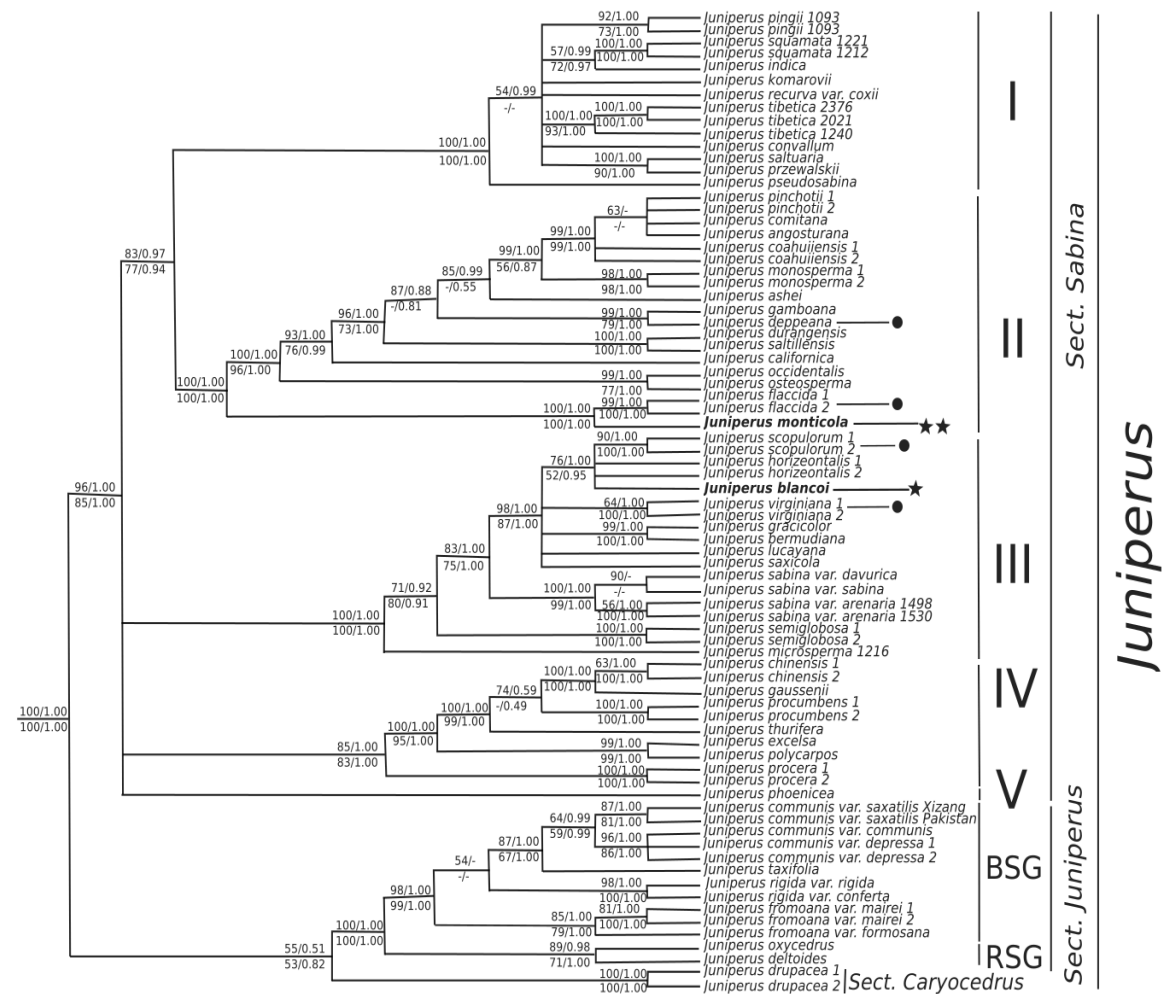


Figura 4. Filogenia de *Juniperus*. Con una estrella se señala a *Juniperus blancoi*, con dos estrellas a *Juniperus monticola* y con círculos a los grupos externos (outgroups) ocupados. Tomado de Mao *et al* (2010).

FAJA VOLCÁNICA TRANSMEXICANA Y SIERRA MADRE OCCIDENTAL

Se han definido 19 provincias biogeográficas dentro del territorio de México con base en cuatro sistemas de clasificación, entre las cuales se encuentran la Faja Volcánica Transmexicana (FVTM) y la Sierra Madre Occidental (SMOc; CONABIO, 1997). Estas provincias se caracterizan por presentar picos de montañas mayores a los 3,000 m de altitud y presentar bosques de coníferas y encinos, sin embargo ambas tienen orígenes geológicos e historias climáticas diferentes.

En la actualidad la FVTM se extiende desde Nayarit, en el Océano Pacífico hasta Veracruz en el Atlántico y se prolonga de nuevo al Pacífico (Mooser & Ramírez 1987). La FVTM se conformó por procesos fisiográficos ocurridos en la Era Cenozoica, hace aproximadamente 20 Ma. Los procesos que definieron su evolución y su configuración involucran apilamientos en un arco volcánico. Los mayores cambios en su perfil de relieve ocurrieron durante el Cuaternario (1.5 Ma al presente) en el Centro de México, donde se produjo la alineación de estratovolcanes como el Nevado de Colima, Sierra Nevada (Iztaccíhuatl-Popocatepetl) y Cofre de Perote en un arco volcánico (Arroyo-Cabrales *et al.* 2008). En su porción este dominan climas templados subhúmedos, semiárido templado y semifrío subhúmedo, mientras que en la porción oeste dominan los climas semicálidos subhúmedos (Gámez *et al.* 2012). Una característica importante de la FVTM es que los grandes estratovolcanes que la conforman representan picos de gran pendiente, inmersos en territorios semicálidos, de modo que alcanzan a formar condiciones alpinas aisladas entre sí como “islas en el cielo” (Mastretta-Yanes *et al.* 2015b).

Por otro lado, la SMOc es el más largo y continuo de los sistemas montañosos de México: una cadena de cerca de 1,160 km de longitud, que corre desde la costa del Océano Pacífico, al Sur de la frontera mexicano-estadounidense hasta la altura de Nayarit y Jalisco, donde converge con la FVTM (Morrone 2005; González-Elizondo *et al.* 2012). La SMOc es más antigua que la FVTM, es resultado de diferentes episodios magmáticos y tectónicos durante el Cretácico-Cenozoico (65 Ma) (Ferrari *et al.* 2005).

HIPÓTESIS

- Existirán diferencias genéticas entre poblaciones dentro de cada especie de *Juniperus* que ocurren en condiciones de alta y de baja altitud.
- SNPs presentes en poblaciones de *J. blancoi* y *J. monticola* en condiciones ambientales similares pueden ser considerados SNPs candidatos que ocurrieron mediante procesos de adaptación local.
- *Juniperus blancoi* y *J. monticola*, a pesar de ser especies distintas y no cercanas dentro del género *Juniperus*, podrán presentar loci asociados a condiciones de alta y de baja montaña independientemente de la especie, debido a procesos de convergencia evolutiva.

OBJETIVOS

OBJETIVO GENERAL

Comparar poblaciones de *J. blancoi* y *J. monticola* bajo diferentes condiciones ambientales (alpinas y bosques de menor altitud), mediante un análisis bioinformático de datos ddRAD, para encontrar posibles loci candidatos sobre los que la selección natural esté actuando.

METAS PARTICULARES

Para cumplir con lo anterior se establecieron las siguientes metas particulares:

- ❖ Optimizar los parámetros de ensamble *de novo* de datos ddRAD con el programa *pyRAD*, minimizando el error y maximizando la recuperación de SNPs.
- ❖ Ensamblar *de novo* y filtrar los datos arrojados por la secuenciación ddRAD con los parámetros optimizados.
- ❖ Examinar la estructura poblacional de ambas especies.
- ❖ Buscar SNPs asociados a condiciones alpinas, mediante la comparación de las poblaciones de alta y de baja montaña de ambas especies.
- ❖ Determinar las relaciones filogenéticas entre poblaciones de *J. blancoi* y *J. monticola*.

MATERIAL Y MÉTODOS

El presente trabajo es un análisis bioinformático de datos ddRAD obtenidos previamente para las especies *J. blancoi* y *J. monticola* según se describe a continuación. Se presentan análisis propios a partir del paso *ensamblado* de novo.

MÉTODOS MOLECULARES

Muestreo y obtención de ADN

El muestreo (Figura 5) y obtención de ADN fue realizado en trabajos previos. En total se colectaron durante 2008-2009 (Tabla 2) ocho poblaciones de *J. blancoi* y sus variedades en la SMOc y la FVTM (Mastretta-Yanes 2009); durante los años 2010-2011 doce poblaciones de *J. monticola* en la Faja Volcánica Transmexicana y una en la Sierra Madre Oriental (Mastretta-Yanes 2014); y como grupos externos se usaron cuatro poblaciones de *J. flaccida*, una de *J. zanonii* y *J. deppeana* en el 2012 (Mastretta-Yanes 2014). También se analizaron muestras de las especies *J. scopolorum* y *J. virginiana* proporcionadas por R. Adams (Mastretta-Yanes 2009).

La obtención de ADN de las muestras de cada individuo de *J. monticola* se realizó a partir de sus hojas con el kit Quiagen DNeasy Plant Mini (69106) y la extracción de *J. blancoi* y los *outgroups* usando el método CTAB modificado de Vázquez-Lobo (1996). La calidad del ADN extraído se confirmó mediante electroforesis.

Diseño experimental de las librerías ddRAD

Las muestras fueron secuenciadas en un trabajo previo (Mastretta-Yanes 2014) con la técnica ddRAD (Peterson *et al.* 2012; Parchman *et al.* 2012) utilizando las enzimas SbfI-HF (TGCA|C) y MseI (T|TA). SbfI es una enzima de corte muy raro, por lo que esta

combinación de enzimas permite generar relativamente pocos loci en un genoma de ~11Gb.

Las extracciones de ADN fueron azarosamente distribuidas en 10 grupos, cada uno destinado en principio a constituir una librería para la secuenciación. Cada grupo fue compuesto por 24 muestras que incluían individuos, replicas y negativos (sin muestra de ADN). Se consideró que en cada grupo hubieran dos replicas de individuos dentro de la librería (denominados como intralibrería), una replica de un individuo dentro de otra librería (denominado como interlibrería) y un negativo (ANEXO I). Las replicas tenían la misma fuente de ADN pero fueron procesadas con un *barcode* independiente. Las posiciones en las placas de PCR fueron azarosamente seleccionadas a un individuo con un *barcode* individual. La digestión, ligación y pasos de PCR fueron realizados en las mismas placas para cada grupo. Las muestras de cada grupo fueron después mezcladas (*pool*) juntas en un mismo tubo y la selección del tamaño para los grupos fue realizada en el mismo gel.

Con el propósito de optimizar los recursos, los grupos fueron secuenciados con single-ends *reads* en 6 *lanes* para las 10 librerías, ubicando las librerías 1 y 10 en *lanes* separadas y a las librerías restantes en pares por cada *lane* (2-3, 4-5, 6-7 y 8-9) de Illumina HiSeq2000, usando el servicio provisto por Lausanne Genomic Technologies Facility, Suiza (Mastretta-Yanes 2014). Los detalles completos utilizados durante el trabajo de laboratorio se encuentran en Mastretta-Yanes (2014).

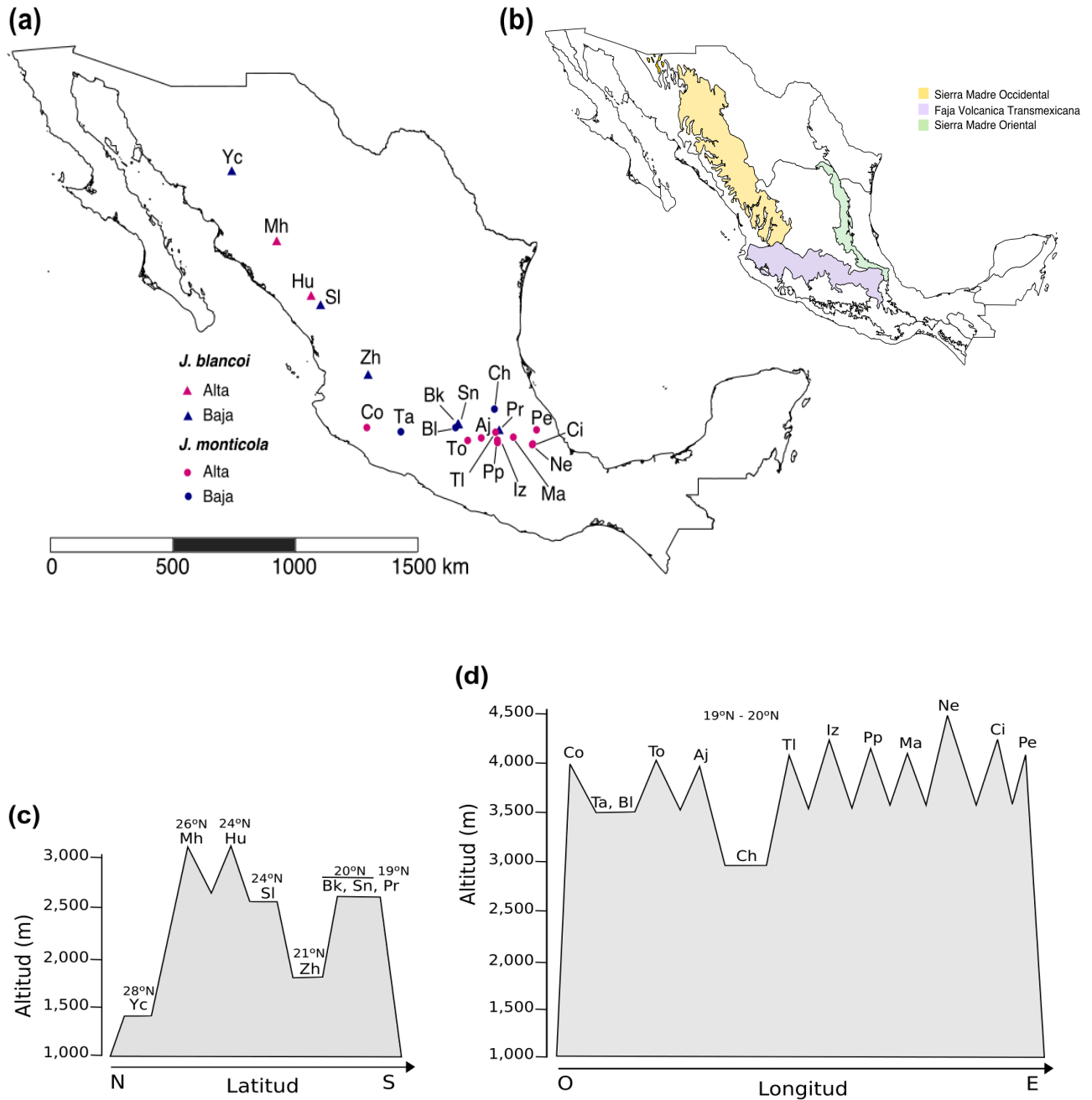


Figura 5. Distribución geográfica de las 21 poblaciones de *Juniperus* en las diferentes regiones biogeográficas. (a) Puntos de muestreo con la categoría altitudinal asignada. (b) Regiones biogeográficas de los puntos de muestreo. Perfil altitudinal de las poblaciones de (c) *J. blancoi* (modificado de Moreno *et al.*, 2014) y (d) *J. monticola*. (ver Tabla 1 y ANEXO II).

Tabla 1. Agrupación de las principales características de cada población.

Población de muestreo	Clave	Especie	Estado de Rep.	Región	Altitud	Cat Alt*	Coordenada
Huehuento	Hu	<i>J. blancoi</i>	Dgo.	SMOc	3150-3200	Alta	24.07528 -105.7395
El Salto	Sl	<i>J. blancoi</i>	Dgo.	SMOc	2500	Baja	23.756317 -105.378333
Mohinora	Mh	<i>J. blancoi</i>	Chih.	SMOc	3150-3200	Alta	25.96011 -107.050288
Preciosita	Pr	<i>J. blancoi</i>	Tlax.	FVTM	2700-2800	Baja	19.445106 -98.574342
Presa Brokman	Bk	<i>J. blancoi</i>	Edo. Méx.	FVTM	2700-2800	Baja	19.652639 -100.149361
Río Patitos	Zh	<i>J. blancoi</i>	Jal.	SMOc	1700	Baja	21.351944 -103.567389
San José Rincón	Sn	<i>J. blancoi</i>	Edo. Méx.	FVTM	2700-2800	Baja	19.652639 -100.124083
Yécora	Yc	<i>J. blancoi</i>	Son.	SMOc	1300	Baja	28.374722 -108.76056
Ajusco	Aj	<i>J. monticola</i>	D.F.	FVTM	3900	Alta	19.212944 -99.256944
Cerro Blanco	Bl	<i>J. monticola</i>	Edo. Méx.	FVTM	3524	Baja	19.569861 -100.236444
Citlaltépetl	Ci	<i>J. monticola</i>	Pue.	FVTM	4210	Alta	19.5006306 -97.294806
Cofre de Perote	Pe	<i>J. monticola</i>	Ver.	FVTM	4058	Alta	19.492067 -97.149448
El Chico	Ch	<i>J. monticola</i>	Hgo.	SMOr	2976	Baja	20.207917 -98.754111
Iztaccíhuatl	Iz	<i>J. monticola</i>	Edo. Méx.	FVTM	4201	Alta	19.140889 -98.642611
La Negra	Ne	<i>J. monticola</i>	Pue.	FVTM	4420	Alta	18.982667 -97.310833
Malinche	Ma	<i>J. monticola</i>	Tlax.	FVTM	4055	Alta	19.24075 -98.034417
Nevado de Colima	Co	<i>J. monticola</i>	Jal.	FVTM	3935	Alta	19.571528 -103.616444
Nevado de Toluca	To	<i>J. monticola</i>	Edo. Méx.	FVTM	4037	Alta	19.5126306 -99.772389
Popocatepetl	Pp	<i>J. monticola</i>	Edo. Méx.	FVTM	4120	Alta	19.05333 -98.6365
Tancítaro	Ta	<i>J. monticola</i>	Mich.	FVTM	3506	Baja	19.425778 -102.319944
Tláloc	TI	<i>J. monticola</i>	Edo. Méx.	FVTM	4075	Alta	19.411583 -98.711306

* La determinación de la categoría altitudinal es relativa a cada especie, considerando cómo ha sido definida por trabajos previos (Moreno-Letelier *et al.* 2014 para *J. blancoi*) y si se encontró creciendo en bosque o pastizal alpino (*J. monticola*).

