



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS**

**FACULTAD DE ESTUDIOS SUPERIORES DE IZTACALA
ECOLOGÍA**

**VARIACIÓN GENÓMICA Y DISTRIBUCIÓN POTENCIAL DE *Quercus
macdougalii* Martínez, ESPECIE ENDÉMICA DE OAXACA**

TESIS

**QUE PARA OPTAR POR EL GRADO DE:
MAESTRA EN CIENCIAS BIOLÓGICAS**

PRESENTA:

Biol. Nelly Jazmín Pacheco Cruz

TUTOR PRINCIPAL DE TESIS:

**DR. JORGE EDUARDO CAMPOS CONTRERAS
FES IZTACALA, UNAM**

COMITÉ TUTOR:

**DRA. ANA ELENA MENDOZA OCHOA
INSTITUTO DE ECOLOGÍA, UNAM**

**DR. JORGE CIROS PÉREZ
FES IZTACALA, UNAM**

CD. MX., NOVIEMBRE, 2019



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

OFICIO CPCB/1121/2019

ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
Presente

Me permito informar a usted que en la reunión ordinaria del Subcomité de Ecología y Manejo Integral de Ecosistemas y Biología Evolutiva y Sistemática, del Posgrado en Ciencias Biológicas, celebrada el día **26 de agosto de 2019** se aprobó el siguiente jurado para el examen de grado de **MAESTRA EN CIENCIAS BIOLÓGICAS** en el campo de conocimiento de **Ecología** de la estudiante **PACHECO CRUZ NELLY JAZMIN**, con número de cuenta **517024935** con la tesis titulada "**Variación genómica y distribución potencial de *Quercus macdougalii* Martínez, especie endémica de Oaxaca**", realizada bajo la dirección del **DR. JORGE EDUARDO CAMPOS CONTRERAS**, quedando integrado de la siguiente manera:

Presidente: DR. ANTONIO GONZÁLEZ RODRÍGUEZ
Vocal: DR. VICTOR MANUEL SALAZAR ROJAS
Secretario: DR. JORGE CIROS PÉREZ
Suplente: DRA. CECILIA LIANA ALFONSO CORRADO
Suplente: DR. RICARDO CLARK TAPIA

Sin otro particular, me es grato enviarle un cordial saludo.

ATENTAMENTE
"POR MI RAZA HABLARA EL ESPÍRITU"
Ciudad Universitaria, Cd. Mx., a 18 de octubre de 2019

COORDINADOR DEL PROGRAMA


DR. ADOLFO GERARDO NAVARRO SIGÜENZA



c. c. p. Expediente del alumno

AGRADECIMIENTOS INSTITUCIONALES

Al Posgrado en Ciencias Biológicas de la UNAM, por permitirme realizar mis estudios de Maestría y desarrollarme profesionalmente. A su planta docente, por compartir sus conocimientos e impulsar nuestro deseo por retribuir a la sociedad. A su área administrativa por mostrar la disponibilidad y compromiso para realizar todo lo necesario para lograr concluir esta meta.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo otorgado mediante la beca de posgrado (CVU: 815856) durante los semestres desde el 2017-2 al 2019-1.

A mi tutor de tesis, el Dr. Jorge Eduardo Campos Contreras, por brindarme su tiempo, asesoría, apoyo, orientación y compartir conmigo sus conocimientos y experiencias.

A los miembros de mi comité tutorial, por su compromiso y valiosas contribuciones durante el desarrollo de este proyecto de tesis.

- Dra. Ana Elena Mendoza Ochoa, por brindarme su apoyo y asesoría, así como por su amistad, confianza, comprensión, consejos y por las agradables reuniones en su laboratorio.
- Dr. Jorge Ciros Pérez, por su asesoría, disponibilidad y apoyo durante la realización de esta tesis y por compartir su tiempo y conocimientos.

AGRADECIMIENTOS A TÍTULO PERSONAL

A mi familia, quienes siempre han sido un pilar que me motiva a seguir adelante. A mi mamá, por siempre escucharme y alentarme a seguir mis aspiraciones, siempre con la frente en alto. A mi padre, por su apoyo y heredarme el gusto por la lectura. A mis queridos hermanitos, Manuel y Luis, por no darse por vencidos ante las adversidades, por enseñarme a no temer ser uno mismo y a valorar y disfrutar la vida. Gracias familia, por esas largas pláticas con el cafecito junto al fuego, por acompañarme al monte a buscar encinos y evitar que me pierda en el bosque y en la vida. Gracias por su tiempo y por estar conmigo en las buenas y en las malas, gracias por todo.

A mis abuelitos, por inculcarme respeto y curiosidad hacia la naturaleza, y por siempre alentarme a aprender, sin miedo a lo desconocido ni al fracaso. A mis tías, tíos, primas y primos quienes, sin importar distancias, siempre han seguido llamando, preocupándose y apoyándome. Muchas gracias por todo su cariño y espero poder estar ahí cuando ustedes lo necesiten.

A mis amigos de la infancia y de la carrera, por mantener el contacto a pesar del tiempo y la distancia, por los recuerdos memorables y por continuar con nuestra amistad, así como por recordarme nuestro compromiso social como personas y como biólogos.

A mis amigos, compañeros y colegas de laboratorio en la FES Iztacala, por permitirme conocer sus perspectivas y aspiraciones, por compartir dudas y frustraciones, así como la curiosidad y alegría por nuestra carrera. Alex, Clau, Ari, Andy, Karen y Ana, gracias por su valiosa amistad.

A mis estimados profesores quienes, me han alentado y apoyado a lo largo de mis estudios. Especialmente, agradezco a mis tutores, la Dra. Cecilia L. Alfonso Corrado y el Dr. Jorge E. Campos Contreras, quienes han marcado mi crecimiento tanto de forma profesional como personal. Gracias por aceptarme como tesista, por la confianza y por todas las oportunidades que me han brindado para seguir aprendiendo, muchas gracias por todo.

Al M. en C. Alejandro Monsalvo y a la Dra. Martha Martínez, por los gratos momentos de convivencia, por su amabilidad y accesibilidad al ayudarme a resolver cualquier duda.

Al Jurado de este proyecto de investigación:

Al Dr. Antonio González Rodríguez, al Dr. Víctor M. Salazar Rojas, al Dr. Jorge Ciro Pérez, a la Dra. Cecilia L. Alfonso Corrado y al Dr. Ricardo Clark Tapia. Gracias por su disposición y amabilidad, y por compartir conmigo sus conocimientos. Agradezco sus asesorías, comentarios y sugerencias, los cuales han contribuido a enriquecer y mejorar el presente trabajo.

A la Universidad de la Sierra Juárez (UNSIJ), por haber sido mi ventana hacia el mundo de la biología y de la investigación. Espero que siga siendo un centro de vinculación entre nuestros pueblos, la naturaleza y la ciencia.

Agradezco a la Universidad Nacional Autónoma de México (UNAM), la máxima casa de estudios del país, gracias por brindarme la oportunidad de conocer y ver más allá de lo que hubiera imaginado. Gracias por seguir siendo un centro que alienta la formación de profesionistas en ciencia y con conciencia. A la Unidad de Biotecnología y Prototipos (UBIPRO), y al Laboratorio de Bioquímica Molecular, gracias por proveer sus recursos e instalaciones.

A las comunidades oaxaqueñas de San Pedro Yólox, Santiago Comaltepec e Ixtlán de Juárez, por permitirme trabajar dentro de sus territorios y estudiar sus bosques. A sus comisariados y población, agradezco su apoyo y confianza.

Gracias

Shkisharhu li

‘Mamba’

DEDICATORIA

A la vida, que me ha dado tanto...

A mi familia, por su apoyo y amor incondicional

A todas aquellas personas que han estado “ahí” para mí y mi familia

Y a mi pueblo, que entre montañas vive, lucha y ama

“Dijo el loco y con noble melancolía, por las breñas del monte siguió trepando, y al perderse en las sombras aún repetía: ¡Hay que vivir sembrando! ¡Siempre sembrando!”

Extracto de la poesía “Sembrando” de Blanco Belmonte

*Lajó jin jan dzea ngan, í ngoo ngī jee mo’, nügī ni kian, ngoor fe’ri:
¡‘ne dzijnea gī ‘ma leje see naa! ¡kontöm ‘ne dzijnea!*

Chinanteco, variante de Santiago Comaltepec, Oax.

Traducción de la Profa. Vilma López Luna.

*Neshuna rapinikiaye dieye le shiato ribtesiyaye:
¡Tudiaba gadarhiu ka yaa! ¡Tudiaba gadarhiu!*

Zapoteco, variante Leaj, Ixtlán de Juárez, Oax.

Traducción de la Sra. Sofía García Hernández.

ÍNDICE

RESUMEN	1
ABSTRACT	2
1. INTRODUCCIÓN	3
2. MARCO TEÓRICO	5
2.1 De la genética a la genómica de poblaciones	5
2.2 Área de distribución de la especie	9
2.3 Endemismo	10
2.3 Los encinos en México	10
3. ANTECEDENTES	14
4. JUSTIFICACIÓN	16
5. HIPÓTESIS	17
6. OBJETIVOS	18
7. MATERIALES Y MÉTODOS	19
7.1 Área de estudio	19
7.2 Especie de estudio	22
7.3 Colecta y georreferenciación de material biológico	24
7.4 Variación genómica	24
7.4.1 Extracción, cuantificación y purificación de ADN	24
7.4.2 Genotipificación basada en secuenciación (SBG)	25
7.4.3 Ensamble e identificación de los SNPs.	25
7.4.3.1 Análisis CGUM	25
7.4.4 Genómica poblacional	28
7.4.4.1 Diversidad genética	28
7.4.4.2 Estructura genética poblacional	29
7.4.4.4 Identificación de SNPs <i>outliers</i>	32
7.5 Distribución potencial	33
7.5.1 Datos de presencia	33
7.5.2 Variables ambientales: CONABIO y WorldClim	33

7.5.3 Modelo de distribución potencial actual	34
7.5.4 Distribución con escenarios futuros: 2050 y 2070	35
7.5.5 Evaluación del modelo	35
7.6 Variación genómica, geografía y ambiente	36
8. RESULTADOS	37
8.1 Variación genómica	37
8.1.1 Diversidad genética	37
8.1.2 Estructura genética poblacional	40
8.1.4 SNPs <i>outliers</i>	47
8.2 Distribución potencial	50
8.3 La relación entre la variación genómica, la geografía y el ambiente	58
9. DISCUSIÓN	60
10. CONCLUSIONES	73
PERSPECTIVAS Y APLICACIONES	74
LITERATURA CITADA	76
Anexo I	92
Anexo II	93
Anexo III	94
Anexo IV	95
Anexo V	95
Anexo VI	96

Índice de figuras y cuadros

Figuras

	Página
Figura 1.- Área de estudio, Sierra Juárez, Oaxaca. Esta área se encuentra en una zona de transición entre dos provincias biogeográficas, la Sierra Madre del Sur y la provincia de Oaxaca. En la zona Norte se ubican cinco sitios de muestreo (CZ, MT, MC, MB y CY) y en la zona Sur se encuentran tres sitios (LS, CR y PZ) (Cuadro 1). -----	20
Figura 2.- Sitios de muestreo: a) Cerro Zacate; b) Carretera Oaxaca-Tuxtepec Km 175; c) Cerro Mirador; d) Parte baja del Cerro Mirador; e) Entrada a Comaltepec a partir de la Carretera Federal Oaxaca-Tuxtepec; f) Laguna del Sacrificio; g) Camino Real; h) Pozuelos. -----	22
Figura 3.- Individuos adultos de <i>Q. macdougallii</i> : a) Hojas de un individuo adulto; b) Hojas de un individuo con características arbustivas en el sitio MC; c) Hojas con evidencia de herbivoría; d) Individuos adultos en el sitio PZ (aprox. 15 m alt.); e) Individuo con apariencia arbustiva en el sitio CZ (3 m de alt.); f) Individuo adulto en el sitio MB cubierto de musgos y epífitas. -----	23
Figura 4.- Diagrama de violín, mostrando la acumulación y distribución de la heterocigosidad por locus. -----	39
Figura 5.- Representación gráfica de alelos únicos en los ocho sitios de muestreo (eje X) y por cada locus (eje Y). -----	40
Figura 6.- Árbol sin raíz construido a partir de distancias genéticas de Nei, por el método Neighbor-joining. Los números sobre las ramas representan las sustituciones nucleotídicas. -----	41
Figura 7.- Árbol construido a partir de distancias genéticas de Nei, por el método Neighbor-joining. En azul se señalan los individuos de la zona Norte y en rojo los de la zona Sur. -----	42
Figura 8.- Red de expansión mínima a partir de distancias genéticas de Nei. -	43
Figura 9.- Análisis de componentes principales (ACP), entre los ocho sitios de muestreo, considerando los dos primeros componentes principales. -----	44
Figura 10.- Análisis Discriminante de Componentes Principales (ADCP), entre los ocho sitios, con 10 componentes principales. -----	45

Figura 11.- Diagramas que representan la estructura poblacional asumiendo un agrupamiento de $K=2$. -----	46
Figura 12.- Promedios de probabilidad de pertenencia por sitios a partir de los datos obtenidos con ADMIXTURE, sobre una imagen satelital del área de estudio. -----	47
Figura 13.- Coeficientes de F_{ST} entre pares de sitios. El gradiente de color representa el rango de valores. -----	48
Figura 14.- Distribución de los SNPs según su valor F_{ST} entre los dos agrupamientos propuestos. En azul se presenta el 1% de los SNPs con los valores más altos de F_{ST} . -----	49
Figura 15.- Análisis de componentes principales (ACP), a partir del conjunto de los 82 SNPs <i>outliers</i> . -----	49
Figura 16.- Cambios de las frecuencias alélicas de los SNPs con los F_{ST} más altos, en los ocho sitios de muestreo. -----	50
Figura 17.- Análisis de correlación entre los dos conjuntos de variables: a) entre el conjunto de variables de CONABIO (Cuadro 4); y b) entre las 19 variables de WorldClim (Cuadro 5). El tamaño del cuadrado indica que tan fuerte es la correlación entre las variables y el color indica si la correlación es positiva (azul) o negativa (rojo). S/C significa que la variable no mostró cambios en sus valores. -----	53
Figura 18.- Mapa de la distribución potencial de <i>Q. macdougalii</i> : a-b) Modelo con las variables de WorldClim, en el área de las provincias biogeográficas SMS y PO; y c) a nivel nacional, señalando el límite del estado de Oaxaca; d-e) Modelo con las variables de CONABIO en el área de las provincias SMS y PO; y f) a nivel nacional, señalando el límite del estado de Oaxaca. La probabilidad de ocurrencia va de alta (rojo) a nula (azul) Las estrellas representan los sitios de muestreo para este estudio. El mapa de elevaciones en el fondo presenta su gradiente de altitud de blanco (mayor altitud) a negro (menor altitud). -----	54
Figura 19.- La prueba de Jackknife mostró la contribución al modelo de cada variable (azul), sin la variable (turquesa) y con todas las variables (rojo): a) CONABIO; y b) WorldClim. Los valores mostrados son promedio de las 100 réplicas ejecutadas. -----	55
Figura 20.- Mapas de distribución potencial proyectadas al futuro: a-b) escenario RCP4.5 y c-d) escenario RCP8.5, en el 2050; e-f) escenario RCP4.5	

y g-h) escenario RCP8.5 en el 2070. En forma de estrella se muestran los sitios de muestreo. ----- 57