Tabla 2. Tamaños de muestra utilizados (número de individuos por población, ver detalles más adelante).

	Población	# Individuos colectados	# Individuos replicados	# Individuos en ensamblado
<i>J. blancoi</i>	Huehuento	10	1	10
	El Salto	8	2	8
	Mohinora	10	1	10
	Preciosita	6	2	6
	Presas Brokman	6	0	4
	Rio Patitos	8	1	7
	San José Rincón	6	1	5
	Yécora	8	1	6
Total Jb		62	9	56
<i>J. monticola</i>	Ajusco	10	1	8
	Cerro Blanco	10	3	9
	Citlaltépetl	10	1	10
	Cofre de Perote	10	0	7
	El Chico	10	1	9
	Iztaccíhuatl	10	1	8
	La Negra	10	3	10
	Malinche	10	2	9
	Nevado de Colima	10	1	10
	Nevado de Toluca	10	1	8
	Popocatépetl	10	2	8
	Tancítaro	10	0	10
	Tlálloc	10	1	10
	Total Jm		130	17
Outgroups		8	4	0
Negativos		10	0	0
Total		210	30	172

*Los outgroups consisten en cuatro poblaciones de *J. flaccida* (dos replicas), una de *J. zanonii* (una replica), *J. deppeana*, *J. scopolorum* (una replica) y *J. virginiana*. Los negativos incluían todos los reactivos con los que se procesaron las muestras, a excepción del ADN y se consideró uno para cada librería.

ENSAMBLAJE DE NOVO Y FILTRADO DE DATOS RAD

Demultiplexeo

Los *reads* de los *Juniperus* fueron previamente demultiplexeadas y filtradas de acuerdo a su calidad usando el software *Stacks* v. 1.17 mediante: (1) el truncado final de los *reads* en una longitud de 87 pb (fundamentado en la disminución de calidad en esta posición para cada *lane* 10), (2) removiendo cualquier *read* con *uncalled base* (N), (3) descartando *reads* con bajos *scores* de calidad (*score* límite 22-28, dependiendo de la *lane* secuenciada), (4) descartando *reads* que hubieran sido marcadas por el filtro de pureza de Illumina, (5) filtrando secuencias de los adaptadores y (6) rescatando *tags* (distancias máximas de 1 entre los *barcodes*) (Mastretta-Yanes 2014).

Optimización del ensamblado de novo

Los datos demultiplexeados en *Stacks* fueron ensamblados y genotipificados usando *pyRAD* v.3.0.5 (Eaton 2014), a partir del paso 2 (i.e. posterior al demultiplexeo, Tabla 3). El software *pyRAD* es una herramienta de ensamblaje y genotipificación que utiliza una agrupación de alineamientos mediante algoritmos que permiten la incorporación de variación de indeles, por lo que es particularmente útil para análisis que involucren especies diferentes (Eaton 2014).

El ensamblaje de novo en *pyRAD* se realizó como se detalla a continuación, usando una computadora Mac OS Yosemite versión 10.10.3, 3.5 GHz Intel Core i7, 16 GB 1600 MHz DDR3. Primero se realizó una corrida con 240 muestras para conocer la cantidad de *reads* obtenidos por cada muestra (Tabla 2, Individuos colectados y replicados). De manera similar al método descrito por Mastretta-Yanes (Mastretta-Yanes *et al.* 2015a) en el software *Stacks*, se optimizaron los parámetros de ensamblaje en *pyRAD*, para minimizar el error y maximizar la recuperación de SNPs. *pyRAD* se corrió primero con diferentes modificaciones de algunos parámetros en el subconjunto de 30 muestras replicadas, (Tabla 2, Individuos replicados).

Posteriormente, se evaluaron los efectos en el modelo de llamado de SNPs sobre las tasas de error, el número de datos perdidos (*missing data*) y el número de SNPs recuperados.

Tabla 3. Descripción de los pasos del procesamiento de ensamblado *de novo* de *pyRAD*

Paso desarrollado por <i>pyRAD</i>	Acción que realiza	Archivos generados/utilizados.
Filtrado	Filtra el llamado de bases de baja calidad. Sitios por debajo del valor establecido son tomados como “N” y <i>reads</i> con más “N” de las permitidas son descartadas.	Usa el nivel de calidad Phred registrado en los archivos FASTQ. Arroja archivos .edit del directorio edits.
Agrupamiento de muestras	Registra el número de veces que se observa un <i>read</i> . Agrupa según su coincidencia dentro de la similitud de secuencia establecida.	Usa a VSEARCH. Arroja archivos .clustS.gz en el directorio clust.xx.
Estimar tasas de error y heterocigosidad	Los estima conjuntamente a partir del recuento de las bases en cada sitio de cada <i>cluster</i> .	Usa la ecuación ML de (Lynch 2008) Arroja un archivo Pi_estimate.txt en el directorio stats.
Crear secuencias consenso	Descarta aquellos con “-“ de la cobertura mínima, “+” del número de sitios indeterminados, “+” del número de sitios heterocigóticos, o “+” del número de alelos permitidos.	Usa el promedio del paso anterior.
Agrupar secuencias consenso a través de las muestras	Si en heterocigotos, un alelo es azarosamente muestreado y utilizado en este paso, de todas formas ambos alelos son retenidos en el conjunto final de datos.	Usa el mismo agrupamiento del paso anterior.
Alinear y filtrar parálogos	Proceso más rápido que facilita la optimización de la cobertura en las muestras.	Arroja archivos .loci tipo FASTA o formatos alternativos para análisis posteriores, estadísticas finales en el directorio stats.

Tabla 4. Descripción de los parámetros *pyRAD* y su efecto en el ensamble (ver Figura 6).

Parámetro	Efecto en el ensamble
Mínimo de cobertura para un cluster (-Mindepth) default 5	Se refiere al mínimo de “ <i>depth-of-coverage</i> ” (<i>reads</i>) para hacer un <i>base call</i> en cada sitio dentro del <i>cluster</i> . Si es bajo los errores pueden pasar por datos reales, considerando mas loci de los existentes. Si es alto los datos reales pueden pasar como errores, pareciendo homocigotos.
Máximo numero de sitios con calidad menor a 20 (-NQual) default 4	Se refiere al máximo numero de sitios de baja calidad, sitios indeterminados (N) en el paso 2. Se selecciona con respecto al umbral de agrupamiento.
Umbral de agrupamiento (WClust) default .90	Umbral de similitud utilizado por <i>Vsearch</i> para agrupar las secuencias, se ingresa como decimal. Se recomienda usar de acuerdo al NQual elegido. Este se calcula como el porcentaje del resultado de la longitud de los <i>reads</i> menos el NQual (ej. Long.=80-NQual=4/80 = 0.95). Si es bajo permite mayor número de sitios indeterminados y si es alto excluye <i>reads</i> con muchos sitios indeterminados
Máximo de individuos con sitios polimórficos (MaxSH) default 4	Se refiere al máximo número de sitios polimórficos compartidos en un locus. Permite detectar parálogos potenciales, como un sitio de heterocigotos compartida a través de muchas muestras, probablemente representa la agrupación de parálogos con una diferencia fija en lugar de un verdadero sitio de heterocigotos.
Mínima cobertura por taxón (MinCov) default 4	Mínimo número de muestras para un locus dado que se retendrá en los datos finales. Si el número es igual al número total de muestras dará loci que solamente compartan todas las muestras. Si es un valor bajo, como 4, dará una matriz más extendida, ya que arrojará valores donde este presente en al menos cuatro muestras de los datos.

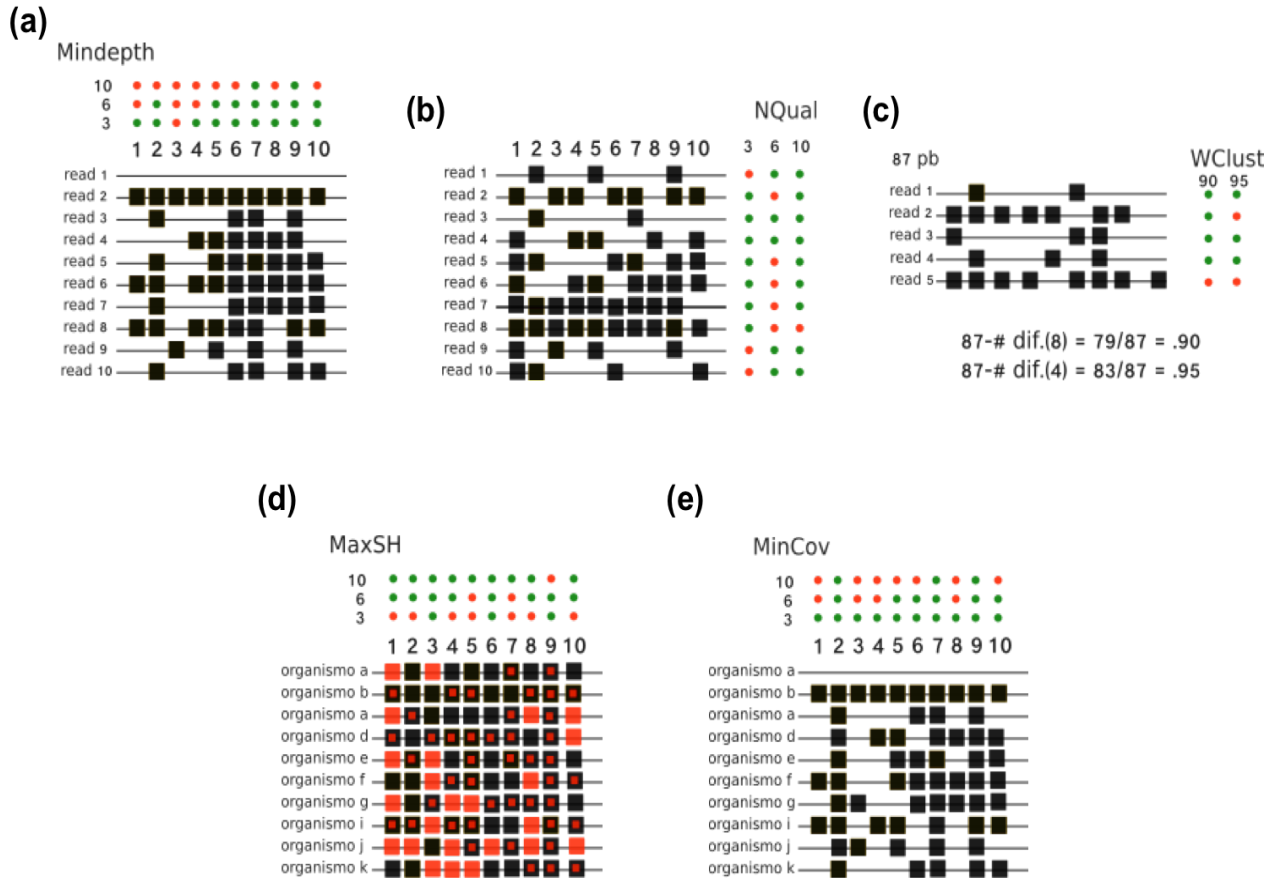


Figura 6. Filtrado y ensamblaje con parámetros *pyRAD*. (a) Mínimo de cobertura para un *cluster*. (b) Máximo número de sitios con calidad menor a 20. (c) Umbral de agrupamiento. (d) Máximo de individuos con sitios polimórficos. (e) Mínima cobertura por taxón. En verde se representa el SNP/*read* aceptado en el ensamblaje y en rojo los rechazados de acuerdo al valor asignado en cada parámetro.

Para el ensamblaje los siguientes parámetros clave (Tabla 4) fueron corridos con los valores especificados en los paréntesis: mínimo de cobertura para un *cluster* (-Mindepth- 8, 9, 10, 11, 12, 13, 14 y 15), máximo número de sitios con calidad menor a 20 (-NQual- 10, 8, 6, 4, 3, 2), umbral de agrupación (-WClust- .88, .90, .93, .95, .96, .97) y porcentaje máximo de individuos con sitios polimórficos compartidos en un locus (-MaxSH- 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%). El WClust se calculó con respecto al NQual. El MaxSH fue calculado para las 60 muestras totales como un porcentaje y usado en el procesamiento como el número entero resultante. Solamente

un parámetro fue variado a la vez, manteniendo los otros fijos de acuerdo a lo sugerido para los valores default de *pyRAD* (Tabla 4).

Ensamblado de novo y análisis de calidad de todas las muestras

Con los cuatro parámetros optimizados (-Mindepth- 12, -NQual- 6, -WClust- .95, y -MaxSH- 30%, ver discusión adelante) se generó el ensamblaje *de novo* y filtrado de los datos arrojados por la secuenciación ddRAD (Tabla 2, Individuos en ensamblado), utilizando tres tipos de compilaciones sin las muestras replicadas: (1) todas las muestras de la población *J. blancoi* y *J. monticola* (nombrada JbyJm con 192 muestras, ANEXO II), (2) solamente las muestras de *J. blancoi* (nombrada Jb con 62 muestras, ANEXO II) y (3) solamente las muestras de *J. monticola* (nombrada Jm con 130 muestras, ANEXO II).

Posteriormente para eliminar la mayor cantidad de *missing data* dentro las muestras se recurrió a modificar el porcentaje mínimo de SNPs presentes por taxón (“cobertura mínima por taxón”) incrementando el parámetro gradualmente cada 5% (-MinCov- de 5% a 85%) en cada compilación, procesando los archivos con los programas VCFtools v.0.1.14 (Danecek *et al.* 2011) para obtener los individuos con mayor cantidad de *missing data* y poder eliminarlos. Los formatos requeridos se obtuvieron con PLINK v.1.07 (Purcell *et al.* 2007). El porcentaje de MinCov se modificó de acuerdo al número de individuos por cada compilación. Sobre esos resultados se evaluaron tres aspectos: el número de SNPs recuperados, el porcentaje de *missing data* y el número de muestras eliminadas por poseer menos del 80% de los SNPs totales.

Con los parámetros de las muestras compiladas establecidos (-Mindepth-12, -NQual- 4, -WClust- .95, -MaxSH- 30% y -MinCov- 65%, ver discusión adelante), se volvió a correr el programa *pyRAD* para las tres compilaciones excluyendo las muestras que no cumplieran con al menos 20% de los SNPs totales.

Una vez que los SNPs pasaron los filtros de calidad de genotificación y llamado, se realizó un análisis exploratorio mediante un Análisis de Componentes Principales (PCA, por sus siglas en inglés) con el uso de la paquetería SNPRelate

v.1.2.0 (Zheng *et al.* 2012) de R versión 3.2.1 (R Development Core Team) para corroborar que la ubicación de las muestras en cada librería y *lane* no interviniera con la variación presente en la distribución de las muestras, ni con la de las especies.

Filtrado por SNPs ligados

Para minimizar el sesgo en la variación de frecuencias se descartaron los SNPs ligados, conservando sólo el primer SNP por cada RAD-locus (secuencia de 86 bases). Esto se realizó mediante scripts de R, PLINK y VCFtools. El efecto en la estructura poblacional se analizó mediante comparación de PCAs con datos incluyendo y excluyendo los SNPs ligados.

GENÓMICA DE POBLACIONES Y ANÁLISIS FILOGENÉTICO

Análisis de estructura poblacional por categorías MAF

Los alelos raros pueden contribuir a la diferenciación poblacional y adaptación local, pero también pueden ser producto de errores en la generación de datos. Para evaluar el efecto de analizar el universo completo de datos contra la exclusión de alelos raros, con los SNPs no ligados se aplicó un filtro de los SNPs con diferentes rangos de frecuencia: excluyendo (1) SNPs con MAF menor a 0.05, (2) SNPs con MAF menor a 0.1, (3) SNPs con MAF mayor a 0.05, (4) SNPs con MAF mayor a 0.1 y (5) considerando todos los SNPs. De esos filtros se realizaron los PCA correspondientes para determinar que filtro permitía diferenciar de mejor manera (mayor porcentaje de variación explicada por los primeros dos componentes) las poblaciones. Los análisis subsecuentes se realizaron con los cinco conjuntos de datos, pero sólo se presenta en resultados los que explican mejor la variación.