Figura 21.- Diagrama de la correlación de Mantel entre una matriz de distancias geográficas (Km) y una de distancias genéticas representadas por flujo génico (Nm), los colores representan donde se acumulan la mayoría de los valores, de azul a rojo. ----- 58

Cuadros

Cuadro 1.- Descripción de los ocho sitios de muestreo. ----- 21

Cuadro 2.- Descripción genética de los ocho sitios. Donde H_o = heterocigosidad observada, H_e = heterocigosidad esperada, $SegSit$ = sitios segregantes, π = diversidad nucleotídica, θ = diversidad nucleotídica poblacional y D = prueba D de Tajima. ----- 38

Cuadro 3.- Número de alelos totales y únicos por sitio. Entre paréntesis, se ordenan de mayor a menor número de alelos. ----- 40

Cuadro 4.- Capas obtenidas del Geoportal de CONABIO, usadas como variables para el modelo de distribución potencial. ----- 51

Cuadro 5.- Capas ambientales obtenidas del portal de WorldClim, usadas como variables para el modelo de distribución potencial. En gris se señalan las ocho variables seleccionadas para la modelación. ----- 52

Cuadro 6.- Áreas de distribución potencial en Km². Actual; con los modelos de CONABIO y WorldClim, Proyección futura, con los modelos de WorldClim bajo los escenarios de cambio climático RCP4.5 y RCP8.5, ambos para los años 2050 y 2070. Se señala el valor de probabilidad de ocurrencia; el mínimo se estableció como 0.2 para cada modelo y el máximo fue el valor más alto reportado para cada modelo. ----- 57

Cuadro 7.- Coeficientes de correlación entre las matrices, obtenidos por medio de la prueba parcial de Mantel. Gen =matriz de flujo génico, Geo =matriz de distancias geográficas, Amb =matriz de distancias euclidianas entre las variables ambientales: CONABIO y WorldClim, con 10 y 8 variables respectivamente. ----- 60

RESUMEN

México es considerado el país con el mayor número de especies del género *Quercus* a nivel mundial. Alberga una gran cantidad de especies endémicas, como *Q. macdougallii*, la cual se distribuye en la Sierra Norte de Oaxaca, y se encuentra catalogada como una especie vulnerable en la IUCN. En este estudio, se analizó la variación genómica y la distribución potencial de la especie, considerando ocho sitios de muestreo ubicados en la zona Norte y Sur de su distribución conocida (aproximadamente de 28 Km). Para el análisis de la variación genómica, se identificaron 8,186 polimorfismos de un solo nucleótido (SNPs). Para la distribución potencial actual se consideraron dos conjuntos de variables ambientales (WorldClim y CONABIO). También, para la distribución potencial futura, se consideraron dos escenarios de cambio climático, RCP4.5 y RCP8.5, ambos para el 2050 y 2070, utilizando las proyecciones de WorldClim. Se encontró una diversidad genética moderada (H_o y $H_e \sim 0.2$), bajos niveles de diferenciación genética y endogamia ($F_{ST}=0.0204$, $F_{IS}=0.0101$). Se observó un gradiente de variación genética entre la zona Norte y Sur. Aunque la especie se mantiene como una sola población, se identificó un grupo genético con estructuración incipiente en la zona Sur. Los SNPs *outliers* no mostraron evidencia de genotipos adaptados a nivel local, sin embargo, se observó la presencia de alelos asociados a cierta zona. La distribución potencial actual de la especie es restringida con una alta especificidad hacia ciertas variables. La distribución potencial a futuro sugiere una reducción en su área de distribución, siendo la zona Sur la más afectada. Debido a lo anterior, se sugiere priorizar la conservación y monitoreo de la especie en toda su área de distribución, particularmente en la zona Sur, donde se observó una alta variación genética y donde los efectos del cambio climático posiblemente serán mayores.

Palabras clave: *Quercus macdougallii*; genómica de poblaciones; SNPs; endemismo; distribución potencial; cambio climático.

ABSTRACT

Mexico is considered the country with the largest number of *Quercus* genus species worldwide. It has a large number of endemic species, such as *Q. macdougalii*, which is distributed in the Sierra Norte de Oaxaca and is listed as a vulnerable species in the IUCN. In this study, we analyzed the genomic variation and the potential distribution of the species, considering eight sampling sites located across the North and South of its known distribution (approximately 28 Km). For the analysis of genomic variation, 8,186 single nucleotide polymorphisms (SNPs) were identified. For the current potential distribution, two sets of environmental variables were considered (WorldClim and CONABIO). Also, future potential distribution, was modelled using WorldClim projections by considering two climate change scenarios, RCP4.5 and RCP8.5, both scenarios for 2050 and 2070. Moderate levels of genetic diversity were found (H_o y $H_e \sim 0.2$), together with low levels of genetic differentiation and inbreeding ($F_{ST}=0.0204$, $F_{IS}= 0.0101$). A gradient of genetic variation was observed between the North and South zones. Although the species can be considered as one population, an incipient structuring genetic group was identified in the South. *Outliers* SNPs showed no sustainable evidence of locally adapted genotypes, however, the presence of unique alleles associated with a certain area was observed. The current potential distribution of the species is restricted with high specificity towards certain variables. The potential future distribution suggests a considerable reduction in its distribution area, with the South zone being the most affected. Therefore, it is suggested to prioritize the conservation and monitoring of the species throughout its distribution range, particularly in the South, where high genetic variation was observed and where the effects of climate change may be more drastic.

Keywords: *Quercus macdougalii*; population genomics; SNPs; endemism; potential distribution; climate change.

1. INTRODUCCIÓN

Los bosques constituyen aproximadamente el 82% de la biomasa continental (Roy *et al.*, 2001), albergan más del 50% de la biodiversidad terrestre y mantienen un rol significativo en el secuestro de carbono, lo que les confiere una gran importancia en el funcionamiento y mantenimiento de sus ecosistemas (Neale & Kremer, 2011; Alberto *et al.*, 2013; Cavender-Bares, 2016).

Los sistemas forestales actualmente enfrentan una serie de desafíos importantes, entre los cuales destacan la demanda de madera, la presión para conservar las áreas forestales, el cambio climático global y las amenazas asociadas. Para hacer frente a estos desafíos, se ha estudiado la interacción de los factores evolutivos, la demografía y la estructura poblacional en especies arbóreas, a partir del conocimiento de las bases genéticas de la variación, la divergencia y la adaptación en sus poblaciones (Eriksson, 1998; Namkoong, 2001; González-Martínez *et al.*, 2006; Plomion *et al.*, 2016b), estudios en los cuales la genómica de poblaciones ha desempeñado un rol significativo (Fady *et al.*, 2016; Holliday *et al.*, 2017).

Dentro los principales grupos de plantas que conforman los bosques en Norte América, el género *Quercus*, comúnmente llamados encinos, constituye un grupo dominante de plantas leñosas, en biomasa y riqueza de especies; asimismo, es considerado un grupo modelo en investigaciones evolutivas, ya que presenta una historia compleja y dinámica (Banda *et al.*, 2016; Cavender-Bares, 2016). Actualmente, México es considerado el mayor centro de diversificación a nivel mundial del género *Quercus*, debido a que presenta el mayor número de especies, de las cuales, más de la mitad son endémicas (Valencia, 2004; Romero-Rangel *et al.*, 2015).

Quercus macdougalii Martínez, es un encino endémico de la Sierra Norte de Oaxaca (Romero-Rangel *et al.*, 2015). Estudios previos indican que cumple un importante rol ecológico dentro del funcionamiento de su ecosistema (Anacleto-Carmona, 2015; Pariona *et al.*, 2017; Clark-Tapia *et al.*, 2018); sin embargo, se encuentra dentro de la categoría de vulnerable (VU D2), en la Lista Roja de Especies Amenazadas de la Unión Internacional para la Conservación de la Naturaleza [UICN] (2010), y está propuesta dentro de la categoría de Amenazada en el Proyecto de modificación del Anexo Normativo III, de la Norma Oficial Mexicana 059 [NOM-059-SEMARNAT-2010] (Diario Oficial de la Federación, 2018).

Debido a lo anterior, se considera importante el estudio de la variación genómica y la distribución potencial de *Q. macdougallii*, ya que, al contribuir al conocimiento de esta especie endémica, a nivel genómico y ambiental, se espera llegar a una mejor toma de decisiones para su conservación.

2. MARCO TEÓRICO

2.1 De la genética a la genómica de poblaciones

La genética de poblaciones analiza la variación y distribución de las frecuencias alélicas en ciertos loci; mediante este enfoque, se puede examinar el estado actual de la especie al igual que su capacidad para responder a factores bióticos, como parásitos, enfermedades, depredadores y factores abióticos, como el cambio climático o la fragmentación del hábitat (Hedrick, 2011; Hedrick & Hurt, 2012; Primack, 2012).

El uso de marcadores moleculares para analizar la variación genética se ha intensificado con la llegada de la genómica, principalmente por el desarrollo de las tecnologías de secuenciación de la siguiente generación (NGS, por sus siglas en inglés) (Fu & Yang, 2017). Este enfoque permite muestrear el genoma de forma profunda ofreciendo ventajas importantes, como la detección de miles de polimorfismos en todo el genoma (Allendorf *et al.*, 2010; Stapley *et al.*, 2010; Li *et al.*, 2012; Narum *et al.*, 2013).

La principal diferencia entre los términos de genética y genómica de poblaciones, es el contraste entre el número de loci que se habían estado analizando previo al uso de las tecnologías NGS (Mariette *et al.*, 2002). La genómica de poblaciones es definida como el estudio de la variación genética a partir del análisis simultáneo de numerosos loci o regiones a nivel del genoma, con el fin de comprender los roles de los procesos evolutivos que influyen en la variación, a través de los genomas y sus poblaciones (Luikart *et al.*, 2003; Weigel & Nordborg, 2015).

Los estudios de genómica de poblaciones desempeñan un papel importante en la conservación de las especies. Por ejemplo, al detectar polimorfismos adaptativos en poblaciones divergentes (Moritz, 2002), al proporcionar un marco espacial para revelar la acción de la selección natural en poblaciones naturales y al correlacionar la variación alélica y la ambiental (González-Martínez *et al.*, 2006).

2.1.1 Fuerzas evolutivas

Las fuerzas evolutivas que influyen en las frecuencias genotípicas de las poblaciones son la mutación, el flujo génico, la deriva genética, la selección natural y la endogamia. Todas ellas

producen cambios direccionales a través del tiempo y generan diferentes niveles y patrones de variación genética (Hedrick, 2011).

La mutación, indispensable para que ocurra la evolución, es considerada como fuente de variación genética, ya que genera cambios en el material genético, los cuales, pueden ocurrir en diferentes niveles, por ejemplo, a nivel puntual, es decir, involucrando el cambio en uno o varios nucleótidos, o a nivel cromosómico (Letelier, 2007; Hedrick, 2011). En las mutaciones puntuales podemos identificar sustituciones, inserciones y supresiones o deleciones. Debido a estos cambios en las secuencias de ADN, las mutaciones pueden ser benéficas, neutrales, deletéreas o letales, dependiendo de sus efectos sobre la eficacia biológica de los organismos y de sus poblaciones (Barrick & Lenski, 2013).

Por otro lado, la selección natural implica la herencia de caracteres que han conferido a sus portadores una mayor eficacia biológica y supervivencia, así como la herencia de parte de la variación genética ligada a esos caracteres. La selección natural puede ser descrita por cuatro modelos: 1) la selección estabilizadora, se produce cuando individuos con un cierto rasgo han tenido éxito con respecto a la gran mayoría de los demás individuos que conforman la población; 2) la selección direccional, que favorece a los individuos con rasgos de un extremo de la distribución y genera una mayor eficacia biológica a los individuos que los portan; 3) la selección disruptiva, donde los fenotipos más aptos están en cada uno de los dos extremos de la variación, siendo las formas intermedias seleccionadas en contra; 4) la selección balanceadora, en la cual los extremos de una característica son seleccionados en contra, siendo los organismos con mayor eficacia aquellos con características intermedias (Zeigler, 2014).

En las poblaciones bajo selección natural se pueden observar tres probables escenarios evolutivos: 1) el especialista, hace hincapié en una alta especificidad de los individuos hacia determinado hábitat, generando una baja adaptación hacia otros ambientes y una migración restringida, lo que le confiere una baja probabilidad de establecerse en nuevos hábitats; 2) el generalista, en el cual, los organismos tienen una mayor tolerancia hacia distintos tipos de hábitats, lo que les presenta mayores oportunidades de colonizar nuevos ambientes; 3) múltiples especialistas, en el cual, se observan adaptaciones locales a los diferentes hábitats alternativos con un flujo génico limitado entre las poblaciones (Collins, 2011).

Otra fuerza evolutiva es la deriva génica, la cual se refiere al cambio aleatorio en la frecuencia de alelos en la transmisión de gametos de una generación a la siguiente (Dobzhansky & Pavlovsky 1957; Hedrick, 2011). Tiende a ocurrir con mayor intensidad, después de que una población experimenta un cuello de botella, donde los alelos raros o con poca frecuencia, tienen una mayor probabilidad de perderse o de fijarse, provocando una disminución en la variación genética (Klug & Cummings, 2013). Esta disminución, está estrechamente vinculada al tamaño efectivo de la población, por lo cual, se considera que mientras mayor sea el tamaño efectivo hay más posibilidades de reducir los efectos de la deriva génica. Sin embargo, existen excepciones en las cuales, poblaciones con bajo tamaño efectivo pueden presentar una alta variación, mayor a la observada en algunas poblaciones de gran tamaño y de rápido crecimiento poblacional (Hedrick, 2011).

Otra fuerza que contribuye a disminuir la variación genética es la endogamia, la cual influye en la estructura poblacional al homogeneizar los genotipos producidos dentro de un linaje, aumentando así el potencial de diferenciación entre ellos (Charlesworth & Wright, 2001). En plantas, el comportamiento de los polinizadores y la escasa dispersión de semillas y polen son algunos de los factores que pueden aumentar los niveles de endogamia (Hedrick, 2011).

El flujo génico (Nm), se refiere al movimiento de genes de una población a otra (Hedrick, 2011). Esta fuerza, juega un papel importante en los aspectos integrales del proceso de adaptación local y estructura poblacional (Kisdi, 2002; Kawecki & Ebert, 2004). En las plantas, la dispersión de semillas y del polen son impulsores esenciales del flujo génico, el cual, dependiendo de sus patrones y niveles, puede determinar hasta qué grado cada población de una especie es una unidad evolutiva independiente (Slatkin, 1994). Por otro lado, la cantidad de flujo génico necesaria para prevenir la evolución independiente de las poblaciones de una especie dependerá de las otras fuerzas evolutivas (Piñero *et al.*, 2008).

El flujo génico o la ausencia de éste, puede influir drásticamente en la distribución geográfica de una especie, ya sea por medio del aislamiento por distancia, del efecto fundador o por migración entre otros procesos (Hedrick, 2011).

2.1.2 Estructura genética poblacional

La estructura genética es la distribución no aleatoria de alelos o genotipos en un tiempo y espacio determinado, la cual, está moldeada por la interacción entre fuerzas evolutivas,

principalmente el flujo génico (Loveless & Hamrick, 1984). La variación en los efectos de estas fuerzas en el área de distribución de la especie, resulta en patrones espaciales característicos de la estructura genética de las poblaciones (Jarvis *et al.*, 2005; Larrañaga & Hormaza, 2016).

Los índices de distancia genética se calculan entre pares de poblaciones y pretenden describir el grado de diferencia genética entre ellas. La medida de distancia genética más usada es la distancia de Nei (Nei, 1987), la cual, pretende estimar el número de mutaciones que a nivel nucleotídico se han acumulado en las secuencias de dos linajes, a partir del tiempo que ha transcurrido desde su divergencia original (Piñero *et al.*, 2008).

Existen adicionalmente una infinidad de índices de distancia genética (Piñero *et al.*, 2008). Por ejemplo, el índice de fijación (F_{ST}) entre dos poblaciones (Wright, 1951) y el flujo génico (Nm) (Rousset, 1997).

2.1.3 Polimorfismo de un solo nucleótido (SNP)

Los polimorfismos de un solo nucleótido (SNPs, por sus siglas en inglés), son marcadores moleculares que representan una diferencia única de nucleótidos entre dos o más individuos, en una ubicación específica del genoma (Appleby *et al.*, 2009). Este marcador es el tipo más frecuente de polimorfismo genético, ya que son los más abundantes y ampliamente distribuidos a lo largo del genoma. Debido a que su aparición y distribución puede variar entre individuos, se puede obtener una alta densidad de marcadores cerca de un lugar específico en el genoma (Agarwal *et al.*, 2008).

Los SNPs pueden ser transiciones (C/T, G/A) y transversiones (C/G, A/T, T/G, C/A), siendo los SNPs C/T los más frecuentes fuera de las regiones transcritas como resultado del aumento de la metilación de la citosina y la desaminación amplificada de la citosina (Edwards *et al.*, 2007).

SNPs outliers

Los métodos para la identificación de loci atípicos o *outliers*, es decir, aquellos que no se encuentran dentro de un patrón de neutralidad esperado y que suelen identificarse como candidatos para una selección divergente (Narum & Hess, 2011), se pueden dividir en dos categorías, la primera basada en el desequilibrio de ligamiento entre los loci y la segunda

basada en las diferencias a nivel de variación y divergencia genética entre los individuos (Vasemägi & Primmer, 2005). Por ejemplo, un valor alto de F_{ST} ha sido considerado como un buen parámetro para identificar SNPs *outliers* (Martins *et al.*, 2018; Pina-Martins *et al.*, 2019).

Genotipificación basada en secuenciación (SBG)

Mediante la genotipificación basada en secuenciación (SBG, por sus siglas en inglés), es posible identificar miles de SNPs. Esta tecnología se caracteriza por ser capaz de genotipificar cualquier estructura poblacional aún si se estudia un organismo no modelo (Truong *et al.*, 2012).

Con este enfoque, que se basa en la secuenciación de fragmentos asociados a sitios de restricción (RAD-seq, por sus siglas en inglés), es posible secuenciar potencialmente miles de loci para cientos de individuos, independientemente del tamaño y conocimiento previo del genoma. Asimismo, permite identificar regiones genómicas específicas que pueden haber experimentado la selección natural y mejorar la precisión de las inferencias demográficas al aumentar en gran medida el número de marcadores putativamente neutrales analizados (Mastretta-Yanes *et al.*, 2015).

2.2 Área de distribución de la especie

Uno de los métodos más recientes para determinar las áreas de distribución de las especies, son los llamados modelos de distribución (Peterson, 2001; Mateo *et al.*, 2011). En general, los modelos de distribución son una representación cartográfica de la idoneidad ambiental para una especie dada (Guisan & Zimmermann, 2000); la mayoría de estos modelos siguen un método similar; utilizan registros puntuales de presencia/ausencia de una especie, incluyen información de variables de tipo ambiental y usan algún algoritmo o criterio que delimita el área de distribución (Peterson, 2001).

Uno de los modelos más eficientes en la predicción de la distribución espacial de las especies es el modelo Maxent (Phillips *et al.*, 2006), el cual, se basa en la máxima entropía y modelos bayesianos (Lehtomäki & Moilanen, 2013). Este modelo se enfoca en caracterizar la idoneidad del hábitat de la especie e identificar espacialmente los sitios más aptos para su distribución pasada, actual y/o futura (Morales, 2012).

En México se han modelado las áreas de distribución geográfica de diversas especies de plantas, con el objetivo de ser una herramienta en la restauración ecológica, la priorización de zonas para la conservación biológica y en la modelación de los efectos del cambio climático sobre estas especies (Villaseñor & Téllez-Valdés, 2004; Contreras-Medina *et al.*, 2010; Naranjo-Luna, 2014; Gorgonio-Ramírez, 2015; Hernández-Ruíz *et al.*, 2016; Alfonso-Corrado *et al.*, 2017).

Para especies arbóreas endémicas, amenazadas y relictos, estos modelos han brindado un escenario de distribución útil para regionalizar y buscar nuevas áreas de distribución (Martínez-Meyer *et al.*, 2004; Mejías *et al.*, 2007; Naranjo-Luna, 2014; Alfonso-Corrado *et al.*, 2017).

2.3 Endemismo

Las especies endémicas son aquellas que crecen naturalmente en un área geográfica única, cuya distribución puede ser local, nacional, regional o continental (Primack, 2012). Una especie endémica puede ocupar solo uno o pocos hábitats especializados, además de presentar la tendencia a formar una sola población en toda su distribución (Işık, 2011; Primack, 2012).

Estos atributos, ya sea de manera individual o en combinación, hacen que una especie sea propensa a una mayor tasa de extinción. Cuando los hábitats de una especie rara y/o endémica se dañan y/o se fragmentan por la mala gestión y otras actividades humanas, las áreas de distribución, el tamaño poblacional y la variabilidad genética de la especie tienden a reducirse, volviendo a sus miembros vulnerables a la extinción a una velocidad más rápida que para otras especies (Işık, 2011).

Debido a ello, y en un esfuerzo por mantener la biodiversidad, las especies endémicas se deben monitorear y administrar, con programas precisos y efectivos de conservación y restauración, los cuales, garanticen la permanencia de sus poblaciones a largo plazo (Navarro & Guitian, 2003; Işık, 2011).

2.3 Los encinos en México

El género *Quercus* (encinos o robles), es el más grande dentro de la familia Fagaceae y el que presenta la mayor distribución en todo el mundo (Valencia, 2004). Se estima que existen entre 400 y 500 especies, y en América se distribuyen desde Canadá hasta Colombia,

incluyendo Cuba (Nixon & Muller, 1997; Manos *et al.*, 1999; Manos & Stanford, 2001; Nixon, 2006).

Una última actualización en la clasificación infragenérica del género *Quercus*, basada en los rasgos morfológicos, las relaciones filogenéticas y la historia evolutiva del género, propone el reconocimiento de dos subgéneros; *Quercus* y *Cerris*, el primero con cinco secciones (*Quercus*, *Ponticae*, *Virentes*, *Lobatae* y *Protobalanus*), y el segundo con tres (*Cyclobalan*, *Ilex* y *Cerris*) (Denk *et al.*, 2017).

Los encinos han sido reconocidos por contribuir significativamente al mantenimiento y funcionamiento de sus ecosistemas (Dickie *et al.*, 2007; Nguyen *et al.*, 2016; Cheeke *et al.*, 2017), así como al bienestar humano (Kroeger *et al.*, 2014), y en los últimos años ha aumentado el número de herramientas y recursos genómicos para su estudio (Konar *et al.*, 2016; Sork *et al.*, 2016a; Plomion *et al.*, 2016a, 2018; Gugger *et al.*, 2017; McVay *et al.*, 2017; Guerrero-Sánchez *et al.*, 2019).

En México, dependiendo del autor, se encuentran entre 150 y 166 especies, lo que es aproximadamente un tercio del género a nivel mundial, por otra parte, cerca de la mitad de las especies que se encuentran en el país, son endémicas (Axelrod, 1983; Pingarroni *et al.*, 2019, en prensa). Debido a estas características, México ha sido considerado como el mayor centro de diversificación en el mundo (Nixon, 1998; Valencia, 2004; Romero-Rangel *et al.*, 2015).

En el país, se encuentran especies de las secciones *Quercus* (encinos blancos, 81 *spp.*), *Lobatae* (encinos rojos, 76 *spp.*) y *Protobalanus* (encinos intermedios o de copa dorada, 4 *spp.*). En la sección *Quercus* más de la mitad de sus especies (47 *spp.*) son endémicas (Valencia, 2004). En Oaxaca, el estado con el mayor número de especies, la única especie endémica de la sección *Quercus*, es *Q. macdougallii* (Valencia, 2004; Romero-Rangel *et al.*, 2015).

Variación genética y genómica

Se ha estudiado la variación genética de diversos encinos por medio de marcadores moleculares tipo microsatélite. Por ejemplo, se han reportado altos niveles de heterocigosidad, como en *Q. grisea* (63 alelos y $H_e=0.877$; Rosas-Osorio, 2010), *Q.*

sideroxylla (84 alelos y $H_e=0.855$; Alfonso-Corrado *et al.*, 2014), *Q. candicans* ($H_o=0.675-0.881$ y $H_e=0.642-0.854$; Oyama *et al.*, 2018), *Q. crassifolia* ($H_o=0.7$; Gorgonio-Ramírez *et al.*, 2017; Oyama *et al.*, 2018), y *Q. castanea* ($H_o=0.599-0.881$; Oyama *et al.*, 2018), *Q. scytophylla* y *Q. hypoleucooides* ($H_o=0.633-0.905$ y $H_e=0.555-0.841$; Peñaloza-Ramírez *et al.*, 2010), y *Q. sartorii*, *Q. meavei* y *Q. paxtalensis* ($H_o=0.7$; De Luna-Bonilla, 2018), entre otros más.

Recientemente, el primer estudio de la variación a nivel genoma para un encino en México, fue para *Q. rugosa*, un encino de amplia distribución, para el cual se identificaron 5,354 SNPs y se observó una diversidad genética relativamente alta ($H_e=0.364$; Martins *et al.*, 2018).

Por otro lado, los estudios genéticos para encinos endémicos son escasos. En Oaxaca se destacan los trabajos de los dos encinos endémicos del estado; *Q. mullerii* (sección *Lobatae*), encino microendémico de la Sierra Sur (Pingarroni *et al.*, 2019, en prensa) y *Q. macdougallii* (sección *Quercus*) encino endémico de la Sierra Norte (Molina-Garay, 2011). En ambos trabajos, por medio del análisis de microsatélites, se encontraron bajos niveles de diversidad genética.

Por consiguiente, resulta necesario determinar o en su caso, confirmar los niveles de variación de especies endémicas con alto valor ecológico y social como *Q. macdougallii*, con el objetivo de contribuir a los planes de conservación y manejo de la especie.

Modelos de distribución potencial

El Panel Intergubernamental del Cambio Climático [IPCC, por sus siglas en inglés] (2007) indica que las formas de vida más vulnerables serán las arbóreas tales como *Pinus*, *Quercus*, *Abies* y *Juniperus*, debido a su intervalo estrecho de tolerancia térmica (Laurent & Vilá, 2003). Las especies del género *Quercus* son elementos clave del ecosistema, por lo cual, la conservación de este importante capital natural requiere el conocimiento de la interacción de las poblaciones y comunidades de encinos con el medio ambiente, especialmente ante el aumento de la temperatura atmosférica detectada en varias partes del mundo (Bacon, 1998; Frich *et al.*, 2002; IPCC, 2007).

Actualmente, no existe información sobre tolerancias térmicas de los encinos mexicanos, sin embargo, las proyecciones de cambio climático para México indican tendencias en el aumento de la temperatura, la disminución general de la precipitación y el cambio en la distribución temporal de la precipitación (Karmalkar *et al.*, 2011; Sáenz-Romero *et al.*, 2009). Bajo este escenario, los climas áridos se expandirán y las especies de bosques templados se encontrarán entre las más vulnerables, debido a su ubicación actual en las partes altas montañosas, considerados como ambientes fríos y húmedos (Sáenz-Romero *et al.*, 2009).

Modelos de distribución potencial de varias especies de encinos en escenarios de cambio climático, han indicado una disminución desde el 7% hasta el 48% en su área idónea, para el año 2050 (Gómez-Mendoza & Arriaga, 2007). El conocimiento acerca del hábitat de la especie es fundamental para determinar qué variables ambientales están limitando la distribución potencial de esta especie y para establecer medidas de conservación eficientes. Por lo anterior, un objetivo importante es identificar la distribución geográfica potencial del hábitat de *Q. macdougalii*.

3. ANTECEDENTES

El estudio de la variación genómica en poblaciones de encinos, cuyo enfoque metodológico se basa en la secuenciación masiva y la identificación de SNPs, ha comenzado recientemente, por otra parte, este enfoque ha impulsado el desarrollo de nuevos estudios, a nivel nacional e internacional, ya que confiere una mayor profundidad en los análisis, lo cual permite detectar patrones de variación e inferir qué posibles mecanismos evolutivos están actuando sobre la especie.