Estructura genética de la distribución de la variación

Para determinar la estructura poblacional y posibles patrones de flujo génico se procedió a analizar ambas especies con ADMIXTURE v.1.3 (Alexander *et al.* 2009), modelo que incluye un procedimiento de validación cruzada que permite identificar el valor de K (número de poblaciones genéticamente diferentes) para el cual el modelo tiene mejor exactitud de predicción, además de que este programa mejora la velocidad de procesamiento en comparación con STRUCTURE (Liu *et al.* 2013). La validación cruzada se calculó con las diferentes K de acuerdo al número total de poblaciones muestreadas más uno, por lo que para *J. blancoi* se utilizaron K's de 1-9 y para *J. monticola* K's de 1-14.

Identificación de SNPs asociados a categorías altitudinales

Las categorías altitudinales se consideraron como bajas desde los 1300-3000 msnm y altas a partir de los 3000 msnm los de *J. blancoi* y en *J. monticola* como bajas desde los 2900-3900 msnm y altas a partir de los 3900 msnm, debido a que las condiciones en el noroeste del país son diferentes a las de la FVTM (Tabla 1).

El índice de F_{ST} es una medida de la diferenciación genética entre subdivisiones de una población (Hedrick 2011), que puede utilizarse para identificar SNPs bajo selección (Excoffier *et al.* 2009) asociados a condiciones ambientales, en este estudio en particular altitud baja y alta, mediante métodos que consideren los *outliers* de los estimados F_{ST} . Para identificar los *outliers* de F_{ST} se obtuvieron los formatos de archivos requeridos con el programa PLINK y posteriormente el análisis de los F_{ST} se realizó con la paquetería snpStats v.1.15.4 (Clayton 2015) de R. Los individuos se agruparon por condición ambiental y por poblaciones, calculando su F_{ST} pareada. Los SNPs con F_{ST} alta se determinaron considerando el 5% de SNPs que mostraran mayor valor de F_{ST} entre el total de SNPs asociados a condiciones ambientales y a sus poblaciones.

Relaciones filogenéticas

El método Neighbor-Joining (NJ) permite agrupar a los individuos y generar dendrogramas considerando las distancias genéticas (Peña 2012). Al realizar un árbol con el método de NJ se espera que los individuos pertenecientes a una misma población formen grupos según sus similitudes poblacionales.

Para realizar los NJ se corrieron nuevamente los datos con *pyRAD* para añadir los grupos externos. En concreto, se utilizaron los individuos de las compilaciones finales (ver *Ensamblado de novo y análisis de calidad de todas las muestras*), añadiendo los grupos externos correspondientes a cada compilación: *J. flaccida*, *J. zanonii* y *J. deppeana* para Jm; *J. scopolorum* y *J. virginiana* para Jb; y todos los *outgroups* para JbyJm. Los valores óptimos de cada parámetro se modificaron de acuerdo al cambio de individuos en cada compilación. Con estas nuevas compilaciones se realizaron arboles NJ con la paquetería *phytools* v.0.5-38 (Revell 2012) de R.

RESULTADOS

ENSAMBLAJE DE NOVO Y FILTRADO DE DATOS RAD

Demultiplexeo

En general se muestra una distribución aleatoria en el número de *reads* por muestra, independiente de la especie o población. Los controles negativos presentan una cantidad de *reads* mínima en comparación a los *reads* del resto de las muestras (representan 1.4%, min= 841, max= 80,842, promedio= 16,644)

Al examinar el número de *reads* por muestra (min= 361, max= 5,888,556, promedio= 1,132,947), no se observa variación asociada a las especies, aunque algunos individuos de *J. blancoi* muestran considerablemente más *reads* que el resto de las muestras (Figura 7). Sin embargo es más importante la variación dependiendo de la librería donde fueron procesadas las muestras (Figura 8). Las librerías que obtuvieron mayor cantidad de *reads* promedio se presentaron en la librería 1 (2,437,127) seguida de la librería 6 (1,434,510).

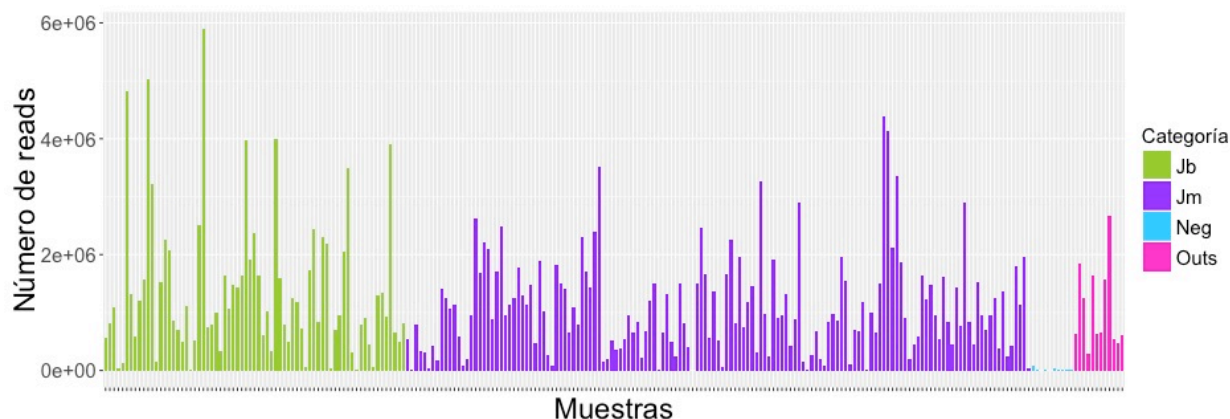


Figura 7. Número de *reads* obtenidos por especie; *J. blancoi* (71 muestras); *J. monticola* (147 muestras); negativos (10 muestras); y outgroups (12 muestras).

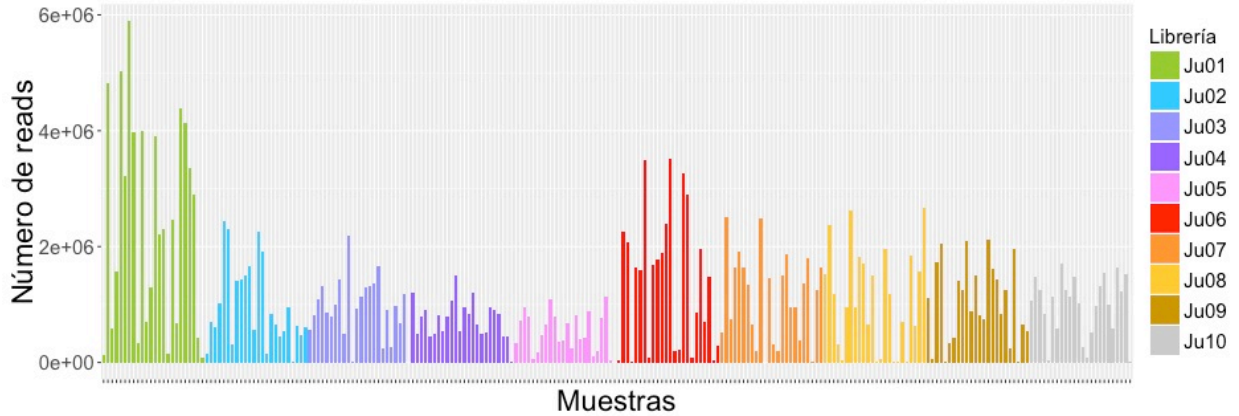


Figura 8. Número de reads obtenidos por cada librería. Diez librerías procesadas en 6 lanes (1, 2-3, 4-5, 6-7, 8-9 y 10).

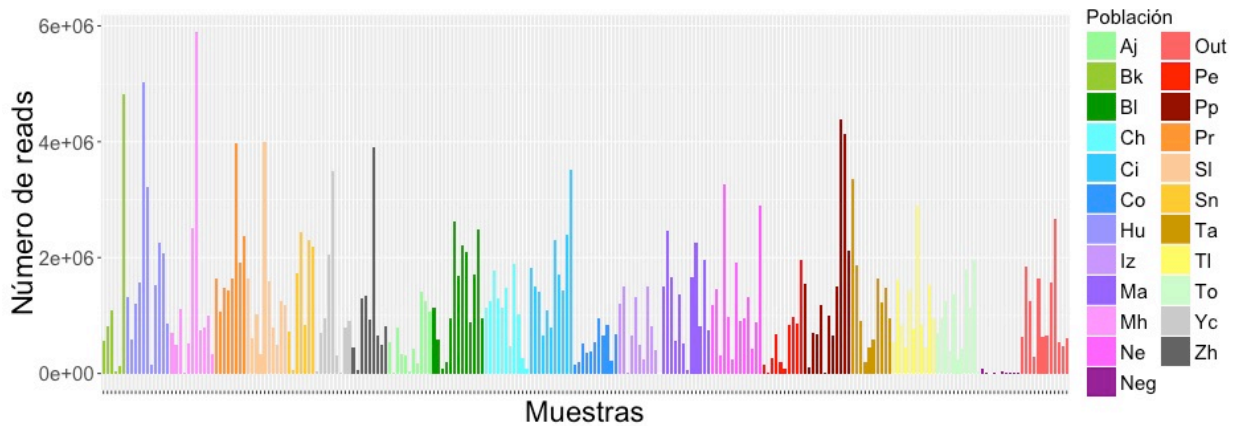


Figura 9. Número de reads obtenidos por cada población. *J. blancoi* (8 poblaciones), *J. monticola* (13 poblaciones), negativos y outgroups. Las claves de las poblaciones corresponden a los señalados en la Tabla 1.

Las muestras que obtuvieron menor cantidad de reads fueron las librerías 5 (521,396) y 4 (745,882). Las poblaciones con menor número de reads fueron el Nevado de Colima (499,067), Ajusco (580,517) y el Cofre de Perote (599,247); las poblaciones con mayor número de reads fueron La Preciosita (1,940,915), Huehuento (1,798,760) y Citlaltépetl (1,691,726) (Figura 9).

Optimización del ensamblado de novo y análisis de calidad

Se determinaron como parámetros optimizados aquellos donde se aumento el número de SNPs, reduciendo el *missing data* y considerando la tasa de error entre los replicados (Mindepth 12, NQual 4, WClust 0.95, MaxSH 30% y MinCov 65%). Al incrementar el valor de los parámetros que afectan el mínimo de cobertura para un *cluster* (Mindepth, 7-15; Figura 10) y el umbral de agrupamiento (WClust, 88-97; Figura 11), la tasa de error en los SNPs así como en el número de SNPs y *missing data* disminuyeron (Tabla 5 y Tabla 6). Al incrementar el número máximo de individuos con sitios polimórficos (MaxSH, 10-90), disminuyó la tasa de error y el porcentaje de *missing data*, sin embargo el número de SNPs incrementó conforme al valor de dicho parámetro (Figura 12, Tabla 7).

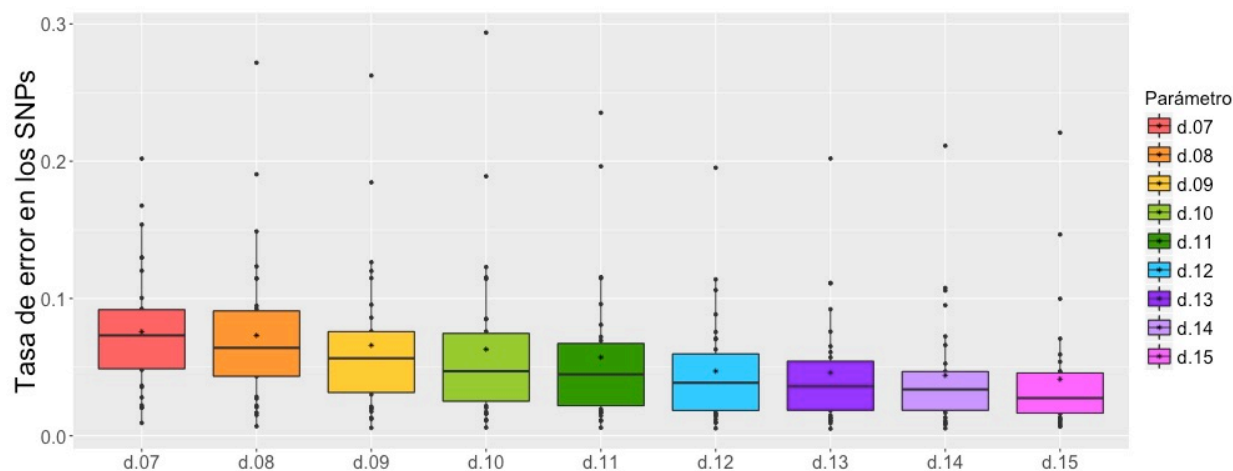


Figura 10. Efecto de variar el mínimo de *reads* para hacer un llamado de base (Mindepth) en *pyRAD* sobre la tasa de error en las muestras replicadas. Con mayor Mindepth disminuye la tasa de error. A partir de valores mayores a 13 la dispersión de los datos comienza a incrementar, por lo cual se utiliza 12 como el valor óptimo. * indica la media de los datos y la línea dentro de la caja la mediana.

Aquellos parámetros que brindaron el mayor número de SNPs y el menor porcentaje de *missing data* se utilizaron para identificar el valor óptimo para cada parámetro: el mínimo de cobertura para un *cluster* (Mindepth) se estableció en 12, el umbral de agrupamiento (WClust) en 0.95, en concordancia con el máximo número de sitios con calidad menor a 20 (NQual) en 4 y un número máximo de individuos con sitios polimórficos (MaxSH) en un valor correspondiente al 30% para todos los datos.

Tabla 5. Valores obtenidos al modificar mínimo de cobertura para un *cluster* (Mindepth) en *pyRAD* sobre las muestras replicadas.

Mindepth	D7	D8	D9	D10	D11	D12	D13	D14	D15
Missing data	94,403 80.9 %	76,683 79.0 %	61,807 77.1 %	50856 75.2 %	42,205 73.5 %	35,178 71.5 %	29,190 69.4 %	25,103 66.8 %	21,558 65.1 %
# de SNPs	116,624	96,956	80,141	67,562	57,372	49,155	42,035	37,577	33,072

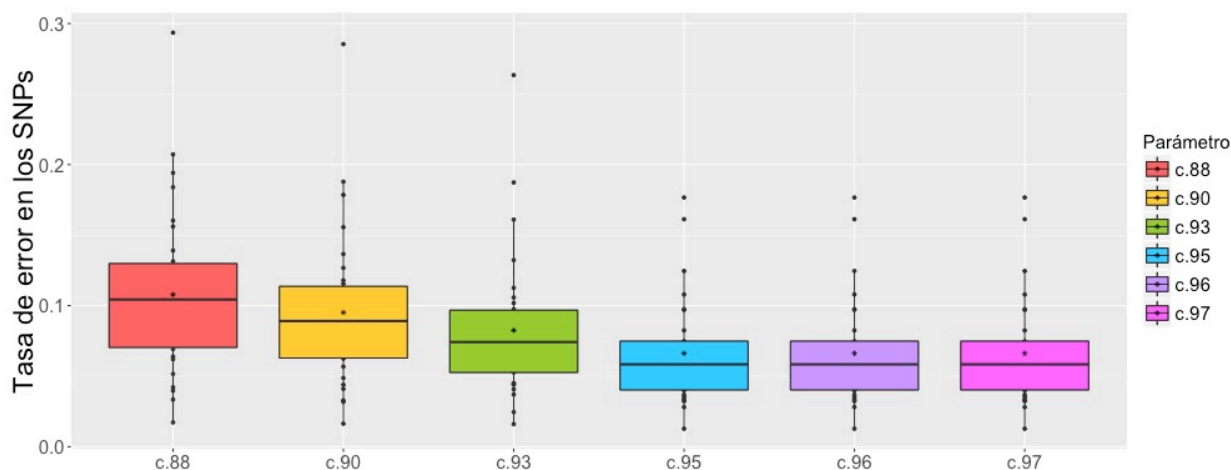


Figura 11. Efecto de variar el umbral de agrupamiento (WClust) en concordancia con el máximo número de sitios con calidad menor a 20 (NQual) en *pyRAD* sobre las muestras replicadas. Incrementar los valores de WClust permite menor número de sitios con calidad menor a 20 (NQual), esto disminuye la tasa de error ya que se excluyen *reads* con muchos sitios indeterminados. Se consideró a 0.95 como el valor óptimo debido a que permitió algunos sitios indeterminados (4) y mediante el cual se obtuvo menor tasa de error. * indica la media de los datos y la línea dentro de la caja la mediana.

Tabla 6. Valores obtenidos al modificar el máximo número de sitios con calidad menor a 20 (NQual) en concordancia con el umbral de agrupamiento (WClust) en *pyRAD* sobre las muestras replicadas.