Recientemente, un estudio ha evaluado la historia evolutiva de *Q. suber*, especie distribuida en la zona occidental del mediterráneo, para obtener información sobre cómo la adaptación local está modelando el reservorio genético de la especie y cómo responderá la especie ante los cambios climáticos globales desde una perspectiva genética. Mediante el uso de la secuenciación por genotipificación (GBS), se identificaron 1,996 SNPs. Los valores de F_{ST} entre los sitios de muestreo fueron de 0.0028 a 0.1216, con una F_{ST} global de 0.0541. La prueba de Mantel no reveló evidencia de aislamiento por distancia entre los individuos. Se detectaron entre 17 y 29 SNPs *outliers*. Los resultados indicaron que la especie no tiene una fuerte estructura poblacional, ya que esencialmente se comporta como una sola población, donde el equilibrio entre el flujo génico y la adaptación local mantiene un grado de homogeneidad en el conjunto de genes de la especie, a lo largo de su distribución. Por otro lado, observaron gradientes de variación, lo cual, puede ayudar a que la especie haga frente a las condiciones locales cambiantes (Pina-Martins *et al.*, 2019).

En México, este enfoque se ha utilizado para identificar la variación genómica que puede conferir adaptaciones locales a *Q. rugosa*, una especie de amplia distribución, así como la predicción del desajuste adaptativo bajo los cambios climáticos futuros. Mediante el uso de la GBS, se identificaron 5,354 SNPs, a partir de 103 individuos analizados, de 17 sitios de muestreo. El valor de F_{ST} entre los SNPs fue de 0.056, y vario de 0.037 a 0.095 entre los sitios de muestreo, la H_e promedio fue de 0.364. Se observó un patrón de aislamiento por distancia con la prueba de Mantel ($r=0.475$, p -valor=0.015). El agrupamiento Bayesiano implementado en STRUCTURE identificó un $K=2$ y la distribución de este agrupamiento en el paisaje muestra un gradiente entre este y oeste. Se identificaron 74 SNPs *outliers*, los cuales presentaron valores de 0.165 a 0.314, y 97 SNPs asociados con la variación del clima. Se identificaron patrones espaciales de variación genética con una fuerte asociación con la estacionalidad de la precipitación y con las distancias geográficas. En conjunto, con los

análisis de distribución futura, se pronosticaron regiones donde las poblaciones de *Q. rugosa*, podrían estar en riesgo debido al cambio climático (Martins *et al.*, 2018).

Los estudios reportados hasta ahora acerca de *Q. macdougallii*, se han enfocado en la zona Norte de su distribución conocida y la mayoría se han centrado en su ecología. En esta zona se han registrado eventos de perturbación antropogénica cercanos a la especie, lo cual influye en la tasa de supervivencia de las plántulas, reportándose diferencias entre sitios con y sin disturbio (10% y 75% respectivamente) (Pariona *et al.*, 2017; Clark-Tapia *et al.*, 2018).

También, se ha observado una distribución agregada de la especie, un número poblacional bajo, una regeneración escasa, una baja tasa de germinación y diferencias significativas en la abundancia de sus individuos entre sitios de muestreo ubicados en la zona Norte, así como entre diferentes niveles de altitud (msnm). Asimismo, se ha sugerido que la altitud y el tipo de suelo pueden afectar considerablemente su distribución (Anacleto-Carmona, 2015).

Por otro lado, mediante el análisis de modelos generalizados lineales y no lineales, se han identificado una serie de variables ambientales que pueden influir ampliamente en la abundancia de la especie. Al considerar a las covariables, se identificó a las temperaturas máximas y mínimas, así como a la precipitación durante periodos de tiempo específicos. Por medio de los modelos aditivos generalizados (GAM) simples, se identificó a la temperatura media en el mes más frío, la temperatura media en el mes más cálido y la precipitación promedio en primavera (abril y mayo). Mientras que los GAM múltiples, identificaron a la pendiente media del terreno, el aspecto dominante del sitio (cenit, noreste, este, sureste, sur, suroeste, oeste, noroeste y norte), la elevación sobre el nivel del mar y la duración media del período libre de heladas (Antúnez *et al.*, 2017).

El único estudio acerca de su variación genética, identificó 26 alelos mediante el uso de tres microsátélites de la serie *quruGA*, y reportó una baja variabilidad genética ($H_o=0.402$; $H_e=0.557$) y una baja estructura genética ($F_{ST}=0.021$), lo cual, sugiere un alto flujo génico entre los sitios muestreados y un posible evento reciente de cuello de botella génico (Molina-Garay, 2011).

4. JUSTIFICACIÓN

Quercus macdougalii, es el único encino endémico de la sección *Quercus* reportado para Oaxaca. Es considerada una especie vinculada a importantes procesos ecológicos, sin embargo, como se ha mencionado previamente, también presenta características que la hacen más vulnerable ante la perturbación de su hábitat, como un bajo número poblacional y baja heterocigosidad.

Debido a lo anterior, es importante el estudio de su variación a una escala más fina, la cual, puede lograrse mediante el uso de la genotipificación basada en secuenciación (SBG) y la identificación de polimorfismos de un solo nucleótido (SNPs). El aumentar el número de marcadores que representan una mayor cobertura del genoma, nos permite inferir los mecanismos evolutivos a los que ha estado sometida la especie y eventualmente, permitirá la identificación y estudio de patrones adaptativos y su relación con aspectos fisiológicos y morfológicos. Es importante resaltar que el presente estudio representa uno de los primeros trabajos en utilizar la secuenciación masiva en el análisis de la variación genómica de un encino endémico de México.

Asimismo, el estudio de la distribución potencial actual y futura de la especie, así como la relación entre su variación genómica, su distribución geográfica y ciertas variables ambientales, podrán contribuir a una mejor toma de decisiones para la conservación de la especie frente al riesgo de posibles cambios ambientales.

5. HIPÓTESIS

- Dada su distribución agregada y bajo número poblacional, se espera observar bajos niveles de variación genómica.
- Dada la ubicación de los sitios de muestreo, además de la presencia de dispersión del polen por viento, se espera encontrar una baja estructura genética.
- Dada la información previa acerca de la especie, se espera que su área de distribución potencial sea restringida y altamente específica. Además, se espera una reducción de esta área bajo dos escenarios de cambio climático.

6. OBJETIVOS

6.1 Objetivo general

- Analizar la variación genómica y caracterizar la distribución potencial actual y futura de *Quercus macdougallii* en la Sierra Norte de Oaxaca.

▪ 6.2 Objetivos específicos

- Identificar polimorfismos de un solo nucleótido (SNP) por medio de genotipificación basada en secuenciación (SBG).
- Analizar los niveles de variación genómica, la estructura y el flujo génico dentro y entre los sitios de muestreo.
- Modelar la distribución potencial y la proyección futura de *Q. macdougallii*, bajo dos escenarios de cambio climático para los años 2050 y 2070.
- Analizar la relación entre la variación genómica, la geografía y el ambiente.
- Proponer sitios prioritarios para la conservación de *Q. macdougallii*, en la Sierra Norte de Oaxaca, a partir del conocimiento de los patrones espaciales de su variación genómica.

7. MATERIALES Y MÉTODOS

7.1 Área de estudio

El área de estudio se encuentra dentro de la Sierra Juárez de Oaxaca, en el Distrito de Ixtlán de Juárez. Esta región también es conocida como la Sierra Norte, y se ubica dentro del sistema montañoso Sierra Madre de Oaxaca. La conformación montañosa de la región se debe a los procesos de orogénesis a los que estuvo sometida desde el Paleozoico (INEGI, 2000; Anacleto-Carmona, 2015). Esta región abarca la transición entre la Provincia de la Sierra Madre del Sur y la Provincia de Oaxaca, asimismo, esta región ha sido descrita como una zona altamente compleja, con alta diversidad y endemismos (Espinosa *et al.*, 2008).

En las partes altas y medias de la Sierra Norte predomina el clima templado subhúmedo seco C(w0) y templado húmedo C(m). La precipitación del mes más seco es menor a 40 mm, con un promedio de 767 mm y con lluvias invernales entre 5% y 10% del total anual (Clark-Tapia *et al.*, 2016).

La distribución de la vegetación está asociada con el clima y la elevación, siendo los principales tipos de vegetación el bosque de pino-encino (de 2,000 a 3,200 msnm), el bosque de encino y el bosque de encino-pino (de 1,400 a 2,000 msnm) y el bosque de niebla (1,000 a 2,600 msnm) (INEGI, 2000).

El área de estudio en la zona norte se encuentra en la vertiente del Cerro Zacate-Humo Grande, en la cual, se reporta un clima templado húmedo con lluvias en verano, una precipitación media anual de 1,600 mm y una temperatura media anual de 14°C (Anacleto-Carmona, 2015), en esta zona colindan los municipios de San Pedro Yólox y Santiago Comaltepec. Por otra parte, la zona Sur del área de estudio se encuentra en el municipio de Ixtlán de Juárez, en la vertiente del Cerro Shiaa Rua Via.

Estos municipios emplean el sistema de bienes comunales, regulando así el acceso, uso y manejo de sus recursos forestales (Gómez-Mendoza *et al.*, 2006; Robson, 2008; Anta-Fonseca *et al.*, 2010; Gorgonio-Ramírez *et al.*, 2017). Sin embargo, los incendios han sido un factor de perturbación en dos períodos, uno ocurrido en 1983 y el otro en 1998, así como heladas severas en el 2008 (Robson, 2008) y a principios de 2018 (N. Pacheco Obs. per.).

De acuerdo con la distribución conocida de la especie, se seleccionaron ocho sitios de muestreo (Fig. 1), cuyas coordenadas y respectiva altitud se muestran en el Cuadro 1.

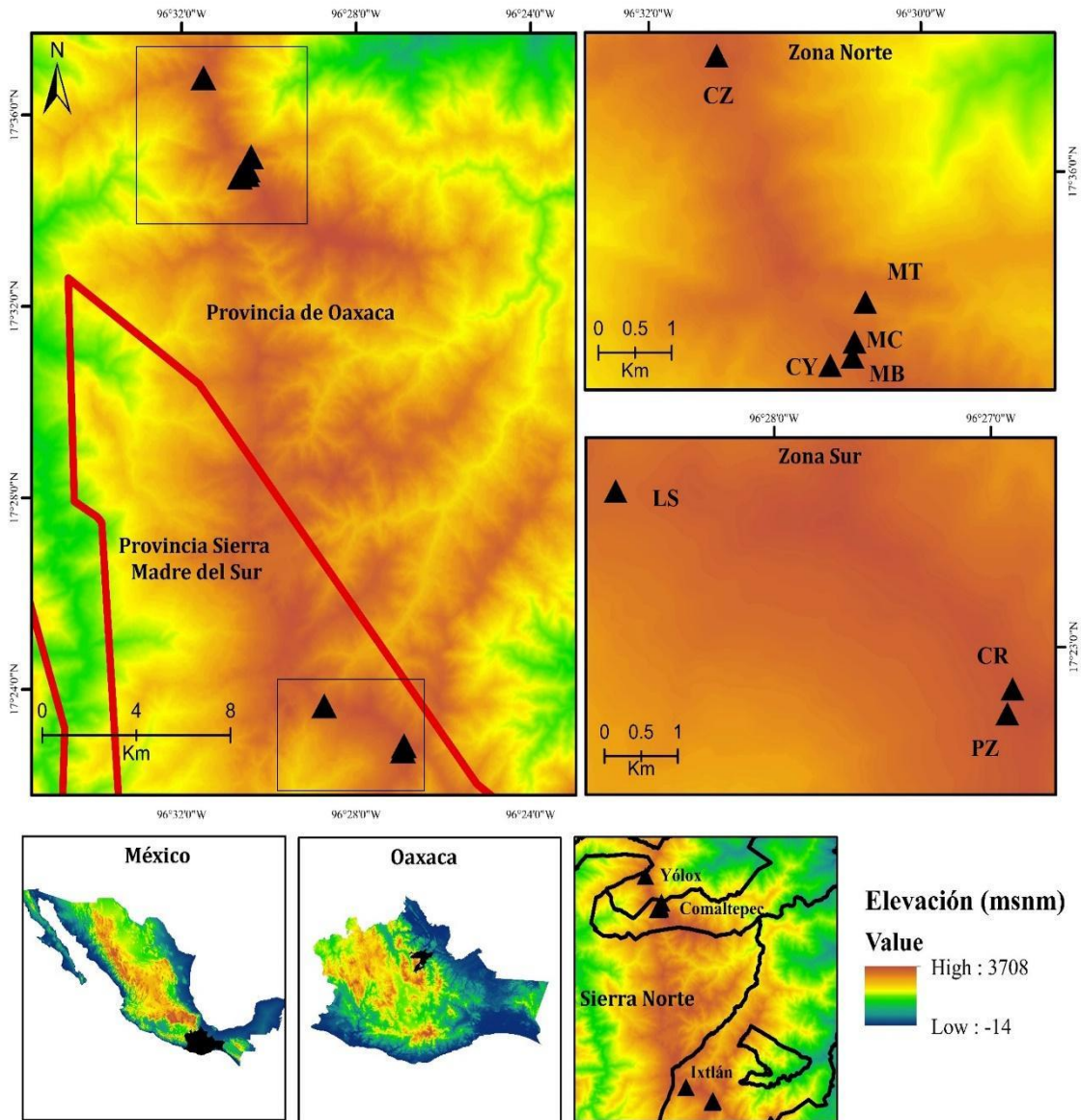


Figura 1.- Área de estudio, Sierra Juárez, Oaxaca. Esta área se encuentra en una zona de transición entre dos provincias biogeográficas, la Sierra Madre del Sur y la provincia de Oaxaca. En la zona Norte se ubican cinco sitios de muestreo (CZ, MT, MC, MB y CY) y en la zona Sur se encuentran tres sitios (LS, CR y PZ) (Cuadro 1).

Cuadro 1.- Descripción de los ocho sitios de muestreo.

	Sitio	Clave	Longitud Oeste	Latitud Norte	Altitud (msnm)	N° Ind
Zona Norte	Cerro Zacate	CZ	96° 31' 30.06''	17° 36' 46.72''	3021	10
	Carretera Oax-Tuxtepec	MT	96° 30' 24.55''	17° 35' 07.56''	2747	10
	Cerro Mirador	MC	96° 30' 29.25''	17° 34' 51.73''	2999	10
	Mirador parte baja	MB	96° 30' 30.18''	17° 34' 45.69''	2930	10
	Entrada a Comaltepec	CY	96° 30' 40.02''	17° 34' 42.27''	2839	10
Zona Sur	Laguna del sacrificio	LS	96° 28' 43.91''	17° 23' 39.35''	2976	5
	Camino Real	CR	96° 26' 54.06''	17° 22' 49.44''	3074	10
	Pozuelos	PZ	96° 26' 55.5''	17° 22' 43.22''	3083	15

El sitio ubicado más al norte del área de estudio es conocido como Cerro Zacate, donde a partir de los 2,600 msnm abundan los encinos y otras especies latifoliadas, siendo nula la presencia de pinos, por otro lado, en la zona más alta se realizan ceremonias religiosas católicas (Fig. 2a). Los otros cuatro sitios de la zona Norte (Fig. 2b-2e) se encuentran cerca de la carretera federal Oaxaca-Tuxtepec N° 175, y cerca de un pequeño aserradero, en esta zona se lleva a cabo un plan de aprovechamiento forestal enfocado a especies del género *Pinus*, por lo cual, es común observar zonas de desmonte para la reforestación con coníferas de interés comercial (Anacleto-Carmona, 2015).