NQual y WClust	C87	C90	C93	C95	C96	C97
Missing data	257,358	148,568	105,943	65,639	61,158	65,639
	85.64%	84.06%	83.01%	81.56%	80.49%	79.61%
# de SNPs	319,217	176,734	105,943	80,475	61,158	42,357

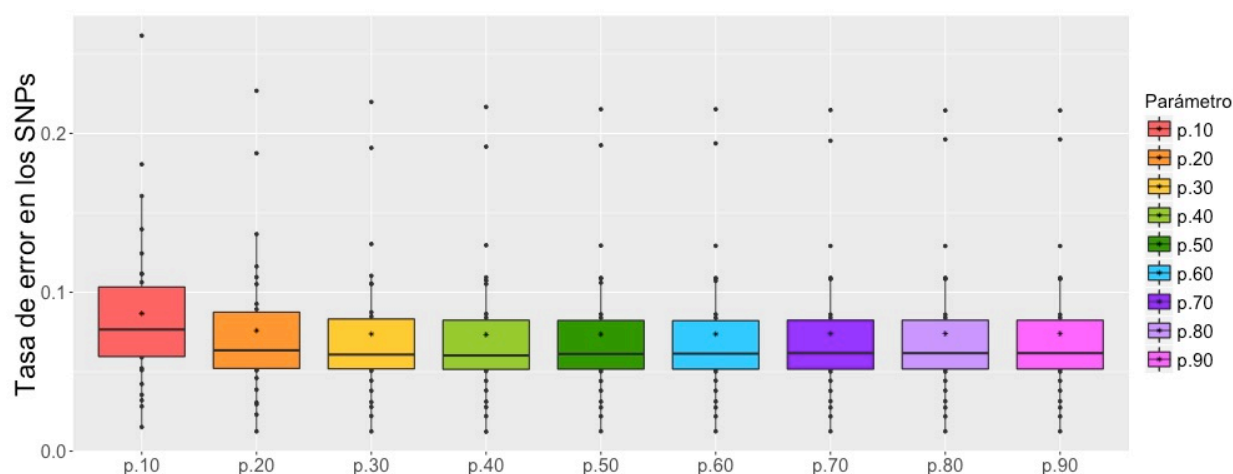


Figura 12. Efecto de variar el número máximo de individuos con sitios polimórficos (MaxSH) en *pyRAD* sobre las muestras replicadas. Aumentar el MaxSH disminuye la tasa de error entre las muestras. La tasa de error al aumentar el MaxSH se mantiene similar a partir del valor 30, por lo cual se utilizó como valor óptimo. * indica la media de los datos y la línea dentro de la caja la mediana.

Tabla 7. Valores obtenidos al modificar el número máximo de individuos con sitios polimórficos (MaxSH) en *pyRAD* sobre las muestras replicadas.

MaxSH	P10	P20	P30	P40	P50	P60	P70	P80	P90
Missing data	151,592	153,538	153,999	154,076	154,136	154,136	154,136	154,136	154,136
	83.75%	82.32%	81.83%	81.74%	81.7%	81.69%	81.67%	81.67%	81.67%
# de SNPs	181,007	186,518	188,190	188,505	188,659	188,694	188,728	188,739	188,739

Tras las pruebas con el parámetro que afecta la mínima cobertura por taxón (MinCov) el número de SNPs recuperados fue mayor en los conjuntos Jm y JbyJm (Figura 13); el porcentaje de *missing data* fue menor en el conjunto Jm (Figura 14); y el número de muestras eliminadas por poseer más del 80% de *missing data* fue menor en los conjuntos Jb y Jm que en el JbyJm (conjunto con ambas especies) (Figura 15).

Se eligió el 65% como valor final para las tres compilaciones, debido a que con este valor se obtuvo menor cantidad de *missing data* (23.8%, 24.4% y 23%; Figura 14), menor número de muestras descartadas (19, 7 y 14; Figura 15) y mayor recuperación de SNPs (7,757, 4,250 y 7,913; Figura 13) para las compilaciones JbyJm, Jb y Jm, respectivamente. De acuerdo a este porcentaje y eliminando las muestras correspondientes (Figura 15), el parámetro MinCov (Mínimo número de muestras para considerar un locus) del conjunto JbyJm se estableció en 112 considerando 173 muestras; para el conjunto Jb se determinó en 36 para 55 muestras; y en el conjunto Jm se fijó en 75 considerando 116 muestras.

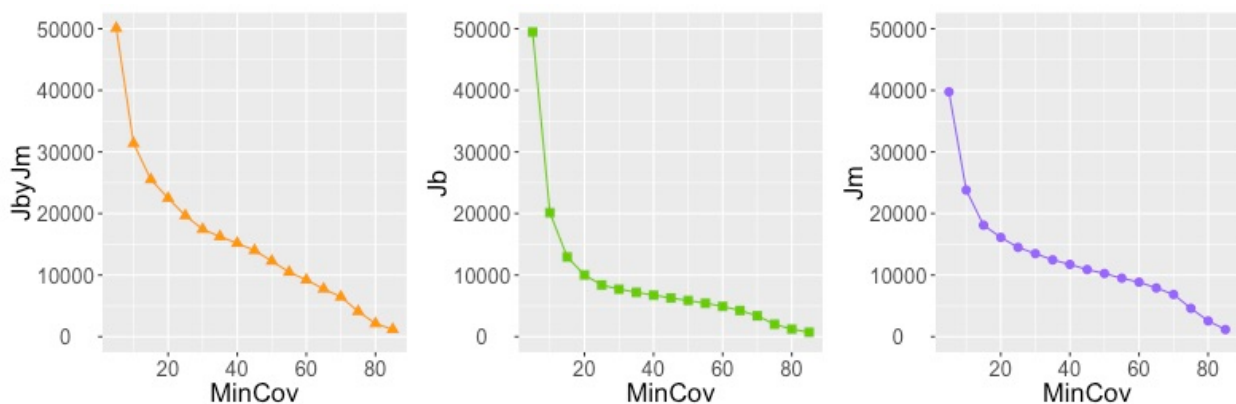


Figura 13. Número de SNPs obtenidos en cada conjunto.

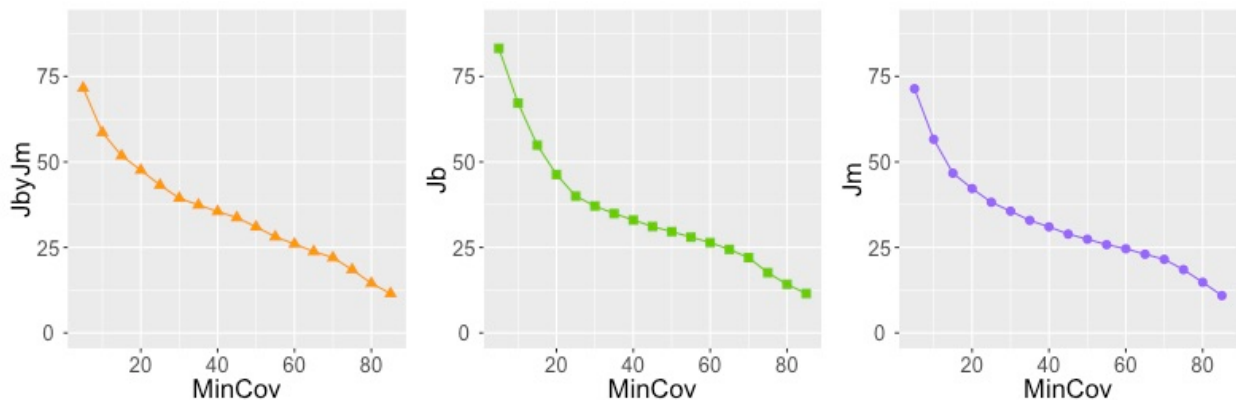


Figura 14. Porcentaje de *missing data* obtenido en cada conjunto.

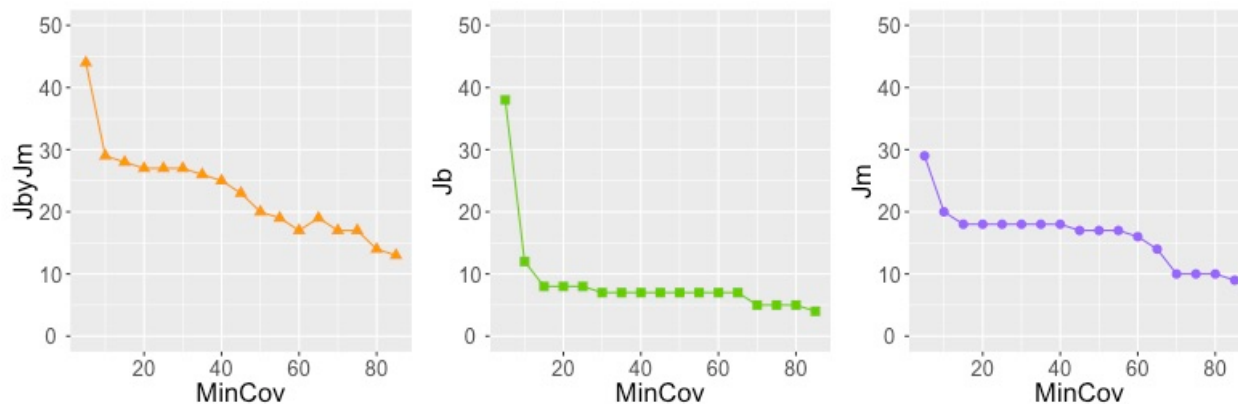


Figura 15. Número de muestras eliminadas en cada conjunto por poseer más del 80% de *missing data* en los SNPs totales.

Tabla 8. Valores obtenidos antes y después de eliminar las muestras con >80% de *missing data* en los SNPs totales. Ajustando el parámetro MinCov en los tres conjuntos de datos.

	Conjunto JbyJm	Conjunto Jb	Conjunto Jm
MinCov	125	40	85
MinCov descartando muestras >80% de <i>missing data</i>.	112	36	75
# SNPs por muestra	7,757	4,250	7,913
# SNPs por muestra descartando muestras >80% de <i>missing data</i>.	9,382	4,973	8,628
Missing data	23.8 %	24.4 %	23%
Missing data descartando muestras >80% de <i>missing data</i>.	18.9%	18.4%	17%
Número de muestras eliminadas (>80% de <i>missing data</i> por muestra)	19	7	14

Los datos finales que se ocuparon para los análisis exploratorios y genómicos después de eliminar las muestras >80% de *missing data* consisten en 9,382 SNPs en la compilación de JbyJm (173 muestras); 4,973 SNPs en Jb (55 muestras); y 8,628 SNPs en Jm (116 muestras). Con un porcentaje de *missing data* de 18.9%, 18.4% y 17% respectivamente (Tabla 8).

El número de poblaciones por especie se mantuvo a lo largo del ensamblaje (ocho en *J. blancoi* y trece en *J. monticola*, Tabla 2) con un promedio de siete y nueve individuos por población en *J. blancoi* y *J. monticola* respectivamente. Se eliminaron máximo tres individuos por población (Cofre de Perote, Tabla 2) y las poblaciones se mantuvieron con al menos cuatro individuos (Presa Brokman, Tabla 2) y máximo diez individuos en *J. blancoi* y *J. monticola* ver Tabla 2.

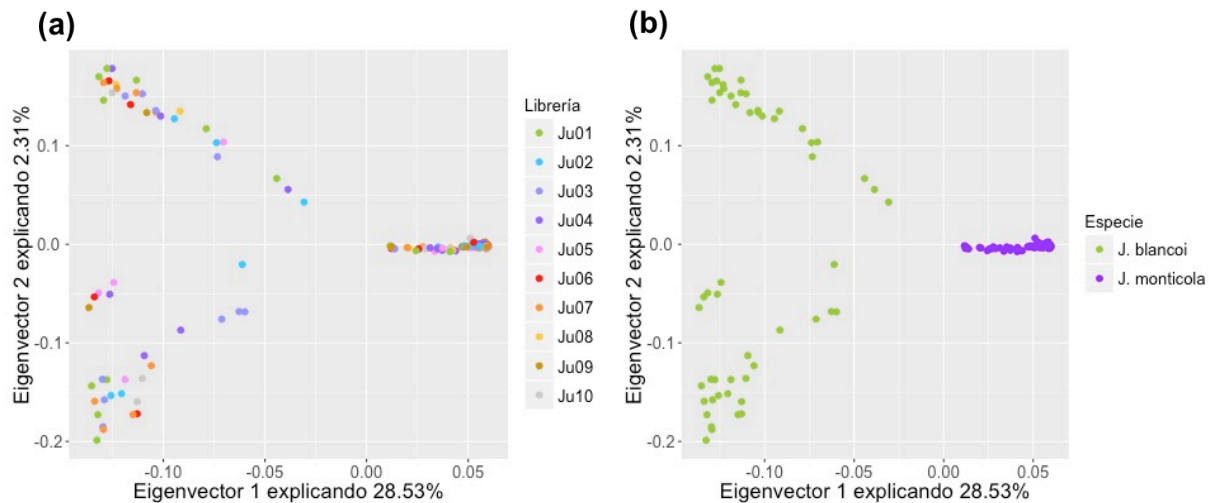


Figura 16. PCA con los dos componentes principales, a partir del conjunto JbyJm coloreados según la librería en la que fueron secuenciadas las muestras. (a) Diez librerías procesadas en 6 lanes (1, 2-3, 4-5, 6-7, 8-9 y 10). (b) Individuos diferenciados por especie.

En el análisis exploratorio con el PCA las muestras finales se distribuyen independientemente de la librería y *lane* en la que fueron procesadas (Figura 16a) y se agrupan en dos grandes conjuntos separados, uno (verde en Figura 16b) en donde se ubican todas las muestras de las poblaciones de la especie *J. blancoi* y otro (morado en Figura 16b) al que pertenecen las muestras de las poblaciones de la especie *J.*

monticola. En estos análisis ~28% de variación está explicada por el primer componente principal, que es el que separa las muestras de ambas especies.

Filtrado por SNPs ligados

Los PCA de cada especie y filtros probados muestran que existe una estructura poblacional y que el porcentaje de la variación explicada por dicha estructura varía dependiendo del filtro utilizado.

En el PCA del conjunto Jb se observa que existen tres grupos separados entre sí asociados al primer componente que explica 10.63% de la variación; y al segundo componente que explica el 7.44% de la variación (Figura 17a). En el conjunto Jm el primer componente del PCA explica el 3.16% de la variación y el segundo un 2.7% de la variación (Figura 18a). En ambas especies, al descartar los SNPs considerados como ligados (2,705 en Jb y 5,752 en Jm, ver métodos) el porcentaje de variación asociada a cada componente principal se incrementó entre el 0.62% y 0.10%, para el primer y segundo componente en Jb (Figura 17b), y entre el 0.06 % y 0.21 % para el primer y segundo componente, respectivamente en Jm (Figura 18b).

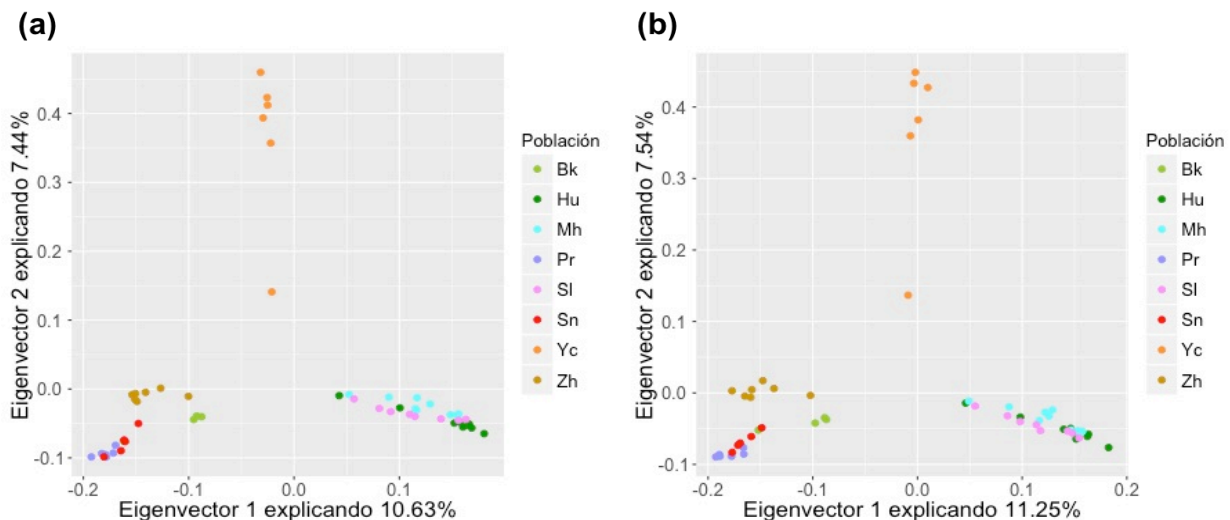


Figura 17. PCA con los dos componentes principales, a partir del conjunto Jb. (a) Considerando SNPs ligados. (b) Descartando SNPs ligados. Códigos de poblaciones como en Tabla 1.

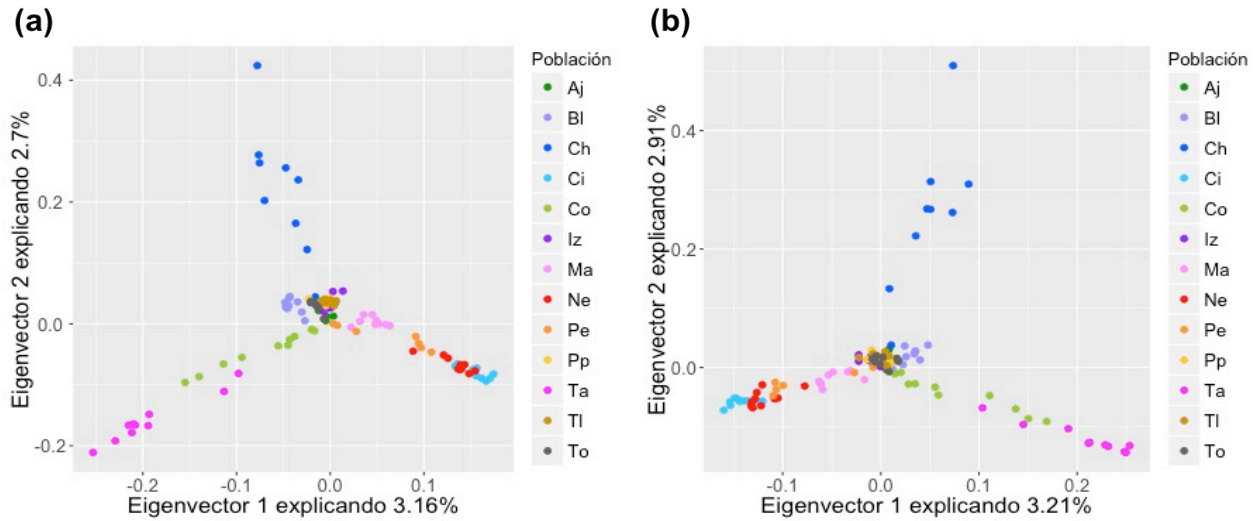


Figura 18. PCA con los dos componentes principales, a partir del conjunto Jm. (a) Considerando SNPs ligados. (b) Descartando SNPs ligados. Códigos de poblaciones como en Tabla 1.

GENÓMICA DE POBLACIONES Y ANÁLISIS FILOGENÉTICOS

Análisis de estructura poblacional por categorías MAF

El espectro de frecuencia permitió observar la distribución de alelos y su efecto sobre la estructura poblacional. Esto permitió elegir el filtro MAF más adecuado considerando los SNPs conservados y descartados en cada filtro, así como sus porcentajes de variación para obtener mayor soporte a la estructura genética propuesta.

En el espectro de frecuencias de *J. blancoi* (Figura 19a) y *J. monticola* (Figura 19b) se muestra que hay mayor cantidad de SNPs con frecuencias menores o iguales a 0.05 y más aún con frecuencias menores o iguales a 0.1 con respecto a los demás filtros.

Los alelos raros suelen clasificarse como aquellos con una frecuencia ≤ 0.05 ó ≤ 0.1 . Al observar la distribución del porcentaje de variación, las $MAF \leq 0.1$ agruparon de manera separada a una población de cada especie de *Juniperus* (Yécora y el Chico, para *J. blancoi* y *J. monticola*, respectivamente; Figura 20), que no muestra esta

distribución cuando se consideran frecuencias comunes (>0.05 y >0.1). En los PCAs con los SNPs $MAF \leq 0.05$ también se diferenció la misma población de cada especie, pero de forma más clara.

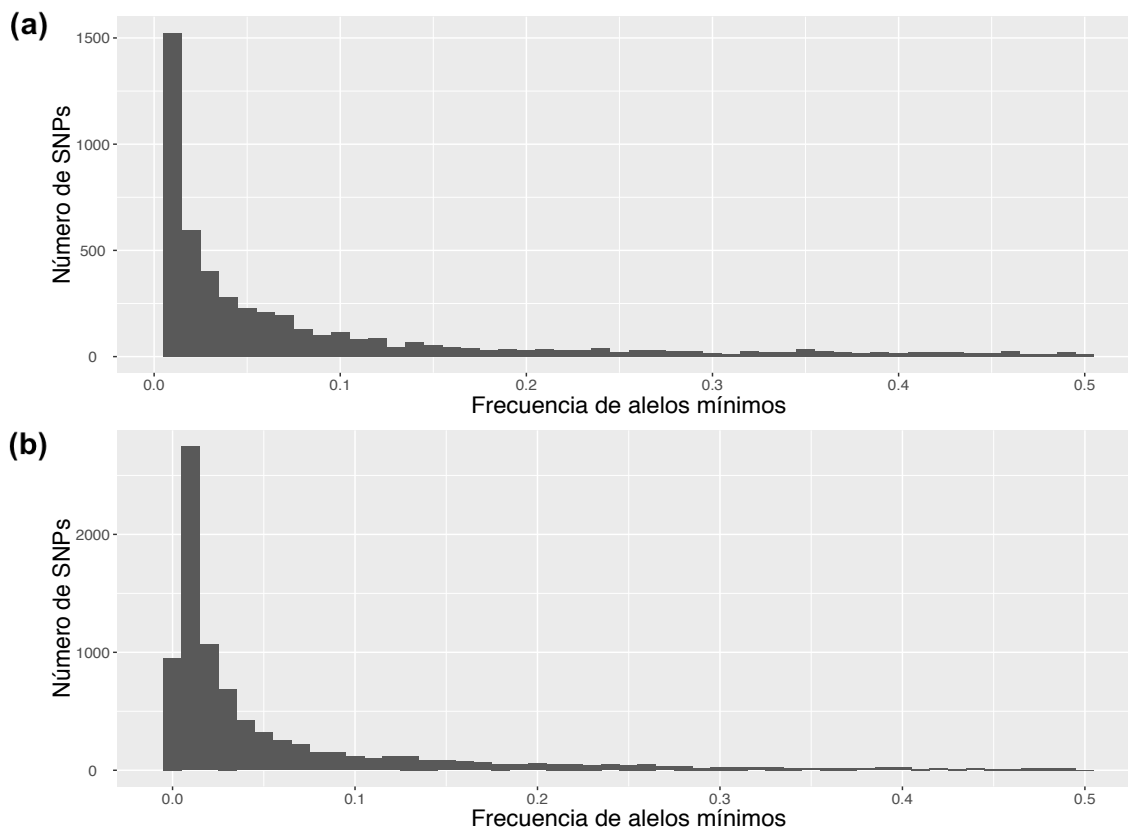


Figura 19. Espectro de frecuencias. Arriba, frecuencias de *J. blancoi* y abajo, frecuencias de *J. monticola*.

Los filtros $MAF > 0.05$ y >0.1 descartan los alelos raros. En ambas especies el porcentaje de variación explicada por el primer componente incremento utilizando el filtro $MAF > 0.1$ (7.7% más para Jb y 0.59% para Jm, con relación a la variación de los SNPs sin algún MAF).

Es interesante destacar que en ambas especies al utilizar los datos ≤ 0.1 el primer componente discriminó a una sola población (Yécora en Jb y a El Chico en Jm). Sin embargo al descartar frecuencias alélicas ≤ 0.1 , la distribución de variación en la

población de Rio Patitos se acercó a Yécora en Jb (Figura 21) y la población de El Chico se aproximó al grupo central del primer componente en Jm (Figura 22).

En el conjunto Jb, el filtro $MAF > 0.1$ permite que el porcentaje de variación en el primer y segundo componente sea de 28.21% y un 8.63%, respectivamente (Figura 21). La distribución de la variación fue similar a la obtenida anteriormente (Figura 17b), pero uno de los grupos se subdividió.

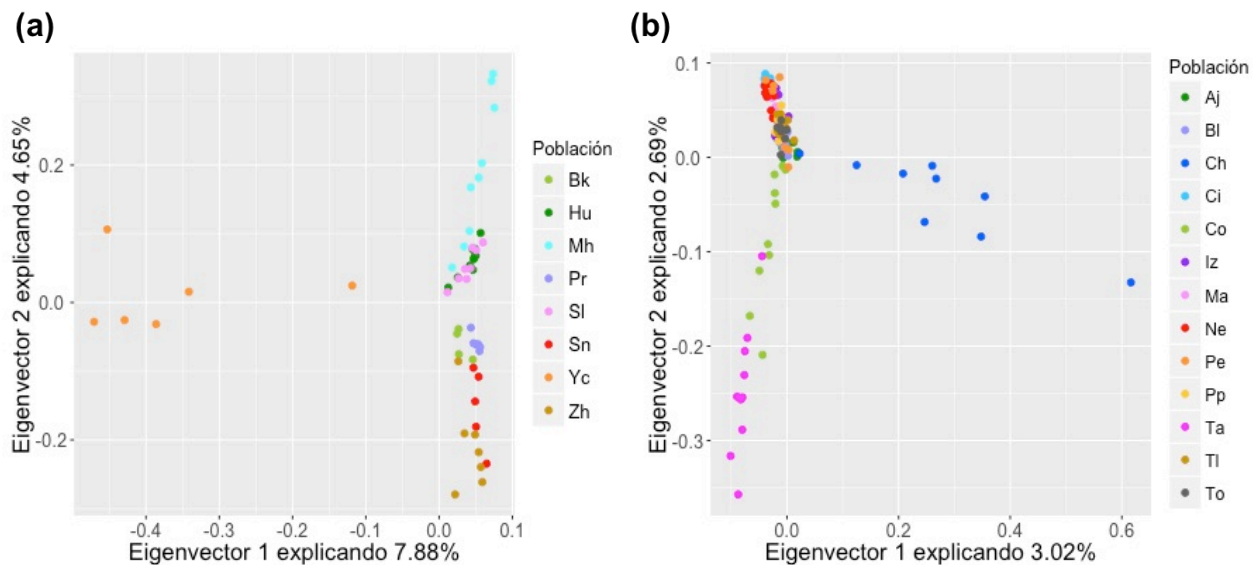


Figura 20. PCA con los dos primeros componentes principales. (a) Conjunto Jb con $MAF \leq 0.1$. (b) Conjunto Jm con $MAF \leq 0.1$. Códigos de poblaciones como en Tabla 1. Nótese que en Jb la población Yécora (Yc) y en Jm la población El Chico (Ch) se diferencian del resto.

Similar a la compilación Jb, en los datos Jm con el filtro $MAF > 0.1$ se incrementó el porcentaje de variación explicada, llegando a 6.53% y 4.25% para el primer y segundo componente, respectivamente (Figura 22). La distribución de la variación se modificó de la agrupación anterior de Jm (Figura 18b) con la inclusión de una población al grupo central. Para los análisis de estructura genética se eligió utilizar las frecuencias > 0.05 y > 0.1 , ya que pese a que disminuyeron el número de SNPs, arrojaron un mayor porcentaje de variación explicada por los primeros dos componentes del PCA, lo que permite visualizar una mayor estructura poblacional dentro de cada especie.

Finalmente se optó por elegir $MAF > 0.1$ ya que junto a los PCA (Figura 21 y 22), vislumbra de mejor manera la estructura genética de los datos.

Estructura genética de la distribución de la variación.

A partir de los PCA anteriores y del posterior análisis ADMIXTURE se encuentra que ambas especies presentan una estructura poblacional asociada geográficamente. Con el análisis de ADMIXTURE se encontró una $K=4$ para *J. blancoi* y $K=3$ para *J. monticola*, independientemente del filtrado MAF utilizado (>0.05 ó >0.1). En *J. blancoi* dicha estructura es más marcada que en *J. monticola* (Figura 23).

En *Juniperus blancoi* no se observa un patrón que divida a las poblaciones por latitud o provincia biogeográfica, sin embargo las poblaciones se agrupan de forma similar a como se encontró en estudios previos (Mastretta-Yanes *et al.* 2012; Moreno-Letelier *et al.* 2014) con datos de cloroplasto y pocos loci nucleares. En concreto se encontró que las poblaciones de Presa Brokman, La preciosita y San José del Rincón se agrupan juntas, al igual que Yécora y Río Patitos. La diferencia con los resultados presentados aquí es que Huehuento forma un grupo distinto de Mohinora y El Salto.

En el conjunto Jm se ve un gradiente en las poblaciones según su ubicación geográfica oeste (Nevado de Colima y Tancítaro), centro (Cerro Blanco, Nevado de Toluca, Ajusco, Tlálloc, Iztaccíhuatl, Popocatépetl, La Malinche) y este (La Negra, Citlaltépetl, Cofre de Perote) dentro de la FVTM (Figura 22).

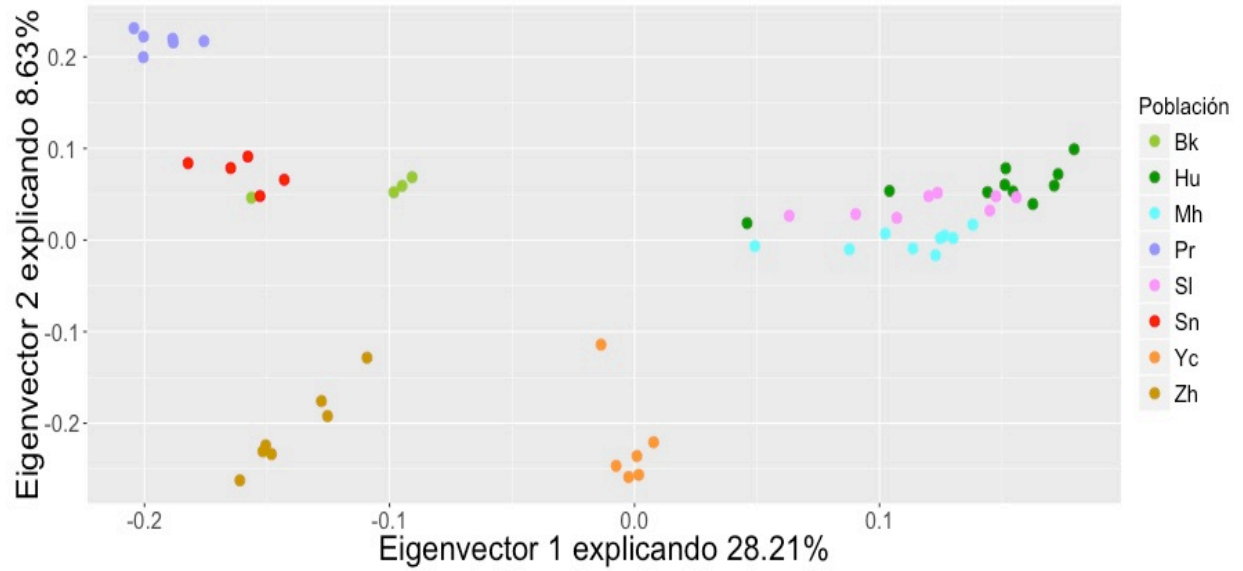


Figura 21. PCA con los dos componentes principales a partir del conjunto Jb, descartando SNPs ligados y utilizando MAF >0.1. Códigos de poblaciones como en Tabla 1.

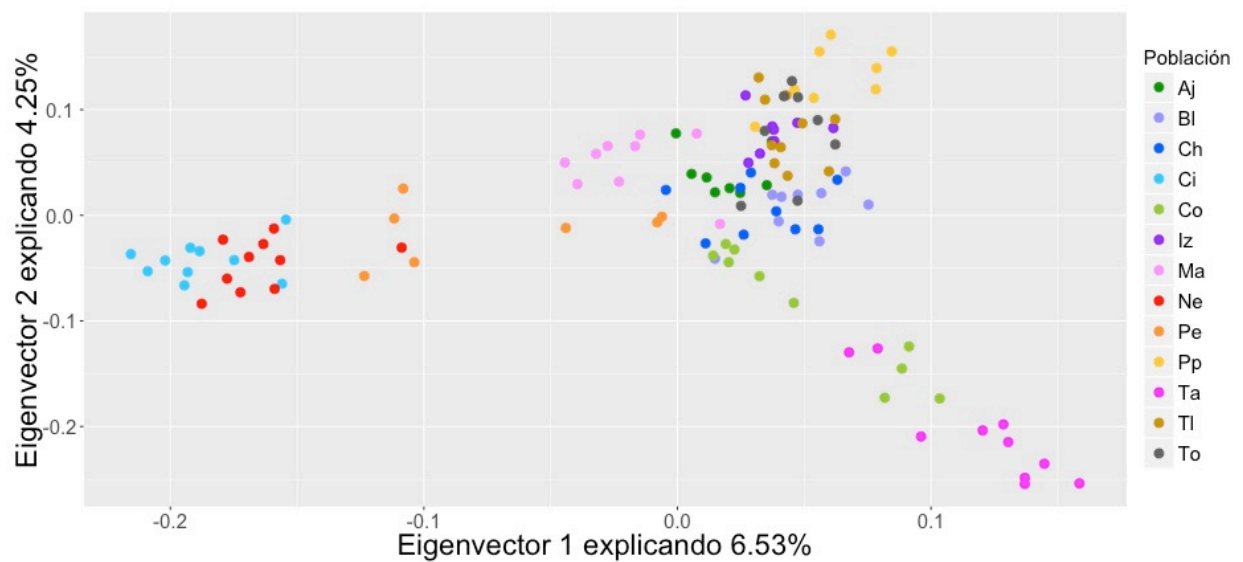


Figura 22. PCA con los dos componentes principales a partir del conjunto Jm, descartando SNPs ligados y utilizando MAF >0.1. Códigos de poblaciones como en Tabla 1.