Por otro lado, en los sitios de muestreo ubicados en la zona Sur (Fig. 2f-2h), se distribuyen principalmente pinos, encinos y algunos abetos. Estos sitios han sido asociados a centros religiosos (católicos y prehispánicos) y a la presencia de manantiales, también conocidos como ojos de agua, lugares en los que se realizan actividades ecoturísticas.

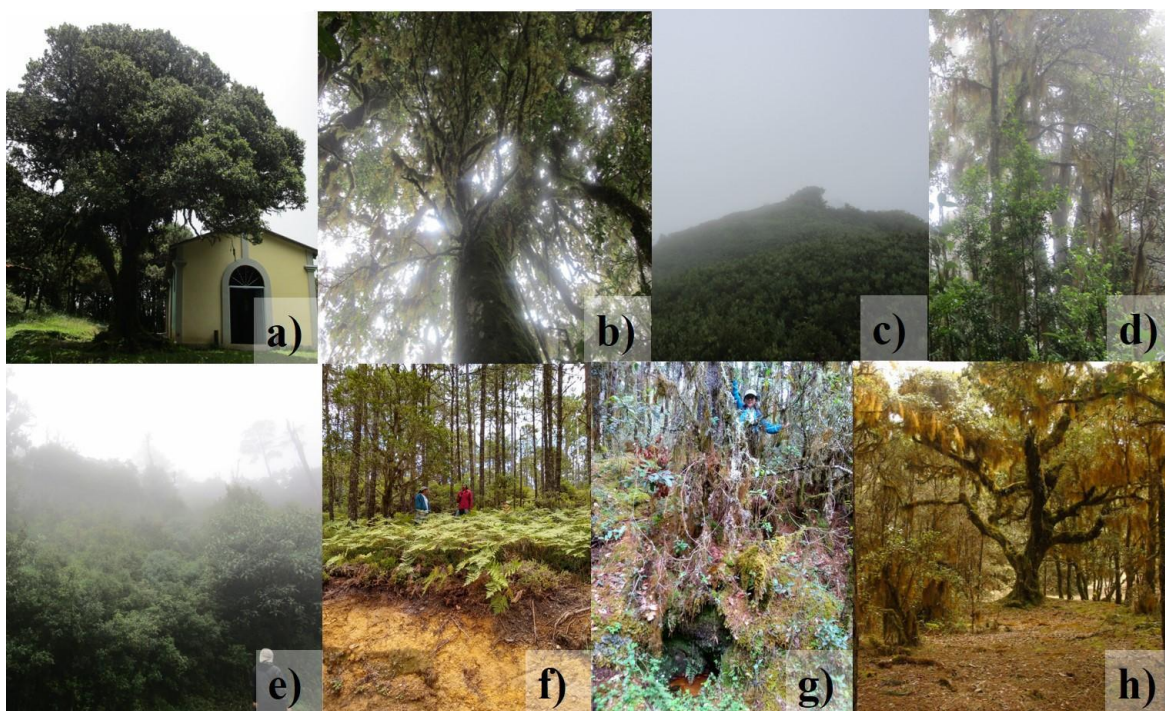


Figura 2.- Sitios de muestreo: a) Cerro Zacate; b) Carretera Oaxaca-Tuxtepec Km 175; c) Cerro Mirador; d) Parte baja del Cerro Mirador; e) Entrada a Comaltepec a partir de la Carretera Federal Oaxaca-Tuxtepec; f) Laguna del Sacrificio; g) Camino Real; h) Pozuelos.

7.2 Especie de estudio

Quercus macdougalii (Martínez, 1963). An. Inst. Biol. UNAM. 34, 147-149. Holotipo Cerro del Humo Chico, Santiago Comaltepec, Oax., Mex. MacDougall, 503 MEXU. Encino conocido localmente como *Yaa dua yu* (árbol para casa o clavado en el suelo, *Zapoteco*) y *'Ma kue* (árbol con hojas en forma de cuchara o jícara, *Chinanteco*).

La especie *Q. macdougalii* pertenece a la familia Fagaceae, género *Quercus*, subgénero *Quercus* y sección *Quercus* (encinos blancos) (Valencia, 2004; Romero-Rangel *et al.*, 2015). Es un árbol monoico de 20 a 25 m de altura, caducifolio, con un tronco de hasta 100 cm de DAP (diámetro a la altura del pecho); corteza gris; ramillas de color gris claro, en ocasiones ennegrecidas, acanaladas, comúnmente aplanadas, de 2 a 2.5 mm de diámetro con indumento de tricomas fasciculados y sésiles. Hojas jóvenes de color verde claro, haz lustroso con indumento de escasos tricomas fasciculados y sésiles. Las láminas de las hojas maduras

presentan una textura coriácea, son oblongo-elípticas, de 2 a 6 cm de largo y de 1 a 3.5 cm de ancho, con ápice agudo u obtuso, mucronado, con base redondeada, cordada o en ocasiones asimétrica. Florece en mayo y fructifica en octubre. Frutos anuales, en grupos de 3 o 4, sésiles. Bellota globosa de 15 mm de largo por 17 mm de diámetro (Romero-Rangel *et al.*, 2015).

Quercus macdougalii es un encino endémico de la Sierra Norte de Oaxaca (Valencia, 2004; Romero-Rangel *et al.*, 2015), que presenta una gran variabilidad en la morfología de sus hojas. En ciertos individuos se observa una apariencia arbustiva (Fig. 3), específicamente en aquellos que se encuentran en los puntos de mayor altitud de la zona Norte (N. Pacheco Obs. per.).

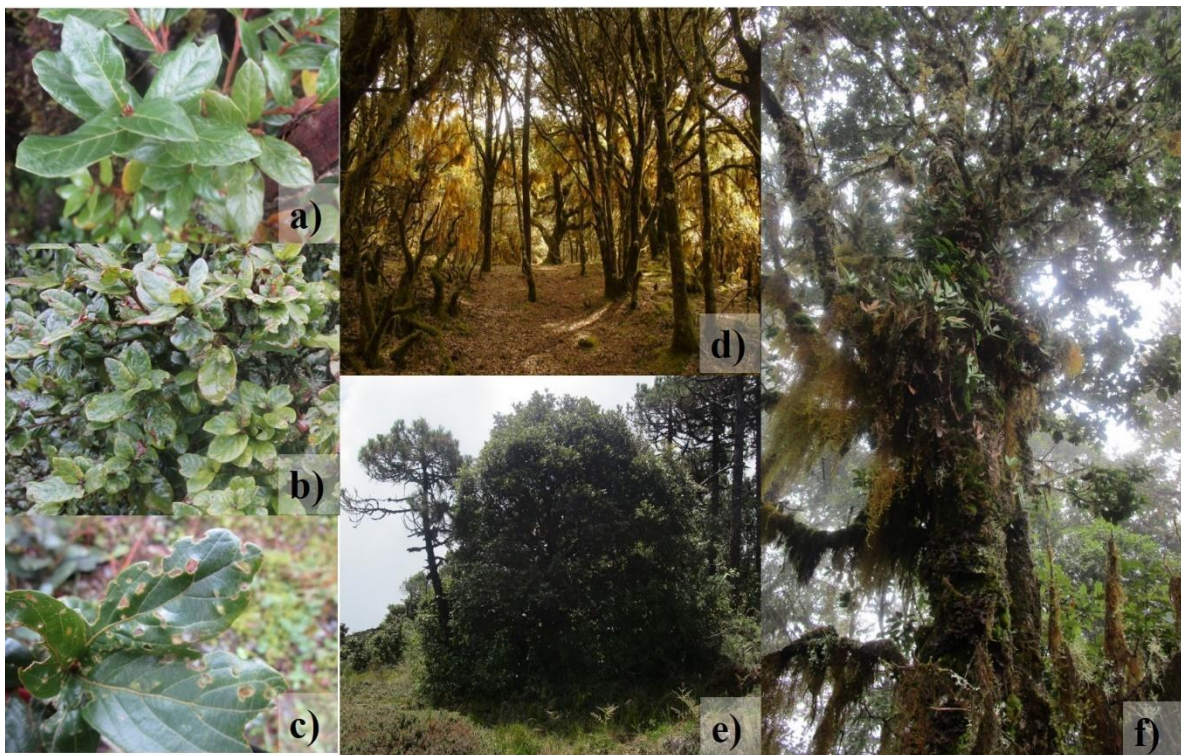


Figura 3.- Individuos adultos de *Q. macdougalii*: a) Hojas de un individuo adulto; b) Hojas de un individuo con características arbustivas en el sitio MC; c) Hojas con evidencia de herbivoría; d) Individuos adultos en el sitio PZ (aproximadamente 15 m de altura); e) Individuo con apariencia arbustiva en el sitio CZ (aproximadamente 3 m de altura); f) Individuo adulto en el sitio MB cubierto de musgos y epífitas.

Esta especie se encuentra dentro de la categoría de vulnerable (VU D2), en la Lista Roja de Especies Amenazadas de la Unión Internacional para la Conservación de la Naturaleza [UICN] (2010). También, está propuesta dentro de la categoría de Amenazada en el Proyecto de modificación del Anexo Normativo III, de la Norma Oficial Mexicana 059 [NOM-059-SEMARNAT-2010] (Diario Oficial de la Federación, 2018).

La especie presenta problemas de supervivencia (Pariona *et al.*, 2017), y se considera sometida a presiones causadas por actividades humanas, como ganadería, agricultura y tala, así como a eventos estocásticos como incendios y heladas (Robson, 2008; Anacleto-Carmona, 2015).

7.3 Colecta y georreferenciación de material biológico

Entre 2017 y 2018 se recolectaron hojas maduras (considerando la textura coriácea de las hojas como indicador de madurez), de 80 individuos de *Q. macdougalii* en ocho sitios. Se colectaron entre 5 y 15 individuos por sitio (Cuadro 1), a lo largo de un transecto de 1.25 km, muestreando un individuo aproximadamente cada 20 m. Estos individuos colectados fueron georreferenciaron por medio de un sistema GPS (Garmin Oregon 750) con datum WGS84. El material biológico de cada individuo se guardó a -20°C en bolsas de plástico con etiquetas para su posterior identificación.

7.4 Variación genómica

7.4.1 Extracción, cuantificación y purificación de ADN

El ADN genómico se extrajo a partir de 100 µg de tejido foliar siguiendo el protocolo de DNeasy Plant Mini Kit QIAGEN. La calidad del ADN fue determinada por medio de electroforesis en gel de agarosa al 1% y su concentración se calculó por medio de la proporción de absorbancia por espectrofotometría con el equipo NanoDrop™ Lite Spectrophotometer (Thermo-Fisher Scientific), y por espectrofluorometría con el equipo Quant-iT PicoGreen dsDNA Assay Kit (Thermo-Fisher Scientific). La pureza del ADN respecto a la relación de absorbancia se determinó con la proporción 260/280 nm (A_{260}/A_{280} para las proteínas) y 260/230 nm (A_{260}/A_{230} para polifenoles). En muestras con valores $260/280 < 1.5$, se aplicó un proceso de purificación con el kit Higher Purity™ DNA

Purification SPRI Magnetic Beads (Canvax Biotech) y posteriormente, las muestras cuya proporción 260/280 continuó siendo < 1.6 fueron extraídas nuevamente.

7.4.2 Genotipificación basada en secuenciación (SBG)

Las muestras de ADN genómico se enviaron al Centro Genómico de la Universidad de Minnesota (CGUM) en EUA, en donde, después de probar un panel de ocho diferentes enzimas de restricción con ocho muestras de ADN (concentración $\geq 2 \mu\text{g}$), de diferentes individuos y del mismo sitio de muestreo, se seleccionaron las enzimas BamHI y NsiI, para el proceso de digestión del ADN genómico, ya que fueron con las que se obtuvo la mayor cobertura del genoma (Elshire *et al.*, 2011).

Posteriormente, se enviaron 80 muestras de ADN genómico (concentración $\geq 0.4 \mu\text{g}$) al CGUM, una por cada individuo colectado (Cuadro 1). Después de digerir el ADN genómico, se crearon las bibliotecas de doble índice, las cuales se combinaron en un solo conjunto y mediante la plataforma Illumina NextSeq 550 se obtuvieron secuencias de 150 pares de bases (pb) de una lectura única (*single-end*).

7.4.3 Ensamble e identificación de los SNPs.

A partir de la información obtenida por la NGS, el CGUM realizó un análisis de ensamble e identificación de polimorfismos de un solo nucleótido (SNPs), utilizando su propia metodología (*pipeline*), la cual en adelante denominaremos análisis CGUM.

De igual forma, a partir de los datos crudos obtenidos por el CGUM, es decir, aquellos sin procesar bioinformáticamente, se realizó un análisis utilizando un *pipeline* propio mediante el programa Stacks (Catchen *et al.*, 2013).

7.4.3.1 Análisis CGUM

El CGUM generó un reporte donde mostró que en promedio se obtuvieron 700 mil lecturas por muestra, con un promedio en puntuaciones de calidad $\geq Q30$. Las secuencias se agruparon por muestra para generar archivos en formato fastq. A partir de estos archivos, el CGUM procesó los datos y se obtuvo una matriz con los SNPs identificados entre las muestras. Para ello, utilizaron los programas Illumina bcl2fastq y Trimmomatic para remover las secuencias de los adaptadores. La calidad de las secuencias se evaluó con el programa

FastQC. Se utilizó el programa bwa para el alineamiento de las secuencias usando como referencia el genoma de *Q. lobata* (Sork *et al.*, 2016a). Con el programa Freebayes se generó un archivo con las variantes identificadas entre las secuencias y entre los individuos (VCF; Variant Call Format, una matriz con los SNPs identificados). Después, mediante el programa VCFTools V. 4.0 se removieron las variantes con una frecuencia alélica mínima < 1%, las variantes con tasas genotípicas < 95%, y las muestras con tasas genotípicas < 50% (Danecek, 2011). Se identificaron 30,141 SNPs entre los 80 individuos.