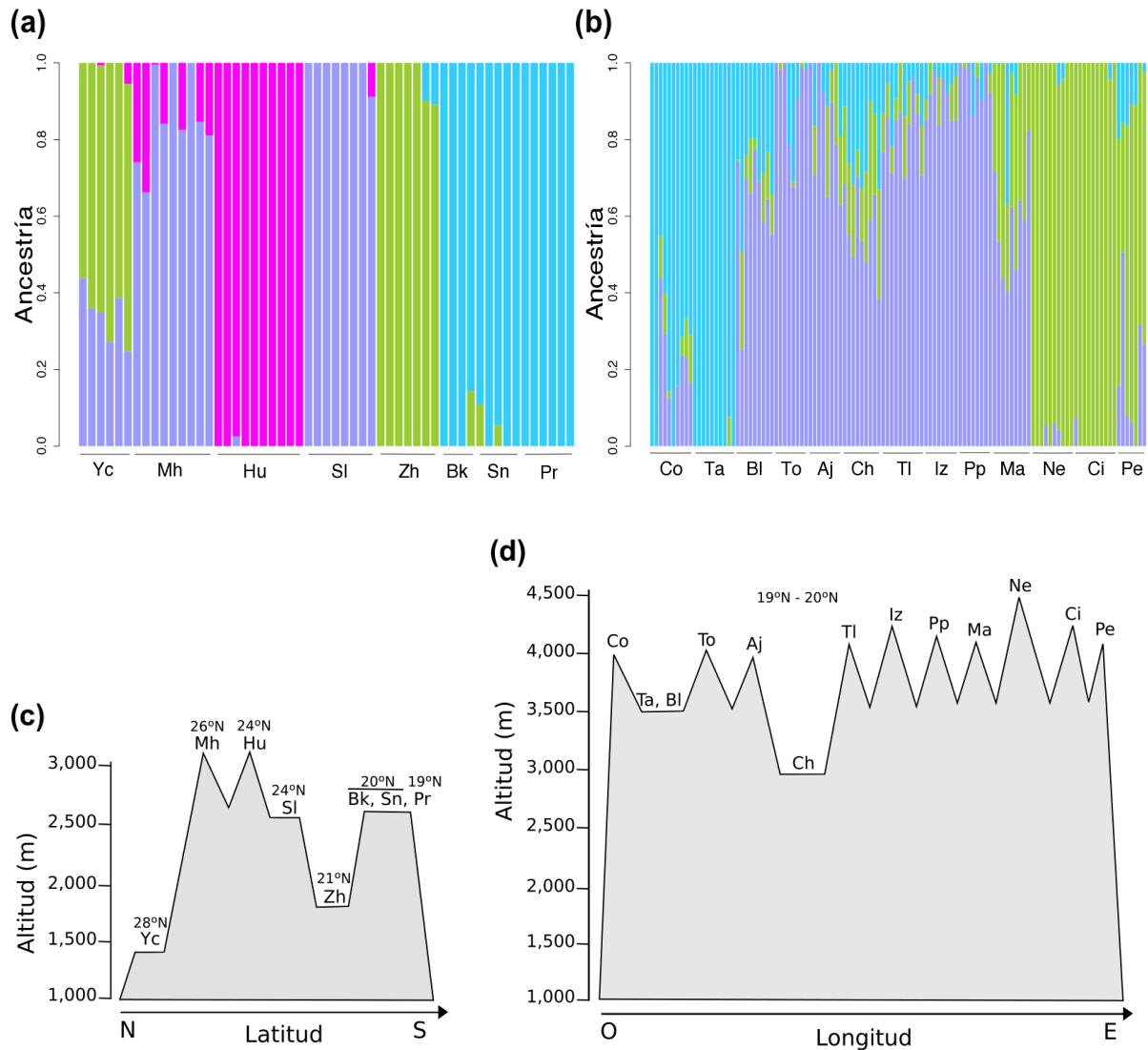


Figura 23. Agrupamiento genético de las poblaciones con los datos filtrando $MAF < 0.1$ para (a) poblaciones de *J. blancoi* acomodadas latitudinalmente de norte (lado izquierdo) a sur (lado derecho); y (b) poblaciones de *J. monticola* acomodadas por longitud de oeste (lado izquierdo) a este (lado derecho). Perfil altitudinal de las poblaciones de (c) *J. blancoi* (modificado de Moreno *et al.*, 2014) y (d) *J. monticola*.

Identificación de SNPs asociados a categorías altitudinales

Las categorías altitudinales se consideraron como bajas desde los 1300-3000 msnm y altas a partir de los 3000 msnm los de *J. blancoi* y en *J. monticola* como bajas desde los 2900-3900 msnm y altas a partir de los 3900 msnm, debido a que las condiciones en el noroeste del país son diferentes a las de la FVTM (Tabla 1). Se identificaron 548 SNPs en *J. blancoi* y 501 SNPs en *J. monticola*. Se determinaron como SNPs informativos el 5% con F_{ST} alta.

Considerando el 5% de los SNPs con F_{ST} más elevada, 28 y 25 SNPs se asocian a categorías altitudinales (Figura 24); mientras que 28 y 26 SNPs se asocian a las diferentes poblaciones (Figura 25) de *J. blancoi* y *J. monticola* respectivamente. Ningún SNP se comparte en las que diferencian categorías altitudinales entre especies. Sin embargo al comparar SNPs con F_{ST} alta de categorías altitudinales y de poblaciones se detectan cuatro SNPs compartidos dentro de cada especie (Tabla 9). Dado que no se cuenta con un genoma de referencia y que estas secuencias son de 86 pb resulta difícil interpretar los resultados de una búsqueda en BLAST.

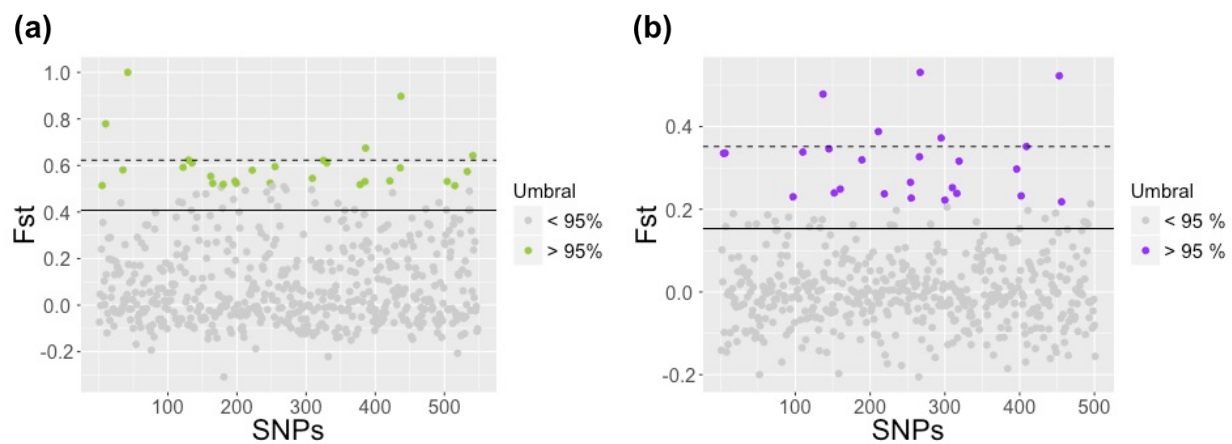


Figura 24. F_{ST} de los SNPs con $MAF > 0.1$. Línea continua representa el 10% de los SNPs con la mayor F_{ST} , línea punteada el 1% y en colores el 5% de los SNPs con F_{ST} más alta, pareando la categoría altitudinal dentro de cada especie. (a) SNPs de *J. blancoi*. (b) SNPs de *J. monticola*.

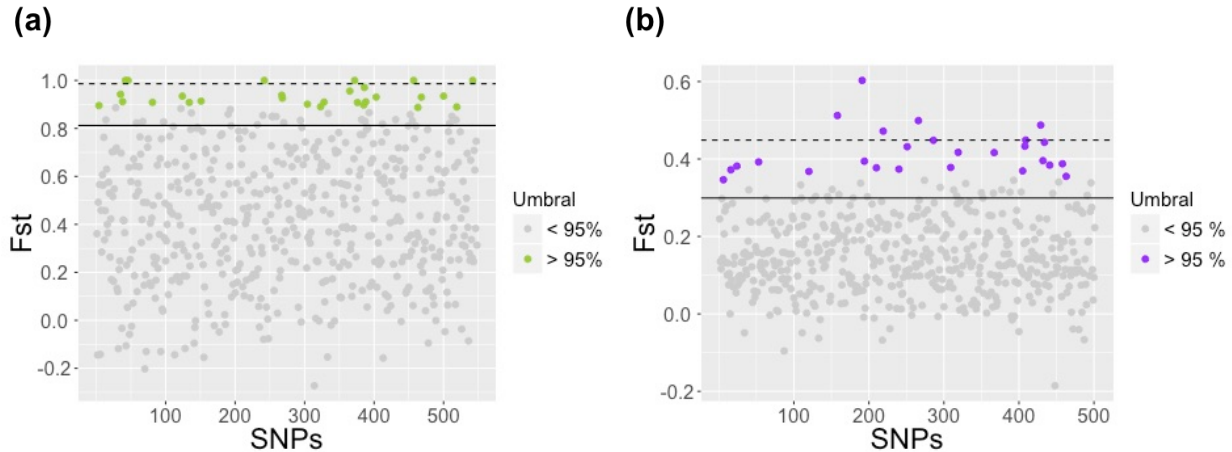


Figura 25. F_{ST} de los SNPs con $MAF > 0.1$. Línea continua representa el 10% de los SNPs más altos, línea punteada el 1% y en colores el 5% de los SNPs con F_{ST} más alta, pareando las poblaciones dentro de cada especie. (a) SNPs de *J. blancoi*. (b) SNPs de *J. monticola*.

Relaciones filogenéticas

Los dendrogramas construidos por NJ (Figura 26b y 27b) recuperan la misma topografía que la filogenia construida por Mao *et al.* (2010).

Dentro de *J. blancoi* los NJ distinguen tres clados donde se agrupan las ocho poblaciones. El primer clado contiene las poblaciones de Huehuento, Mohinora y El Salto. Yécora conforma un solo clado. Finalmente individuos de San José Rincón, Presa Brokman, La Preciosita y Río Patitos forman otro clado (Figura 26a).

En *J. monticola* la longitud e las ramas es menor en comparación a *J. blancoi*. El primer nodo agrupa a Tancítaro y el Nevado de Colima en una rama y en la otra puede diferenciarse El Chico de todas las demás poblaciones. Además se observa que los individuos de Citlaltépetl y La Negra se agrupan juntos (Figura 27).

Los grupos externos, *J. deppeana* (1) *J. flaccida* (6), *J. zanonii* (2), se encontraron cercanos al agrupamiento de *J. monticola* (Figura 26a) y las especies de *J. virginiana* (1) y *J. scopolorum* (2), se ubicaron más próximas a *J. blancoi* (Figura 27a) lo cual coincide con la proximidad presente en el árbol filogenético desarrollado por (Adams & Schwarbach 2013).

Tabla 9. Ubicación (número de RAD-loci y SNP) de SNPs con F_{ST} alta en la división por categoría altitudinal o población. En negritas se indican los SNPs compartidos al considerar poblaciones y hábitats en *J. blancoi* y *J. monticola*.

Categoría altitudinal		Población	
<i>J. blancoi</i>	<i>J. monticola</i>	<i>J. blancoi</i>	<i>J. monticola</i>
68:53	258:47	242:13	179:04
206:31	701:29	286:61	243:28
233:16	719:10	350:70	265:40
356:76	782:13	478:15	282:11
587:34	843:32	518:60	320:63
596:07	884:09	596:07	538:16
955:53	1073:30	620:45	648:58
986:42	1421:15	918:03	777:54
1067:03	1432:03	1000:30	1030:23
1162:44	1617:62	1162:44	1078:31
1407:13	1632:18	1197:40	1191:55
1470:24	1692:12	1286:32	1632:18
1490:83	1693:03	1649:77	1677:49
1717:21	1807:13	1732:13	1685:06
1885:03	1933:17	1836:83	1688:13
1893:13	1971:49	1892:33	1693:03
1934:06	1977:23	1934:06	1799:48
2124:44	2015:60	2124:44	1805:56
2322:81	2112:23	2139:06	1854:17
2354:17	2130:31	2157:57	1971:49
2356:24	2477:32	2211:07	2146:21
2472:50	2555:24	2244:09	2336:64
2576:70	2823:45	2350:25	2420:04
2778:15	2870:24	2376:41	2477:32
2786:12	3138:11	2425:25	2555:24
2874:32		2523:47	2658:15
2905:18		2536:23	
2951:18		2728:41	

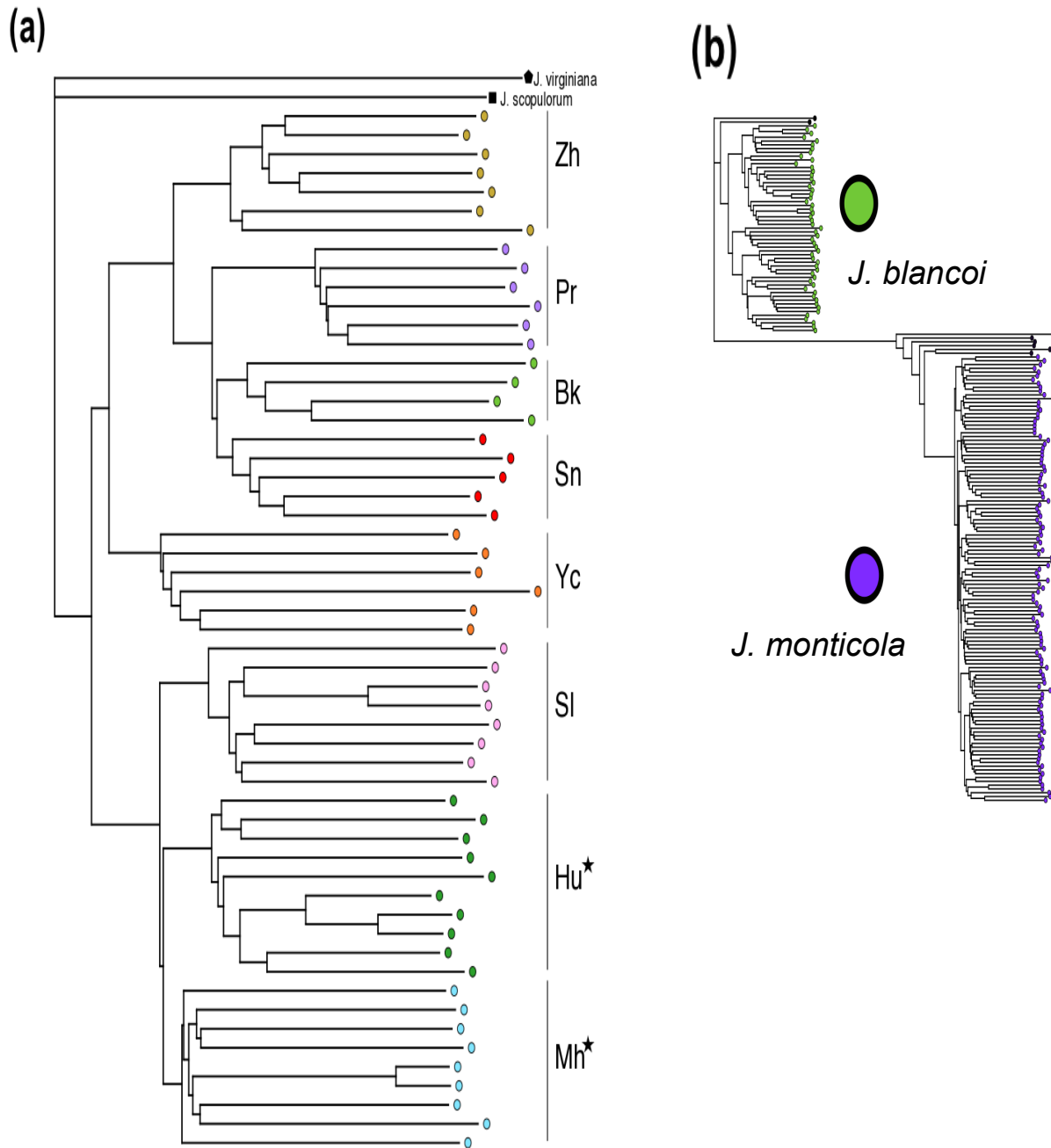


Figura 26. (a) Árbol NJ de 55 individuos de *J. blancoi*. Cada población es representada en un color diferente, en negro se representan los grupos externos cercanos a *J. blancoi*. La raíz es *J. virginiana*. (b) Árbol NJ de ambas especies. La raíz es *J. virginiana*.