Selección de individuos

A partir de estos SNPs identificados, se realizó un Análisis de Componentes Principales (ACP), considerando las distancias genéticas entre los individuos, como una prueba exploratoria de los datos, entre los individuos de *Q. macdougallii* y 14 individuos de *Q. mullerii*, un encino de la sección *Lobatae* (Anexo I). Se esperaba que los individuos se separaran por especie, sin embargo, aunque se observó un fuerte agrupamiento entre los individuos de *Q. mullerii*, en los individuos de *Q. macdougallii*, el individuo LS_30, uno de los cinco individuos que se encontraron en el sitio LS, mostró diferencias significativas con respecto a los demás individuos. Esto podría indicar que el individuo es otra especie, parecida morfológicamente a *Q. macdougallii*, que sea un híbrido, o que haya habido algún error durante su secuenciación.

Debido a lo anterior, para este proyecto se decidió trabajar con 79 individuos, y se sugiere que para la correcta identificación del individuo LS_30 se consideren búsquedas en bases de datos genéticos de otros encinos, así como su recolecta para una descripción taxonómica más extensa.

7.4.3.2 Análisis con Stacks

Se aplicó una *pipeline* para realizar un ensamble e identificación de SNPs con el programa Stacks V. 2.0 (Catchen *et al.*, 2013), usando como referencia al genoma de *Q. lobata* (Sork *et al.*, 2016a), y con los archivos en formato *fastq* enviados por el CGUM.

Primero se construyó un genoma híbrido alineando las lecturas obtenidas de la especie contra el genoma publicado de *Q. lobata* para generar un genoma consenso entre las dos especies. Después se utilizó el programa bowtie2 V. 2.3.4.1 (Langmead & Salzberg,

2012), para alinear las lecturas de los 79 individuos con el genoma consenso. Posteriormente, se utilizó el programa Stacks con la *pipeline ref_map.pl*, para alinear hacia un genoma de referencia y posteriormente identificar los SNPs, con los parámetros y valores sugeridos por los creadores del programa (Catchen *et al.*, 2013).

7.4.3.3 Filtrado de SNPs

En ambas *pipelines*, para evaluar los SNPs identificados se aplicaron filtros bioinformáticos con el programa VCFTools V. 4.0, considerando la mínima frecuencia alélica (*maf*) y del valor máximo de datos perdidos (*max-missing*) (http://vcftools.sourceforge.net/man_latest.html; Danecek, 2011).

Las combinaciones de sus valores fueron: *maf* 0.01 y 0.008, para incluir a los SNPs con frecuencias alélicas mayores o iguales al valor, donde la frecuencia alélica se define como el número de veces que aparece un alelo para todos los individuos en ese locus, dividido por el número total de alelos en ese locus; *max-missing* 0.9 y 1, donde 1 es para remover todos los datos faltantes. Se observó que el número de SNPs obtenidos no se alteró al cambiar la mínima frecuencia alélica, por lo cual se decidió utilizar el valor de 0.01 pero con dos enfoques. El primero permitiendo un 10% de datos faltantes (enfoque laxo, *max-missing*=0.9), y el segundo removiendo todos los datos faltantes (enfoque estricto, *max-missing*=1) (Anexo II).

Al visualizar la calidad de las matrices de SNPs y sus datos faltantes (Anexo III), se consideró que las matrices generadas con la *pipeline* de Stacks no fueron lo suficientemente robustas para los siguientes análisis y que se necesitan más pruebas para evaluar su confiabilidad en la correcta identificación de las variantes.

También, debido al riesgo de identificar falsos positivos mediante el enfoque laxo y para tener una mayor certeza en cuanto a la identificación de las variantes, para este proyecto se consideró utilizar la matriz del análisis del CGUM con el enfoque estricto (8,186 SNPs).

7.4.4 Genómica poblacional

7.4.4.1 Diversidad genética

Se calculó la heterocigosidad observada (H_o), heterocigosidad esperada (H_e), la D de Tajima, y la diversidad nucleotídica π y poblacional θ con el programa TASSEL-GBS V. 5.0 (Glaubitz *et al.*, 2014), para cada sitio y para el conjunto total de individuos. La prueba D de Tajima es una prueba de neutralidad, la cual, puede detectar eventos evolutivos dentro de una población (Tajima 1989; Zhang *et al.*, 2015).

Alelos totales y únicos

Se calculó el número de alelos totales y únicos por sitio de muestreo con el programa R V.3.5 con los paquetes *poppr* y *adegenet*, con la función *private_alleles* (Jombart & Ahmed, 2011). Para el número de alelos únicos se consideró el número de alelos a partir de SNPs bialélicos, por lo cual, si en un sitio de muestreo todos los individuos poseen el mismo alelo, es decir (T:T), y otro sitio de muestreo tiene dos individuos donde se observa un cambio de bases (C:T), el número de alelos únicos para el primer sitio es cero, mientras que para el segundo sitio es dos, ya que en dos individuos, se ha observado otra variante a parte de la timina, aportan dos citosinas.

Diversidad y diferenciación poblacional

Se calculó la diversidad y diferenciación poblacional, en el programa R V. 3.5.3 con el paquete *vcfR* y la función *genetic_diff* (Knaus & Grünwald, 2017), considerando de manera artificial los ocho sitios de muestreo como poblaciones. El método de Nei empleado para este análisis se basa en los métodos reportados por Hedrick (2005).

La diferencia es que las heterocigosidades se ponderan por el número de alelos observados en cada población, lo cual se basa en el cálculo de la F_{ST} por pares (pairwise F_{ST}) de Nei, que utiliza el número de individuos observados en cada sitio para ponderar las heterocigosidades. Al usar la cantidad de alelos observados en lugar de la cantidad de individuos, se elimina la suposición acerca de cuántos alelos puede contribuir cada individuo. Se calculó la heterocigosidad por loci entre los ocho sitios de muestreo, así como la G_{ST} de Nei (Nei, 1987) y la G'_{ST} de Hedrick (Hedrick, 2005).

7.4.4.2 Estructura genética poblacional

Distancias genéticas

Se construyó un árbol con el método Neighbor-joining a partir de una matriz con las distancias genéticas de Nei (Nei, 1987) entre los individuos. La similitud entre individuos y grupos de individuos está representada por la longitud de la rama, la cual, está representada por el número de sustituciones por sitio para un grupo o individuo. Cuando los individuos o grupos son muy similares, se agrupan por ramas cortas. Por ello, cuanto más larga sea la rama, mayor será el número de sustituciones y mayor será la distancia genética entre los individuos o los grupos. Estos análisis se realizaron en el programa R V. 3.5.3, con los paquetes y funciones *poppr::aboot* y *dartR::gl.tree.nj*, considerando 100 réplicas con el método de remuestreo Bootstrap.

Red de expansión mínima

También se realizó un análisis de red de expansión mínima (MSN), el cual agrupa genotipos multilocus (MLG) por distancias genéticas entre ellos. Cada MLG es un nodo y la distancia genética es representada por las aristas (líneas que conectan los nodos). Al basarse en datos de secuenciación de alto rendimiento y con una densidad alta de marcadores moleculares, cada individuo se relaciona a un genotipo único. Los nodos se encuentran conectados por la distancia mínima entre los individuos, lo que permite observar partes reticulares, es decir, un conjunto de nodos conectados entre ellos por distancias genéticas idénticas (Kamvar *et al.*, 2014).

Estos análisis se realizaron en el programa R V. 3.5.3, con el paquete *poppr* y las funciones *bitwise.dist* y *poppr.msn*, a partir de las distancias de disimilitud de Hamming (https://grunwaldlab.github.io/Population_Genetics_in_R/Minimum_Spanning_Networks.html; <https://www.rdocumentation.org/packages/poppr/versions/2.8.3/topics/bitwise.dist>).

Análisis de Componentes Principales (ACP)

El ACP o PCA por *Principal Component Analysis*, es un método estadístico de análisis multivariante, el cual se utiliza para el tratamiento de datos compuestos por numerosas variables con el objetivo de reducir su dimensionalidad, conservar la mayor parte de su

variación y perder la menor cantidad de información posible mediante la identificación de componentes principales (CPs) (Peres-Neto *et al.*, 2005).

Los vectores propios (eigenvectores) y los valores propios (eigenvalores) de una matriz de covarianza (o correlación) representan el núcleo de un ACP. Los eigenvectores o componentes principales (CPs) determinan las direcciones de las nuevas características en el espacio, y los eigenvalores determinan su magnitud. En otras palabras, los eigenvalores explican la varianza de los datos a lo largo de los ejes de las nuevas características.

Los resultados obtenidos son utilizados para explorar las similitudes o diferencias significativas entre los individuos, lo que permite visualizar patrones de agrupamiento entre el conjunto de datos (Ringnér, 2008). Este análisis se realizó en el programa R V. 3.5.3, con el paquete *adegenet* y la función *glPca* (<https://rdrr.io/cran/adegenet/man/glPca.html>). A partir de los valores obtenidos en cada eigenvalor, se seleccionaron los tres componentes principales con los valores más altos para graficarlos.

Análisis Discriminante de Componentes Principales (ADCP)

Se realizó un ADCP o Discriminant Analysis of Principal Components (DAPC). Uno de los principales atributos de este análisis es que puede realizarse en ausencia de cualquier suposición sobre el modelo genético de la población subyacente, en particular con respecto al equilibrio de Hardy-Weinberg o el equilibrio de ligamiento. El ADCP se usó para identificar y describir grupos de individuos genéticamente relacionados. Este análisis transforma los datos utilizando un ACP y luego realiza un análisis discriminante en los componentes principales (CPs) retenidos utilizando un método de validación cruzada. Este método multivariado es adecuado para analizar grandes cantidades de SNP de todo el genoma y proporciona la asignación de individuos a grupos, así como una evaluación visual de la diferenciación entre poblaciones.

Se utilizó el procedimiento de optimización propuesto por el paquete *adegenet* de R para evaluar el número óptimo de CPs a retener, con 1,000 repeticiones utilizando un 90% de los datos como un conjunto de datos de entrenamiento. El procedimiento de validación cruzada implementado con la función *xvalDapc* realiza una validación cruzada estratificada del ADCP utilizando un número variable de CPs, y manteniendo el número de funciones discriminantes fijas (https://grunwaldlab.github.io/Population_Genetics_in_R/DAPC.html).

El número de componentes principales se eligió en función de los criterios para producir el más alto porcentaje promedio de reasignación exitosa y el error cuadrático medio más bajo (Jombart *et al.*, 2010).

fastStructure y ADMIXTURE

Para caracterizar la estructura genética también se utilizaron los enfoques evolutivos de agrupamiento, basados en los modelos de los programas bioinformáticos: fastStructure (Raj *et al.*, 2014) y ADMIXTURE (Alexander *et al.*, 2009). fastStructure utiliza la inferencia aproximada del modelo bayesiano del programa STRUCTURE (Pritchard *et al.*, 2000) mientras que ADMIXTURE calcula las estimaciones de máxima probabilidad de los parámetros para estimar el número más probable de grupos evolutivos, K.

Se utilizó el programa fastStructure V. 1.0 para cada valor de K, de 1 a 4, utilizando el modelo estándar simple (*simple prior*). Se utilizó también una secuencia de comandos de fastStructure (*choiceK.py*) para calcular las puntuaciones heurísticas y así detectar el rango de los valores más probables de K. Después de identificar los valores más probables de K, se volvió a ejecutar fastStructure para cada K, con el modelo logístico (*logistic prior*), el cual es más exigente computacionalmente.

Con el programa ADMIXTURE V. 1.3.0, se volvió a analizar la matriz de SNPs, para cada valor de K, de 1 a 4. Los valores más probables de K se determinaron utilizando su procedimiento de validación cruzada (Alexander & Lange, 2011). Para visualizar y comparar los resultados de agrupación de fastSTRUCTURE y ADMIXTURE, se utilizaron los paquetes *ggplot2*, *eeptools*, *tidyr* y *plyr* en el programa R V. 3.5.3.

Varianza molecular y coeficientes de fijación (F_{IS} y F_{ST})

El análisis de varianza molecular (AMOVA, por sus siglas en inglés), es un método estadístico que permite cuantificar la diversidad genética mediante pruebas estadísticas F en diferentes niveles jerárquicos, de manera que la diversidad dentro del grupo se minimiza mientras que la diversidad entre los grupos se maximiza. Este análisis también puede ayudar a detectar posible diferenciación en un conjunto de datos genéticos.

El AMOVA se realizó con el paquete *PopGenome* en el programa R V. 3.5.3. Se probaron desviaciones significativas de la estructura poblacional aleatoria en AMOVA

utilizando la función *randtest* del paquete *ade4* con 9,999 repeticiones con el método de remuestreo Bootstrap (Dray & Dufour, 2007; <https://grunwaldlab.github.io/poppr/reference/poppr.amova.html>).

El coeficiente de endogamia F_{IS} , es una medida de la desviación de las frecuencias genotípicas de las esperadas por panmixia en términos de deficiencia o exceso de heterócigos. F_{IS} puede variar entre -1 y 1. Valores cercanos a -1 indican que hay un exceso de heterócigos, valores cercanos a 1 indican que hay un exceso de homócigos y valores cercanos a 0 indican que las poblaciones están en equilibrio Hardy-Weimberg (Wright, 1943).

El coeficiente de coancestría F_{ST} , compara el nivel de variación genética que existe dentro de dos o más subpoblaciones. Se define como la probabilidad de asociación de gametos idénticos dentro de las subpoblaciones con respecto al sorteo aleatorio de gametos de toda la población (Wright, 1943).

Si el apareamiento en la población total es aleatorio, entonces se espera que T y S tengan el mismo valor y F_{ST} será 0, lo que indica identidad genética entre las subpoblaciones. Sin embargo, si el apareamiento entre poblaciones no es aleatorio, la variación genética en cada subpoblación será menor que la variación en la población total y el valor de F_{ST} será de 1, lo que indica completa diferenciación entre las subpoblaciones. Por lo tanto, F_{ST} es una medida de la desviación del apareamiento aleatorio causado por la estructura genética de la población. Aunque F_{ST} es informativo sobre las diferencias promedio dentro de las poblaciones y entre ellas, nos dice poco sobre la consistencia o varianza de estas diferencias (Bamshad *et al.*, 2004).

Se calcularon los coeficientes de fijación, F_{IS} y F_{ST} , para todos los individuos entre los ocho sitios de muestreo, con las funciones *hierfstat::pairwise.fst* y *StAMPP::stamppFst* en el programa R V. 3.5.3.