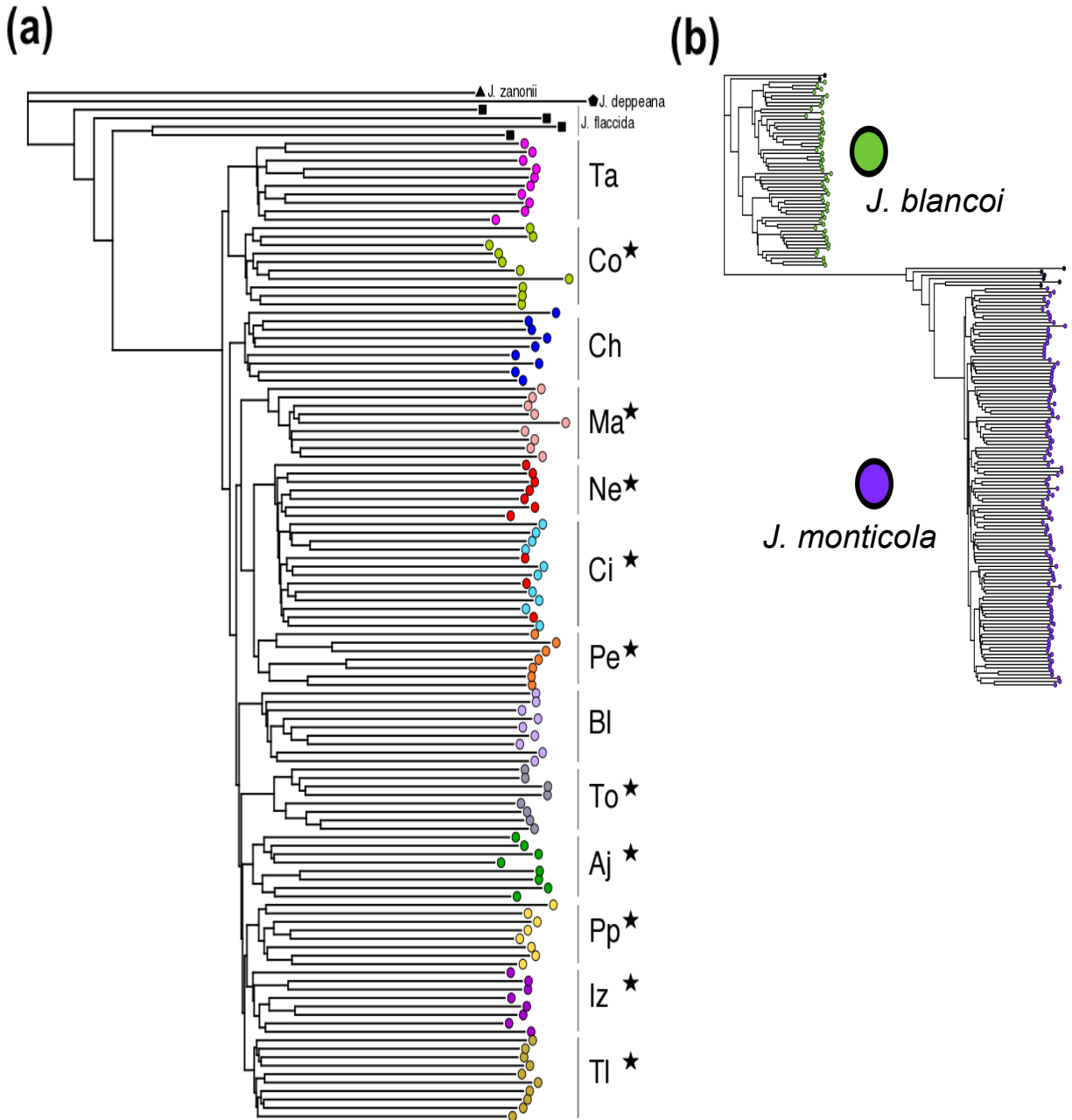


Figura 27. (a) Árbol NJ de 116 individuos de *J. monticola*. Cada población es representada en un color diferente, en negro se representan los grupos externos cercanos a *J. monticola*. La raíz es *J. zanonii*. (b) Árbol NJ de ambas especies. La raíz es *J. virginiana*.

DISCUSIÓN

ENSAMBLAJE DE NOVO Y FILTRADO DE DATOS RAD

Demultiplexeo

Se encontró variación en el número de *reads* obtenidos por muestra. Esto parece estar asociado primero a la librería en la que fueron procesadas (Figura 8), principalmente porque la librería 1 produjo una mayor cantidad de *reads*. Lo anterior puede deberse a variación durante el trabajo de laboratorio, ya que la librería 10 también se secuenció en una sola *lane* y no generó una cantidad de *reads* equivalente, y la librería 6 (segunda con mayor *reads*) fue secuenciada en conjunto con otra librería (librería 7). Dado que las muestras fueron procesadas al azar posterior a la extracción, posiblemente la diferencia en la cantidad de *reads* se deba a la calidad y cantidad del ADN asociado a cada muestra (Bonin *et al.* 2004).

Optimización del ensamblado de novo

Modificar los parámetros para alcanzar la optimización de las muestras ya se ha reportado en otros estudios utilizando el software *Stacks* (Mastretta-Yanes *et al.* 2015a), sin embargo el presente trabajo representa la primera vez que se efectúa con los parámetros de *pyRAD*. Esto permitió una mayor precisión al elegir los valores óptimos según las tasas de error en nuestras muestras, por lo cual se comprueba que dicho método es igual de relevante para el software *pyRAD*.

Al modificar el mínimo de cobertura para un *cluster* (*Mindepth*) estamos afectando directamente el llamado de loci. Entre más laxo sea este parámetro es más probable que errores de secuenciación sean tomados como variación alélica, incrementando erróneamente la estimación de heterocigosis. Conforme disminuimos el valor de este parámetro permitimos mayor incertidumbre en el ensamblaje (Figura 6), por lo que obtenemos mayor número de SNPs a costa de mayor *missing data*, lo cual

se ve reflejado en una mayor tasa de error de los SNPs. Lo inverso sucede al aumentar los valores en este parámetro. Si se establecen valores bajos es posible obtener estimaciones heterocigóticas pobres, ya que es muy difícil distinguir los errores de la secuenciación de heterocigotos en profundidades bajas (Mindepth, por debajo de 5). Si este parámetro es incorrecto podría conducir a un exceso de llamadas de heterocigotos por llamadas ambiguas de SNPs (N).

El máximo número de sitios con calidad menor a 20 (NQual) permite considerar determinado número de diferencias entre los *reads* de un individuo. Considerar valores altos de NQual genera mayor incertidumbre en el ensamblaje ya que se aceptan mayor número de llamados de SNPs de baja calidad que aumentan la tasa de error. Estos resultados son contrarios a lo encontrado por Hapke & Thiele (2016), quienes indican que modificar el NQual no tiene mayor influencia en los análisis con pyRAD. El umbral de agrupamiento de las secuencias (WClust) debe ser considerado según el NQual elegido. El WClust es el valor utilizado para agrupar a los *reads* dentro de los individuos y entre los individuos. La reducción en los valores de WClust corresponde a un aumento en la tasa de error que incrementa de manera drástica cuando se permiten umbrales de tolerancia menores a 0.90.

Las tasas de error cuando varían los valores de MaxSH no modificaron en gran medida la proporción de la obtención de SNPs en comparación a los demás parámetros evaluados, similar a lo encontrado a través de similitudes entre muestras pares (Hapke & Thiele 2016).

En los tres conjuntos (Jb, Jm y JbyJm) se obtuvieron mayor número de SNPs al considerar valores bajos en la cobertura mínima (MinCov, <20). El MinCov considera el número de muestras (individuos) que comparten la presencia de un SNP. Disminuir drásticamente el MinCov genera una matriz de datos más laxa donde existen falsos positivos y falsos negativos al generar matrices más estrictas al aumentar el MinCov (Eaton 2014).

El valor de los parámetros se ajustó con el número de muestras en cada compilación. Los conjuntos JbyJm y Jm presentan mayor número de SNPs que en el conjunto Jb, sin embargo la proporción de SNPs obtenidos con respecto a la cantidad de muestras comprobó que los conjuntos Jb y Jm obtuvieron mayor cantidad de SNPs.

Los conjuntos con muestras de la misma especie aumentaron la compatibilidad de SNPs con respecto al conjunto JbyJm que agrupa dos especies diferentes. El conjunto JbyJm obtuvo más incompatibilidades de SNPs, debido a que las mutaciones en el sitio de restricción de la enzima varían entre las especies (Mastretta-Yanes *et al.* 2014).

Incrementar el valor de MinCov cerca del número total de muestras favorece que los SNPs detectados estén en la mayor cantidad de muestras. Sin embargo, valores >50% de los individuos disminuyen la probabilidad de que esos SNPs existan en ambas especies. Contrario a elegir valores muy laxos, que incluyen a SNPs poco relevantes e incrementan el *missing data* en la matriz de datos. Las muestras eliminadas por tener >80% de *missing data* en los SNPs totales aumentan cuando se utilizan valores bajos en el MinCov.

Con los filtrados finales el *missing data* de los tres conjuntos de datos disminuyó ~6% en los 3 conjuntos, haciendo notoria la relevancia de ajustar adecuadamente el MinCov, descartando las muestras con al menos >80% de *missing data* en los SNPs totales.

Filtrado SNPs ligados

En este estudio nos referimos a los SNPs ligados como aquellos que comparten el mismo RAD-locus (86pb). Se han detectado errores de desequilibrio de ligamiento (DL) en poblaciones de reciente mezcla. Y debido a que el software ADMIXTURE no considera el DL (Alexander *et al.* 2009) se decidió analizar sólo el primer SNP de cada RAD-loci lo cual garantiza la calidad de llamado.

El porcentaje de variación asociado en los PCA aumentó al descartar los SNPs considerados ligados, a pesar de descartar más de 8,000 SNPs entre ambos conjuntos. Esto probablemente se debe a que el desequilibrio de ligamiento causado por la mezcla entre poblaciones puede provocar resultados erróneos, como la sobrestimación de ancestría que se ve reflejado en una mayor cantidad de grupos (Kaeuffer *et al.* 2007) o mezclas poblacionales falsas (Falush *et al.* 2003). Los SNPs descartados disminuyen el porcentaje de variación y son considerados redundantes.

GENÓMICA DE POBLACIONES Y ANÁLISIS FILOGENÉTICOS

Poblaciones ancestrales de J. blancoi y J. monticola.

Si se descartan frecuencias alélicas mínimas $MAF < 0.1$, se está disminuyendo el ruido que interviene al diferenciar los grupos por posibles errores de secuenciación. Sin embargo también pueden ser efecto de estructura por categorías altitudinal o poblacional. Las variantes raras pueden ser alelos derivados que surgen en poblaciones específicas después de una división del ancestro común. Por lo tanto pueden ser SNPs privados o particulares de una población, importantes para atribuir características fenotípicas específicas (Choudhury *et al.* 2014).

Juniperus blancoi y *J. monticola* tienen a Yécora y El Chico respectivamente como poblaciones diferenciadas por SNPs privados. Dichos SNPs podrían señalar a Yécora y El Chico como las poblaciones ancestrales de cada especie. Para *J. blancoi* esto se ha observado antes en un estudio con marcadores de cloroplasto donde se encontró que *J. blancoi* proviene de una distribución ancestral establecida en el noroeste de México en la SMOc (Mastretta-Yanes 2009). Lo cual fortalece el supuesto de que Yécora, la población más al noroeste de México, pueda ser considerada como la población ancestral de *J. blancoi*.

De esta forma, aunque descartar alelos raros pueda servir para descartar ruido, en este caso parece también descartar los loci con variación privada de las poblaciones ancestrales de estas especies. Esto es similar a lo propuesto por Choudhury *et al.* (2014).

Estructura genética de dos especies de Juniperus asociada a contrastantes historias evolutivas

Los árboles NJ construidos para ambas especies recuperan la topología filogenética ya reportada en estudios con marcadores nucleares, cpADN y nrITS (Mao *et al.* 2010). De manera similar los PCA de *J. blancoi* recuperan una estructura similar a la encontrada

por Moreno-Letelier *et al.* (2014). Sin embargo, estas especies difieren en el nivel de estructura genética encontrado entre sus poblaciones, lo que indica que tienen historias evolutivas contrastantes.

El PCA de *J. blancoi* muestra tres posibles grupos diferenciados (Figura 21). El primer grupo incluye a Huehuento, Mohinora y El Salto. El segundo a La Preciosita, San José Rincón, Presa Brokman y Rio Patitos. Finalmente el tercero incluye a Yécora. Se encontró que Rio Patitos podría compartir ancestría con Yécora o bien con las poblaciones de la FVT (Figura 23a).

Similar a lo encontrado por Moreno-Letelier *et al.* (2014) en *J. blancoi* se nota un agrupamiento de las poblaciones Presa Brokman, San José Rincón y La Preciosita. Estas poblaciones comparten condiciones altitudinales y latitudinales (Figura 23c). Las poblaciones de Mohinora (*var. huehuetensis*) y El Salto (*var. blancoi*) comparten la misma ancestría. De igual forma Rio Patitos (*var. blancoi*) y Yécora (*var. mucronata*) parecen haber tenido una ancestría común a pesar de estar distantes geográficamente (Jalisco y Sonora, respectivamente), además de que existen otras poblaciones intermedias entre ellas.

Los grupos ancestrales identificados para *J. blancoi* no corresponden con los grupos definidos por las variedades taxonómicas y presentan una distribución geográfica disyunta, similar a lo reportado con marcadores nucleares y de cloroplasto (Mastretta-Yanes *et al.* 2012; Moreno-Letelier *et al.* 2014).

Acorde al estudio de Moreno-Letelier *et al.* (2014) cada clado dentro de *J. blancoi* se asocia a un tiempo de divergencia relativamente antiguo. El Salto, Huehuento y Mohinora a 7.7 Ma, Presa Brokman, San José Rincón y La Preciosita a 6.3 Ma y Yécora y Rio Patitos a 6.1 Ma, lo cual indica que se trata de poblaciones muy aisladas entre sí por periodos largos. Esto apunta a que *J. blancoi* pudo ser abundante en el pasado y después haber atravesado por sucesivas fragmentaciones como lo discute Mastretta-Yanes *et al.* (2012).

El PCA de *J. monticola* muestra tres posibles grupos (Figura 22). El primer grupo incluye a Citlaltépetl, La Negra y Cofre de Perote, que son las poblaciones más al este de la FVTM. El segundo a La Malinche, Popocatépetl, Iztaccíhuatl, Tláloc, El Chico, Ajusco, Nevado de Toluca y Cerro Blanco, que son poblaciones del centro de la FVTM.

Finalmente el tercer grupo formado por Tancítaro y el Nevado de Colima, que se encuentran en el extremo oeste de la FVTM. Este agrupamiento coincide con las variedades morfológicas previamente descritas: *J. monticola* var. *orizabensis*, var. *monticola* y var. *compacta* respectivamente (Adams *et al.* 2007). La división en grupos de poblaciones (oeste, centro y este), que son cercanas entre sí, pero que se encuentran más alejadas de poblaciones de otros grupos es congruente con una dinámica de islas en el cielo como la propuesta para la FVTM (Mastretta-Yanes *et al.* 2016). Lo anterior sugiere que la dinámica de flujo génico de estas poblaciones posiblemente fue influenciada por los cambios climáticos del Pleistoceno.

Aunque en este estudio no realizamos análisis de reloj molecular, Mao *et al.* (2010) reporta una divergencia entre *J. monticola* y *J. flaccida* de aproximadamente 5 Ma, lo cual es más reciente que la divergencia de 9 Ma entre las poblaciones de la FVTM y el resto de las poblaciones de *J. blancoi*. Esta divergencia entre especies es notoria en el PCA y los NJ (Figura 16b y Figura 27b-28b). Dentro de *J. blancoi* encontramos una diferenciación marcada dentro de las poblaciones (Figura 23a), atribuida al tiempo de divergencia y a su condición fragmentada. Características que podrían considerarse dentro de la taxonomía de *J. blancoi* para elevar a nivel de especie a algunas de sus poblaciones.

Los *Juniperus* son tolerantes a condiciones áridas y semiáridas (Adams 2014). De hecho *J. flaccida*, la especie más cercana a *J. monticola*, crece en condiciones templadas a semiáridas del centro de México (Adams *et al.* 2007). Por ende es posible que la disminución de la temperatura del Mioceno-Plioceno pudiera haber favorecido la especiación de *J. monticola* y su posterior colonización de los ambientes alpinos donde hoy se encuentra.

SNPs asociados a categorías altitudinales

Los cambios en la distribución durante diferentes condiciones climáticas se ha reportado en México para oyameles y pinos (Moreno-Letelier & Piñero 2009). Sin embargo, el origen adaptativo del género *Juniperus* a climas secos y cálidos (Mao *et al.* 2010) los convierte en un modelo interesante para analizar la adaptación a hábitats más húmedos y fríos presentes en montañas alpinas.

Uno de los retos de este trabajo fue detectar SNPs que diferencian tanto poblaciones como categorías altitudinales, pues los primeros ocasionan incertidumbre debido al efecto demográfico que interviene en la diferenciación asociada a la altitud (cuatro encontrados en este trabajo). Para aminorar el efecto de esto en la presente tesis, primero se determinó la estructura poblacional dentro de cada especie.

En el presente estudio buscamos loci diferenciados por categorías altitudinales entre dos especies de *Juniperus*. Estos loci pueden considerarse SNPs candidatos que se diferenciaron mediante procesos de adaptación local. Sin embargo, hace falta complementar estos supuestos con análisis particulares de adaptación local (Wang *et al.* 2014; Vangestel *et al.* 2016).

Los resultados obtenidos permiten rechazar la hipótesis de que *J. blancoi* y *J. monticola* compartan loci asociados a condiciones de alta y de baja montaña independientemente de su especie. Encontrar diferentes SNPs que diferencian a poblaciones alpinas de las de baja montaña en cada especie nos habla de al menos dos vías por las cuales cada especie se adaptó. Esto contrasta con los resultados encontrados por Yeaman *et al.* (2016) donde se encontró convergencia entre dos especies de coníferas separadas hace más de 140 Ma. Una posible razón es que en dicho estudio se utilizó secuenciación de genomas completos, incluyendo loci parálogos, mientras que aquí se examinó un submuestreo del genoma y los loci parálogos se descartaron mediante los filtrados.