7.4.4.4 Identificación de SNPs *outliers*

Los SNPs *outliers* se identificaron como los SNPs con los valores más altos de F_{ST} . Los valores se obtuvieron después de calcular el F_{ST} para cada SNP entre las dos zonas (Norte y Sur), después se seleccionó el 1% de los SNPs con los valores más altos (Caldu-Primo *et al.*, 2017). Se calcularon las distancias genéticas de Nei y se realizó un ACP.

A partir de los ocho *outliers* con los valores más altos, se calcularon las frecuencias alélicas entre los ocho sitios de muestreo. Se utilizaron los paquetes SNPStats (Clayton, 2015) y adegenet (Jombart & Ahmed, 2011) en el programa R V. 3.5.3.

7.5 Distribución potencial

La distribución potencial de *Q. macdougallii*, se generó con el programa MaxEnt V. 3.4.0 (Phillips *et al.*, 2006), de disponibilidad libre (http://biodiversityinformatics.amnh.org/open_source/maxent/), el cual permite analizar la relación entre las ubicaciones de las especies y ciertas características de su ambiente, para determinar la idoneidad general de su hábitat.

El modelo genera una estimación de la probabilidad de la presencia de la especie, que va desde cero a uno, donde cero representa la probabilidad más baja y uno la más alta (Phillips *et al.*, 2006; Yanga *et al.*, 2013). Este procedimiento ofrece una alta precisión y es particularmente adecuado para especies con pocos registros de datos (Elith *et al.*, 2006).

7.5.1 Datos de presencia

Para la recopilación de datos de presencia de la especie, coordenadas geográficas, se realizó una búsqueda en la base de datos de GBIF (www.gbif.org), en la base personal de Molina-Garay (2011) y en el portal de datos abiertos del Herbario Nacional del Instituto de Biología de la UNAM (<http://www.ib.unam.mx/botanica/herbario/>). A partir de estas bases se recopilaron 11 coordenadas geográficas de presencia de *Q. macdougallii*. Estos datos se cotejaron con las coordenadas de los 80 individuos georreferenciados en este trabajo.

7.5.2 Variables ambientales: CONABIO y WorldClim

Para evaluar las diferencias entre el uso de variables con datos a nivel nacional y mundial, se realizó un modelo utilizando como variables a 10 capas vectoriales del Geoportal de la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO; <http://www.conabio.gob.mx/informacion/gis/>) y otro modelo utilizando las 19 capas bioclimáticas del proyecto WorldClim (Hijmans *et al.*, 2005; Fick & Hijmans, 2017).

Las capas de CONABIO, se obtuvieron en formato ESRI, con el datum WGS84. La información en que se basan proviene de estaciones meteorológicas a lo largo del país. Las

capas bioclimáticas de WorldClim son resultados de una interpolación espacial de datos climáticos provenientes de una serie de estaciones meteorológicas, recabados de 1960 a 1990 (Hijmans *et al.*, 2005). Estas capas en formato raster (un tipo de formato SIG, es decir, una matriz de datos georreferenciados) se obtuvieron con una resolución espacial de 30 arc-seg (~1 km).

Las capas se procesaron conforme al área de la región de accesibilidad (M) (Hortal *et al.*, 2012; Soberón *et al.*, 2017), la cual se determinó como el área de las provincias biogeográficas de Oaxaca y de la Sierra Madre del Sur (CONABIO, 1997; Espinosa *et al.*, 2008). La selección de esta área fue debido a la dispersión histórica del género *Quercus*, desde el norte hacia el Sur del país (Hipp *et al.*, 2018).

Selección de las variables

Con el fin de seleccionar las variables no correlacionadas y que contribuyeron significativamente al modelo, se realizó un análisis de correlación para cada conjunto de variables utilizando el coeficiente de correlación de Pearson, y se generaron modelos de distribución potencial utilizando todas las variables para evaluar su contribución al modelo.

Los análisis de correlación se realizaron con el programa R V. 3.5.3 y Rstudio V. 1.2.1335, con los paquetes *raster*, *rgdal*, *dismo*, *rJava*, *sp*, *maptools*, *spocc*, *rgbif*, *rgeos*, *usdm*, *foreign* y *corrplot*. El procesamiento de las capas se realizó con los programas ArcMap V. 10.5 y QGIS V. 3.6.3-Noosa.

7.5.3 Modelo de distribución potencial actual

Los modelos se realizaron con el programa Maxent V. 3.3.3 (Phillips *et al.*, 2006), a partir de la región de accesibilidad (M), para después proyectar los resultados a nivel país. Se consideraron 100 réplicas, 500 iteraciones, el método de remuestreo Bootstrap, se eliminaron los registros duplicados de presencias y se consideró un muestreo aleatorio diferente de los puntos de referencia para cada réplica (Ponce-Reyes *et al.*, 2012).

Después, mediante un análisis de la prueba de Jackknife, la cual consiste en retirar secuencialmente y con reemplazo de un par de datos de una base de n pares de datos correspondientes, que en este caso son las variables, se determinó la influencia de dichas variables en el modelo de distribución potencial (Phillips *et al.* 2006).

7.5.4 Distribución con escenarios futuros: 2050 y 2070

El proyecto Worldclim (Hijmans *et al.*, 2005) también dispone de capas en formato GeoTIFF, con información sobre sus variables bioclimáticas para cuatro diferentes escenarios de cambio climático (RCP2.6, RCP4.5, RCP6.0 y RCP8.5), para los años, 2050 y 2070. Estas capas son la combinación de información sobre el cambio climático proveniente de modelos de circulación global (GCM) y las capas provenientes de una interpolación que describen las características actuales (Varela *et al.*, 2015).

Se realizó una proyección futura con dos escenarios (RCP4.5 y RCP8.5) de emisiones de gases de efecto invernadero (CO₂, CH₄, N₂O y F), que influyen en el cambio climático, correspondientes a los años 2050 y 2070 (Stocker, 2013), usando el modelo de circulación global MIROC-ESM (Watanabe *et al.*, 2011).

El escenario RCP4.5 plantea un aumento en la temperatura de 0.9 a 2°C para el periodo del 2046 al 2065, y de 1.1 a 2.6°C para el 2081 al 2100. Por otro lado, el escenario RCP8.5 plantea un aumento en la temperatura de 1.4 a 2.6°C para el 2046 al 2065, y de 2.6 a 4.8°C para el 2081 al 2100.

Los datos se procesaron en el programa Maxent V. 3.3.3 (Phillips *et al.*, 2006), con 100 réplicas, 500 iteraciones, utilizando el método de remuestreo Bootstrap, removiendo registros de presencia duplicados y tomando en consideración un muestreo aleatorio diferente de los puntos de referencia para cada réplica (Ponce-Reyes *et al.*, 2012).

7.5.5 Evaluación del modelo

El ajuste del modelo fue calificado con el criterio de variación bajo la curva (AUC) de la curva ROC (análisis de sensibilidad y especificidad) (Phillips *et al.*, 2006). El AUC mide la habilidad del modelo de discriminar entre sitios donde la especie está presente, versus donde está ausente.

El AUC fluctúa entre cero y uno, en donde el valor de uno indica perfecta discriminación, un valor de 0.5 indica que la discriminación predictiva entre sitios no es mejor que una selección al azar, y un valor menor de 0.5 indica un desempeño inferior que el azar (Pearce & Ferrier, 2000; Araujo & Guisan, 2006).

Adicionalmente, mediante un análisis de la prueba de Jackknife, se determinó la contribución relativa de cada variable para su respectivo modelo generado. Los resultados son expresados con una medida llamada “ganancia”, esta información es de suma importancia para evidenciar los requerimientos ecológicos de la especie (Phillips *et al.* 2006).

7.6 Variación genómica, geografía y ambiente

Para observar si existe una relación entre las distancias geográficas y las distancias genómicas se realizó una prueba de Mantel (Mantel, 1967), la cual, mediante un análisis de correlación entre dos matrices simétricas, se obtiene el coeficiente de correlación de Pearson.

A partir de las coordenadas de cada sitio de muestreo se generó una matriz de distancias Euclidianas y la matriz de distancias genéticas se generó a partir de la medida del flujo génico ($F_{ST}/1-F_{ST}$) entre los ocho sitios de muestreo.

Para analizar las relaciones entre la variación genómica y los factores geográficos y ambientales, se realizó una prueba parcial de Mantel, la cual, analiza la correlación entre las tres matrices. Los datos ambientales para cada sitio se obtuvieron como promedios a partir las mismas variables utilizadas en el modelo de la distribución potencial.

La prueba Mantel se realizó con el paquete *dartR* (<https://rdrr.io/cran/dartR/man/gl.ibd.html>) y la prueba parcial de Mantel con el paquete *ncf* (<https://www.rdocumentation.org/packages/ncf>).

8. RESULTADOS

8.1 Variación genómica

Se identificaron 8,186 SNPs entre los 79 individuos de los ocho sitios de muestreo de *Q. macdougallii*; cinco sitios distribuidos en la zona Norte y tres en la zona Sur del área de estudio.

8.1.1 Diversidad genética

Al analizar el conjunto total de SNPs a través de los ocho sitios y los 79 individuos, se encontró una diversidad nucleotídica $\pi = 0.200$ y poblacional $\theta = 0.202$, y una D de Tajima = -0.036.

En el Cuadro 2 se muestra la descripción genética por sitio. La heterocigosidad observada (H_o) es similar en todos los sitios (~0.21) y fue mayor a la heterocigosidad esperada (H_e), lo cual sugiere un leve exceso de heterocigos. Se observó una diversidad nucleotídica parecida entre los ocho sitios (~0.2), siendo en MT donde se observó el valor más bajo (0.194), y en MC el valor más alto (0.2003). Los valores observados de la prueba D de Tajima para los ocho sitios fueron negativos, la mayoría cercanos a -1, con excepción del sitio LS que mostró el valor más bajo (-2.886) y el sitio PZ con el valor más alto (-0.422).

Cuadro 2.- Descripción genética de los ocho sitios. Donde H_o = heterocigosidad observada, H_e = heterocigosidad esperada, $SegSit$ = sitios segregantes, π = diversidad nucleotídica, θ = diversidad nucleotídica poblacional y D = prueba D de Tajima.

Sitios	H_o	H_e	$SegSit$	π	θ	D
CZ	0.2139	0.1888	5576	0.1979	0.2408	-0.8942
MT	0.2152	0.1853	5372	0.194	0.232	-0.8219
MC	0.2158	0.1909	5643	0.2003	0.2437	-0.8951
MB	0.2153	0.1869	5465	0.1958	0.236	-0.8547
CY	0.2158	0.1874	5444	0.1962	0.2351	-0.8307
LS	0.2142	0.175	4103	0.1979	0.2734	-2.8862
CR	0.2178	0.1894	5518	0.1983	0.2383	-0.8424
PZ	0.2152	0.1893	5739	0.1952	0.2156	-0.422

Se calculó la diversidad considerando de manera artificial los ocho sitios de muestreo como poblaciones. Se analizó la heterocigosidad por locus y entre los sitios de muestreo (Fig. 4). Se observan bajos niveles de heterocigosidad, la mayoría acumulada en valores menores a 0.1.

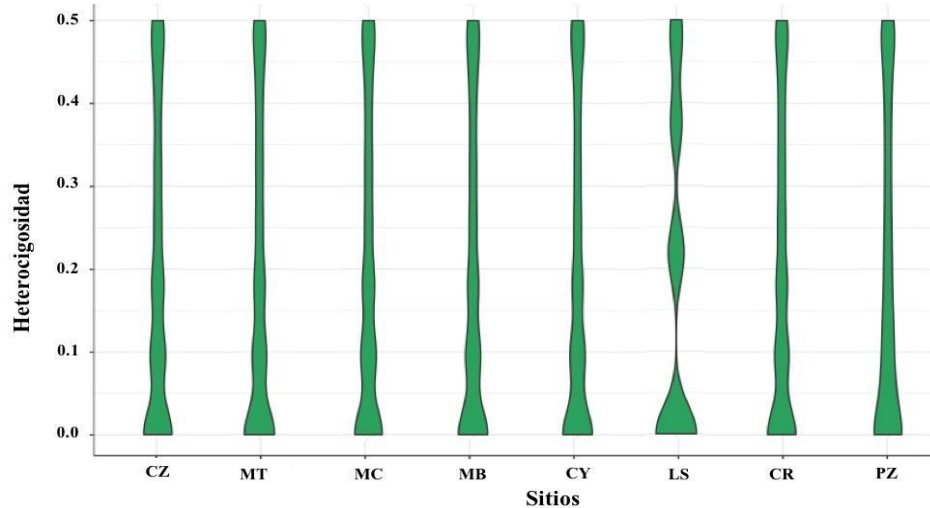


Figura 4.- Diagrama de violín, mostrando la acumulación y distribución de la heterocigosidad por locus.

El análisis de los alelos únicos y totales por locus, para cada sitio de muestreo se observa en el Cuadro 3. Los dos sitios ubicados en los extremos del área de estudio, PZ (Sur) y CZ (Norte), presentan el mayor número de alelos únicos. En ambos sitios también se encontraron a los individuos de la especie con las mayores altitudes, por encima de los 3,000 msnm. Después, los sitios de mayor a menor número de alelos únicos son: MT, MB, CY, MC, CR y LS. En cuanto al número de alelos totales, PZ es el sitio con el mayor valor, seguido de MC, CZ, CR, MB, CY, MY y LS.

En la Fig. 5, se representan los alelos únicos por sitio. El eje Y indica la localización de cada locus en el ensamble que realizó el CGUM, a partir de las secuencias de cada individuo y usando como referencia el genoma de *Q. lobata*.

Para el sitio PZ se observa el mayor número de alelos únicos en comparación con los demás sitios, además de una distribución más homogénea de sus alelos. Por otra parte, los sitios MC y CR presentan una sección de SNPs sin alelos únicos, lo cual, podría indicar que estos dos sitios mantienen alelos similares en los mismos loci. En el sitio LS, también se

observa una amplia sección de SNPs sin alelos únicos, los cuales, además, en el ensamble se encuentran cercanos entre sí.

Cuadro 3.- Número de alelos totales y únicos por sitio. Entre paréntesis, se ordenan de mayor a menor número de alelos.