CONCLUSIONES

- 1) Este trabajo se muestra la importancia de estimar la tasa de error a partir de muestras duplicadas, ya que esto nos permite elegir los parámetros más adecuados, considerando el número de SNPs obtenidos y el menor porcentaje de *missing data*.
- 2) Optimizar los parámetros en agrupamientos separados por especie permite recuperar mayor número de SNPs que considerando a *J. blancoi* y a *J. monticola* dentro de una misma matriz de datos.
- 3) La estructura poblacional nos habla de procesos de divergencia y especiación ocurridos recientemente en la historia evolutiva de las coníferas. Los periodos de especiación y divergencia que tuvieron *J. blancoi* y *J. monticola* están ligados a variación ambiental producto de procesos históricos.
- 4) La estructura poblacional es mucho más marcada en *J. blancoi* que en *J. monticola*.
- 5) Mediante el análisis bioinformático de datos ddRAD se logró comparar poblaciones de *J. blancoi* y *J. monticola* bajo diferentes condiciones ambientales (alpinas y bosques de menor altitud).
- 6) Se encontraron 24 y 21 loci candidatos sobre los que la selección natural posiblemente está actuando en *J. blancoi* y *J. monticola*, respectivamente. Estos loci difirieron entre especie, por lo que no se apoya la hipótesis que la adaptación a condiciones alpinas se dió por convergencia evolutiva entre loci ortólogos.

REFERENCIAS

- Adams DRP (2014) *Junipers of the World: The Genus Juniperus, 4th Edition*. Trafford Publishing.
- Adams RP, Schwarzbach AE (2013) Phylogeny of *Juniperus* using nrDNA and four cpDNA regions. *Journal of Biogeography*, **95(2)**, 179–187.
- Adams RP, Schwarzbach AE, Morris JA, Socorro M, Elizondo G (2007) *Juniperus compacta* (Cupressaceae) A new species from México. *ResearchGate*.
- Aguirre-Planter E (2007) Flujo génico: métodos para estimarlo y marcadores moleculares. In: *Ecología molecular*, p. . Instituto Nacional de Ecología, México.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.
- Anderson JT, Lee C-R, Rushworth CA, Colautti RI, Mitchell-Olds T (2013) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology*, **22**, 699–708.
- Arroyo-Cabrales J, Carreño AL, Lozano-García S, Montellano-Ballesteros M (2008) La diversidad en el pasado. In: *Capital natural de México*, pp. 227–262. Conabio, México.
- Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, **3**, e3376.
- Blanquart F, Kaltz O, Nuismer SL, Gandon S (2013) A practical guide to measuring local adaptation. *Ecology Letters*, **16**, 1195–1205.
- Bonin A, Bellemain E, Bronken Eidesen P *et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.
- Cabrero J, Camacho JPM (2002) Fundamentos de genética de poblaciones. , 83–126.

- Castillo A (2007) La selección natural a nivel molecular. In: *Ecología molecular*, p. 608. México, D.F.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular ecology*, **22**, 3124–3140.
- Chong Z, Ruan J, Wu C-I (2012) Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics*, **28**, 2732–2737.
- Choudhury A, Hazelhurst S, Meintjes A *et al.* (2014) Population-specific common SNPs reflect demographic histories and highlight regions of genomic plasticity with functional relevance. *BMC Genomics*, **15**, 437.
- Clayton D (2015) *snpStats: SnpMatrix and XSnpmatrix classes and methods*.
- Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, **27**, 2156–2158.
- Darwin C (1859) *On the Origin of the Species by Natural Selection*. Murray.
- Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.
- Dennett DC (1996) *DARWIN'S DANGEROUS IDEA: EVOLUTION AND THE MEANINGS OF LIFE*. Simon & Schuster, New York.
- Dobzhansky T, Pavlovsky O (1957) An experimental study of interaction between genetic drift and natural selection. , **11**, 311–314.
- Dodsworth S, Leitch AR, Leitch IJ (2015) Genome size diversity in angiosperms and its influence on gene space. *Current Opinion in Genetics & Development*, **35**, 73–78.

- Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.
- Eguiarte LE, Aguirre-Liguori JA, Jardón-Barbolla L, Aguirre-Planter E, Souza V (2013) Genómica de poblaciones: nada en evolución va a tener sentido si no es a la luz de la genómica, y nada en genómica tendrá sentido si no es a la luz de la evolución. *Tip Revista Especializada en Ciencias Químico-Biológicas*.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.
- Ferrari L, Valencia-Moreno M, Bryan S (2005) Magmatismo y tectónica en la Sierra Madre Occidental y su relación con la evolución de la margen occidental de Norteamérica. *Boletín de la Sociedad Geológica Mexicana*.
- Futuyma DJ (2013) *Evolution*. Sinauer Associates.
- Gámez N, Escalante T, Rodríguez G, Linaje M, Morrone JJ (2012) Caracterización biogeográfica de la Faja Volcánica Transmexicana y análisis de los patrones de distribución de su mastofauna. *Revista mexicana de biodiversidad*, **83**, 258–272.
- Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, **11**, 759–769.
- González-Elizondo MS, González-Elizondo M, Tena-Flores JA, Ruacho-González L, López-Enríquez IL (2012) Vegetación de la Sierra Madre Occidental, México: una síntesis. *Acta botánica mexicana*, 351–403.

- Hapke A, Thiele D (2016) GIBPSs: a toolkit for fast and accurate analyses of genotyping-by-sequencing data without a reference genome. *Molecular Ecology Resources*, **16**, 979–990.
- Hedrick PW (2011) *Genetics of Populations*. Jones & Bartlett Learning.
- Hohenlohe PA, Catchen J, Cresko WA (2012) Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. *Methods in Molecular Biology (Clifton, N.J.)*, **888**, 235–260.
- Kaeuffer R, Réale D, Coltman DW, Pontier D (2007) Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. *Heredity*, **99**, 374–380.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.
- Khlestkina EK, Salina EA (2006) SNP markers: Methods of analysis, ways of development, and comparison on an example of common wheat. *Russian Journal of Genetics*, **42**, 585–594.
- Kisdi É (2002) Dispersal: Risk Spreading versus Local Adaptation. *The American Naturalist*, **159**, 579–596.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.
- Liu Y, Nyunoya T, Leng S *et al.* (2013) Softwares and methods for estimating genetic ancestry in human populations. *Human Genomics*, **7**, 1.
- Loewe L (2008) Negative Selection. *Nature Education*, **1(1):59**.
- Losos JB (2011) Convergence, Adaptation, and Constraint. *Evolution*, **65**, 1827–1840.

- Lynch M (2008) Estimation of Nucleotide Diversity, Disequilibrium Coefficients, and Mutation Rates from High-Coverage Genome-Sequencing Projects. *Molecular Biology and Evolution*, **25**, 2409–2419.
- Mao K, Hao G, Liu J, Adams RP, Milne RI (2010) Diversification and biogeography of Juniperus (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. *New Phytologist*, **188**, 254–272.
- Martínez M (1946) Los Juniperus mexicanos. *17(1-2): 3-128*.
- Mastretta-Yanes A (2009) Las poblaciones solitarias: fragmentación, diferenciación y rareza de un junípero ribereño y sus implicaciones para la conservación. Licenciatura Biología Thesis. Universidad Nacional Autónoma de México, México, D.F.
- Mastretta-Yanes (2014) Landscape genomics of tropical high altitude plant species. Doctor of Philosophy Thesis. University of East Anglia, United Kingdom.
- Mastretta-Yanes A, Arrigo N, Alvarez N *et al.* (2015a) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, **15**, 28–41.
- Mastretta-Yanes A, Moreno-Letelier A, Piñero D, Jorgensen TH, Emerson BC (2015b) Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. *Journal of Biogeography*, n/a-n/a.
- Mastretta-Yanes A, Wegier A, Vázquez-Lobo A, Piñero D (2012) Distinctiveness, rarity and conservation in a subtropical highland conifer. *Conservation Genetics*, **13**, 211–222.

- Mastretta-Yanes A, Zamudio S, Jorgensen TH *et al.* (2014) Gene Duplication, Population Genomics, and Species-Level Differentiation within a Tropical Mountain Shrub. *Genome Biology and Evolution*, **6**, 2611–2624.
- Metzker ML (2010) Sequencing technologies — the next generation. *Nature Reviews Genetics*, **11**, 31–46.
- Mitchell-Olds T, Willis JH, Goldstein DB (2007) Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics*, **8**, 845–856.
- Mooser F, Ramírez T (1987) Faja Volcánica Transmexicana, Morfoestructura, Tectónica y Vulcanotectónica (texto)., (mapa). *Boletín de la Sociedad Geológica Mexicana*, **48**, 75–80.
- Moreno-Letelier A, Mastretta-Yanes A, Barraclough TG (2014) Late Miocene lineage divergence and ecological differentiation of rare endemic *Juniperus blancoi*: clues for the diversification of North American conifers. *New Phytologist*, **203**, 335–347.
- Moreno-Letelier A, Piñero D (2009) Phylogeographic structure of *Pinus strobiformis* Engelm. across the Chihuahuan Desert filter-barrier. *Journal of Biogeography*, **36**, 121–131.
- Morin PA, Luikart G, Wayne RK, Group the S workshop (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.
- Morrone JJ (2005) Hacia una síntesis biogeográfica de México. *Revista mexicana de biodiversidad*, **76**, 207–252.
- Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**, 2841–2847.

- Nielsen R, Slatkin M (2013) *An Introduction to Population Genetics: Theory and Applications*. Sinauer Associates, Sunderland, Mass.
- Nystedt B, Street NR, Wetterbom A *et al.* (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature*, **497**, 579–584.
- Parchman TL, Gompert Z, Mudge J *et al.* (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, **21**, 2991–3005.
- Peña C (2012) *Metodos de inferencia filogenetica*.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, **7**, e37135.
- Plascencia RL, Castañón A, Raz-Guzmán A (2011) La biodiversidad en México, su conservación y las colecciones biológicas. , **101**, 36–43.
- Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, **81**, 559–575.
- Puritz JB, Matz MV, Toonen RJ *et al.* (2014) Demystifying the RAD fad. *Molecular Ecology*, **23**, 5937–5942.
- Rai HS, Reeves PA, Peakall R, Olmstead RG, Graham SW (2008) Inference of higher-order conifer relationships from a multi-locus plastid data set. *Botany*, **86**, 658–669.
- Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

- Schlötterer C (2004) The evolution of molecular markers — just a matter of fashion? *Nature Reviews Genetics*, **5**, 63–69.
- Templeton AR (2006) *Population Genetics and Microevolutionary Theory*. John Wiley & Sons, Hoboken, N.J.
- Templeton AR, Hemmer H, Mace G *et al.* (1986) Local adaptation, coadaptation, and population boundaries. *Zoo Biology*, **5**, 115–125.
- Teotónio H, Chelo IM, Bradić M, Rose MR, Long AD (2009) Experimental evolution reveals natural selection on standing genetic variation. *Nature Genetics*, **41**, 251–257.
- Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, **29**, 673–680.
- Vangestel C, Vázquez-Lobo A, Martínez-García PJ *et al.* (2016) Patterns of neutral and adaptive genetic diversity across the natural range of sugar pine (*Pinus lambertiana* Dougl.). *Tree Genetics & Genomes*, **12**, 51.
- Vázquez-Lobo A (1996) Filogenia de hongos endófitos del género *Pinus*: Implementación de técnicas moleculares y resultados preliminares. Universidad Nacional Autónoma de México, México.
- Wang G-D, Fan R-X, Zhai W *et al.* (2014) Genetic Convergence in the Adaptation of Dogs and Humans to the High-Altitude Environment of the Tibetan Plateau. *Genome Biology and Evolution*, **6**, 2122–2128.
- Wright S (1949) The Genetical Structure of Populations. *Annals of Eugenics*, **15**, 323–354.
- Wright AF (2001) Genetic Variation: Polymorphisms and Mutations. In: *eLS*, p. . John Wiley & Sons, Ltd.

Yeaman S, Hodgins KA, Lotterhos KE *et al.* (2016) Convergent local adaptation to climate in distantly related conifers. *Science*, **353**, 1431–1433.

Zeigler D (2014) Chapter 2 - Natural Selection. In: *Evolution* (ed Zeigler D), pp. 9–22. Academic Press, San Diego.

Zimin A, Stevens KA, Crepeau MW *et al.* (2014) Sequencing and Assembly of the 22-Gb Loblolly Pine Genome. *Genetics*, **196**, 875–890.

Zonneveld BJM (2012) Conifer genome sizes of 172 species, covering 64 of 67 genera, range from 8 to 72 picogram. *Nordic Journal of Botany*, **30**, 490–502.

ANEXO II

<i>J. blancoi</i> (62)		<i>J. monticola</i> (130)	
Huehuento (10)	Mohinora (10)	Yécora (8)	Río Patitos (8)
JbHu02.01	JbMh12.01	JbYc10.02	JbZh05.01
JbHu04.01	JbMh01.02	JbYc14.03	JbZh14.01
JbHu05.01	JbMh02.04	JbYc16.04	JbZh13.03
JbHu11.01	JbMh19.05	JbYc06.05	JbZh02.04
JbHu12.02	JbMh04.06	JbYc15.05	JbZh04.05
JbHu01.03	JbMh06.07	JbYc08.06	JbZh12.07
JbHu15.03	JbMh09.07	JbYc07.09	JbZh16.07
JbHu03.04	JbMh17.07	JbYc02.10	JbZh19.04
JbHu13.08	JbMh03.09		
JbHu14.06	JbMh18.03		
P. Brokman (6)	El Salto (8)	San José R. (6)	Preciosita (6)
JbBk19.01	JbS112.03	JbSn12.02	JbPr13.01
JbBk20.01	JbS111.04	JbSn07.02	JbPr04.03
JbBk22.01	JbS110.06	JbSn19.03	JbPr01.06
JbBk04.03	JbS101.07	JbSn01.05	JbPr11.07
JbBk09.03	JbS120.08	JbSn11.10	JbPr02.10
JbBk01.03	JbS118.10	JbSn03.09	JbPr19.07
	JbS07.02		
	JbS09.01		
Nev. Toluca (10)	Ajusco (10)	C. Blanco (10)	Tláloc (10)
JmToi09.01	JmAj24.04	JmBU14.01	JmTI16.01
JmToi02.02	JmAj04.04	JmBU02.06	JmTI01.02
JmToi11.05	JmAj01.04	JmBU05.06	JmTI18.04
JmToi13.05	JmAj22.05	JmBU03.07	JmTI05.05
JmToi10.07	JmAj06.08	JmBU17.07	JmTI20.07
JmToi01.08	JmAj10.08	JmBU18.08	JmTI02.09
JmToi06.09	JmAj03.09	JmBU15.09	JmTI04.09
JmToi08.09	JmAj05.09	JmBU04.08	JmTI17.09
JmToi12.09	JmAj21.09	JmBU16.09	JmTI19.10
JmToi07.07	JmAj23.09	JmBU01.10	JmTI03.04
C. de Perote (10)	Citlaltépetl (10)	La Negra (10)	Malinche (10)
JmPei05.01	JmCUI14.01	JmNeJ17.02	JmMaJ20.02
JmPei02.02	JmCUI10.02	JmNeJ05.03	JmMaJ19.03
JmPei19.02	JmCUI16.02	JmNeJ03.06	JmMaJ05.04
JmPei04.03	JmCUI09.04	JmNeJ21.06	JmMaJ22.08
JmPei20.03	JmCUI11.05	JmNeJ01.08	JmMaJ06.08
JmPei07.06	JmCUI12.05	JmNeJ04.10	JmMaJ02.09
JmPei21.06	JmCUI13.05	JmNeJ19.10	JmMaJ21.09
JmPei22.06	JmCUI08.08	JmNeJ18.03	JmMaJ23.09
JmPei06.07	JmCUI15.08	JmNeJ20.05	JmMaJ03.01
JmPei03.08	JmCUI17.06	JmNeJ02.07	JmMaJ04.02
Muestras r (15)	Muestras r (19)	Muestras r (15)	Muestras r (10)
JmChJ12_ir.06	JmPp16.01	JmChJ17_r.06	JmChJ04.03
JmCUI17.06	JmMa03.01	JmCUI17.06	JmChJ01.03
JmToi07.07	JmPp16_r.01	JmToi07.07	JmChJ03.06
JmNeJ02.07	JmMaJ04.02	JmNeJ02.07	JmChJ02.10
JmNeJ02_r.07	JmNeJ18.03	JmNeJ02_r.07	JmChJ05.10
JmToi07_r.07	JmMa04_ir.03	JmToi07_r.07	JmChJ11.10
JmBU04.08	JmPp04.03	JmBU04.08	JmTI03_r.04
JmBU04_r.08	JmPp04_r.03	JmBU04_r.08	JmBU16_ir.10
JmBU116.09	JmIz113.04	JmBU116.09	JmCoJ05.05
JmAJ23.09	JmNeJ18_ir.04	JmAJ23.09	JmNeJ20_r.05
JmAJ23_r.09	JmTI03_r.04	JmAJ23_r.09	JmIz113_ir.05
JmBU16_ir.10	JmCoJ05.05	JmBU16_ir.10	JmNeJ20.05
JmBU01.10	JmNeJ20_r.05	JmBU01.10	JmChJ12.05
JmBU01_r.10	JmIz113_ir.05	JmBU01_r.10	JmChJ20.10
	JmChJ12.05		JmCol05_r.05
	JmChJ20.10		
	JmCol05_r.05		

Muestras Jb y Jm sin r-----166
 Muestras con r-----60
 Outs sin r-----4
 Negativos-----10
 Total-----240