Sitios	Alelos totales	Alelos únicos
CZ	13,563 (3)	178 (2)
MT	13,361 (7)	159 (3)
MC	13,628 (2)	119 (6)
MB	13,451 (5)	134 (4)
CY	13,425 (6)	133 (5)
LS	12,102 (8)	78 (8)
CR	13,499 (4)	90 (7)
PZ	13,722 (1)	370 (1)

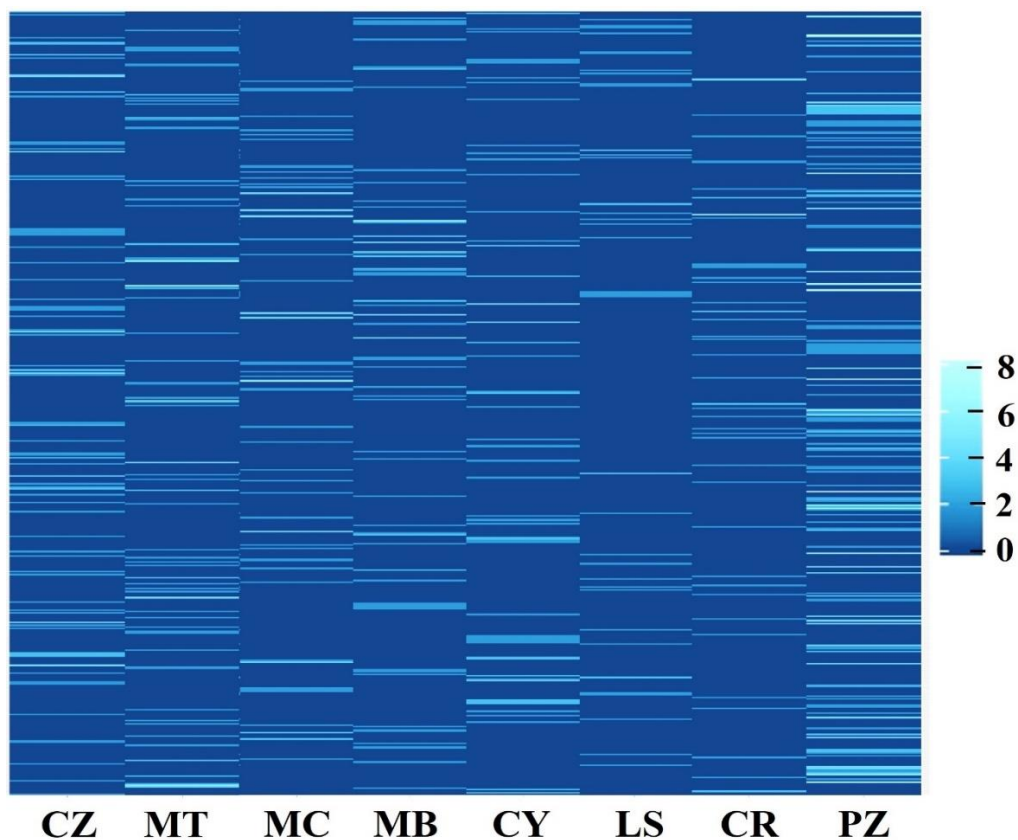


Figura 5.- Representación gráfica de alelos únicos en los ocho sitios de muestreo (eje X) y por cada locus (eje Y).

Los sitios PZ y LS presentan el mayor y el menor número de alelos totales y únicos, sin embargo, esto puede estar fuertemente influenciado por el número de individuos que se analizaron (PZ=15 y LS=4).

Descartando a estos dos sitios, en cuanto al número de alelos únicos, es el sitio CZ quien presenta el mayor número (178) y el sitio con el menor número fue CR (90). Sin embargo, considerando a los alelos totales el sitio con la mayor cantidad fue MC y el sitio con la menor cantidad fue MT. Por lo cual, no se observa una relación entre mayor número de alelos totales y mayor número de alelos únicos.

8.1.2 Estructura genética poblacional

Los valores de G_{ST} de Nei y G'_{st} de Hedrick para todo el conjunto de SNPs, fueron 0.063 y 0.081 respectivamente, lo que indica valores bajos de diferenciación poblacional.

Para analizar las distancias genéticas, se construyeron árboles a partir de las distancias genéticas de Nei (1987). Se observó su comportamiento entre sitios (Fig. 6) y entre individuos (Fig. 7).

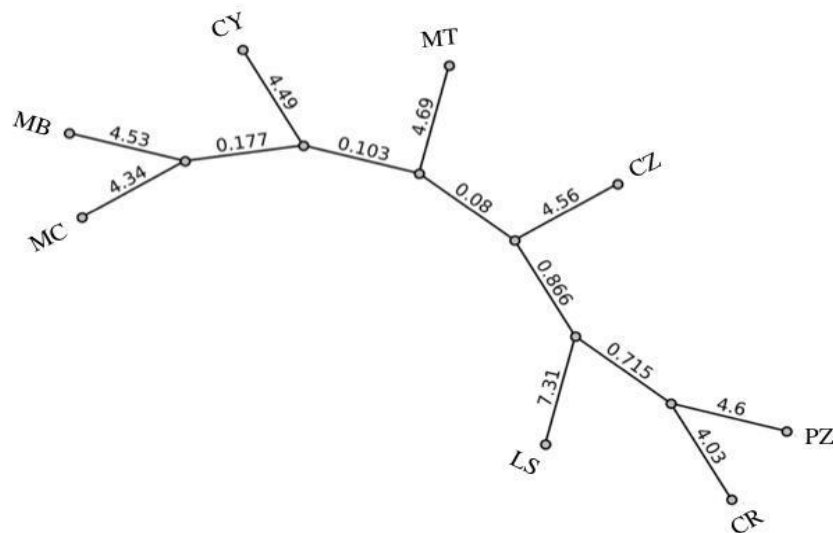


Figura 6.- Árbol sin raíz construido a partir de distancias genéticas de Nei, por el método Neighbor-joining. Los números sobre las ramas representan las sustituciones nucleotídicas.

Las distancias genéticas entre los sitios de la zona Norte son similares. Los sitios MC y MB, así como PZ y CR, son geográficamente los más cercanos entre ellos, lo que también se observa en sus distancias genéticas. El sitio con la mayor diferencia en sus distancias es LS, ya que, aunque se agrupa con los demás sitios de la zona Sur, su valor de sustituciones nucleotídicas es el más alto de todos (7.31).

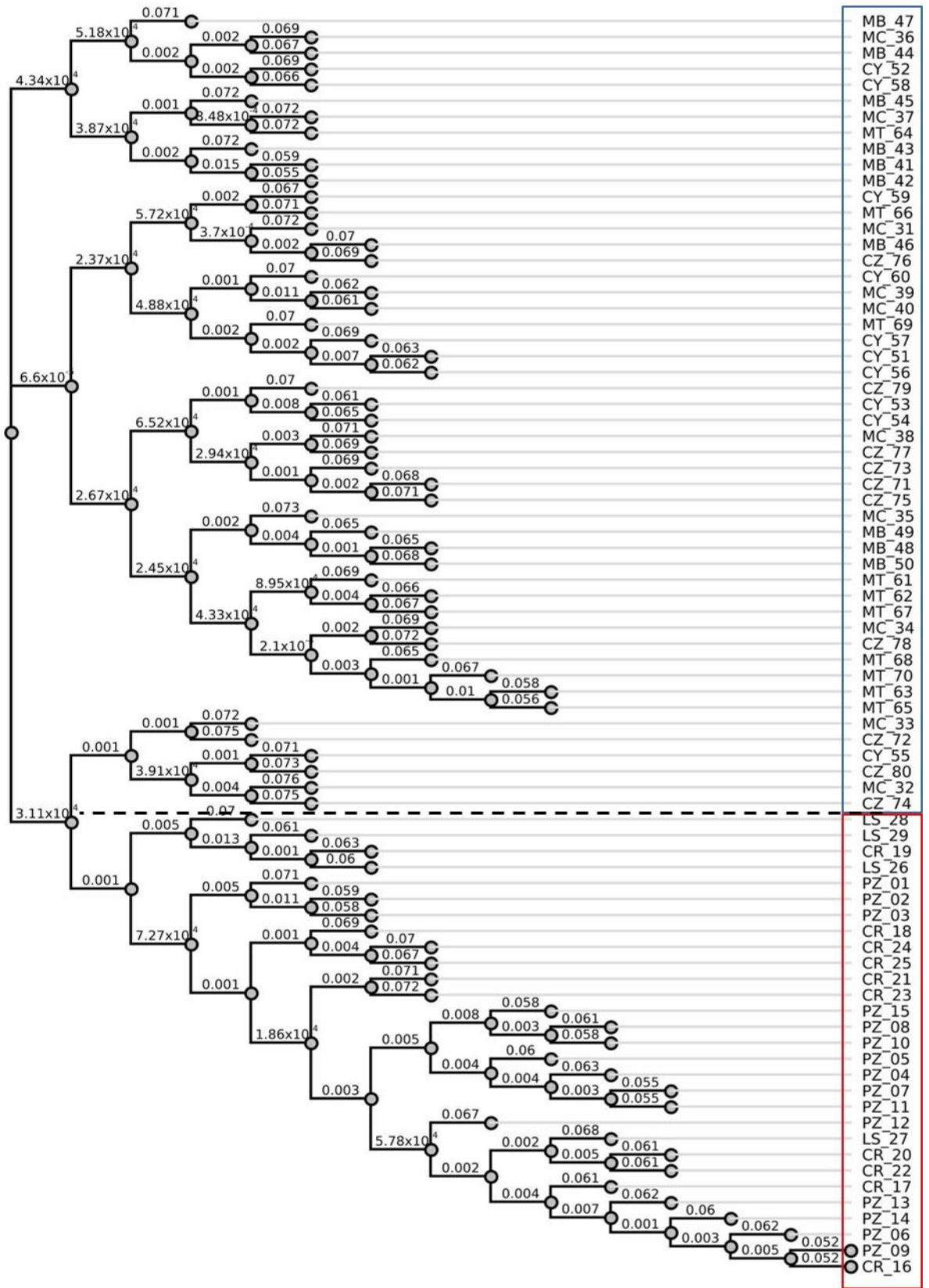


Figura 7.- Árbol construido a partir de distancias genéticas de Nei, por el método Neighbor-joining. En azul se señalan los individuos de la zona Norte y en rojo los de la zona Sur. Los números sobre las ramas representan las sustituciones nucleotídicas.

A nivel de individuos, podemos observar tres grupos, dos de ellos abarcan a los individuos de la zona Norte, mientras que el otro agrupamiento abarca a todos los individuos de la zona Sur, junto con seis individuos de la zona Norte: MC_32 y MC_33 (sitio MC); CY_55 (sitio CY); CZ_72, CZ_74 y CZ_80 (sitio CZ).

En el análisis de red de expansión mínima (Fig. 8), donde se optimizó la visualización de las relaciones entre los individuos basadas en las distancias genéticas, no se identificó individuos clonales. Asimismo, se observa que el individuo CR_16 conecta dos ramificaciones que se forman entre los individuos del Sur.

Por otro lado, se observa que ciertos individuos de la zona Sur presentan distancias genéticas más cercanas hacia otros individuos de la zona Norte. Por ejemplo, los individuos CR_19, LS_26, LS_28 y LS_29, se agrupan en un extremo de la red, más cerca a los individuos del sitio MT, mientras que los individuos PZ_01, PZ_02 y PZ_03 se ubican más cerca a los sitios MC y MB.

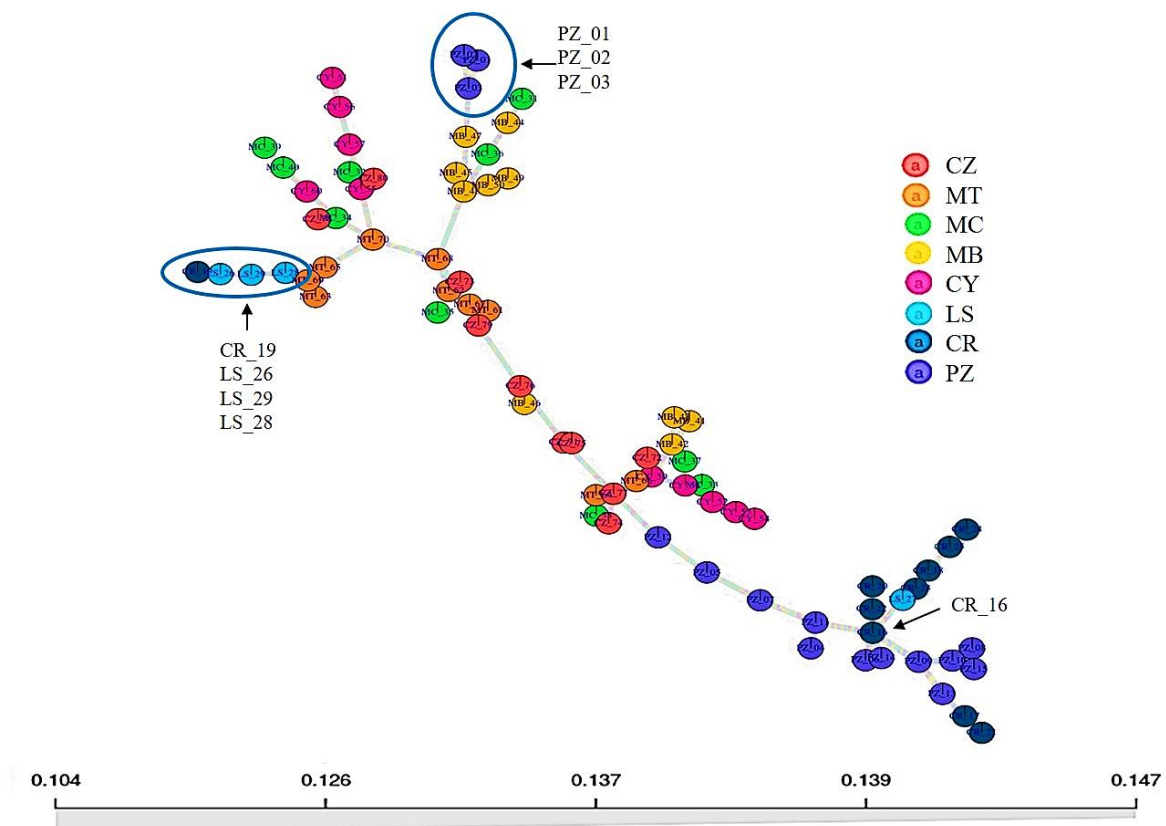


Figura 8.- Red de expansión mínima a partir de distancias genéticas de Nei.

En el ACP (Fig. 9), se observa una tendencia de separación entre los individuos de la zona Norte y los del Sur. Mientras que los individuos de la zona Norte se agrupan fuertemente entre ellos, los de la zona Sur se distribuyen de manera dispersa. Por otro lado, el primer CP separa claramente a los individuos de la zona Norte, agrupando también a los individuos LS_26, LS_28, LS_29 y CR_19 de la zona Sur. Estos mismos individuos también son aislados por el segundo CP junto con otros de los sitios MC y CZ.

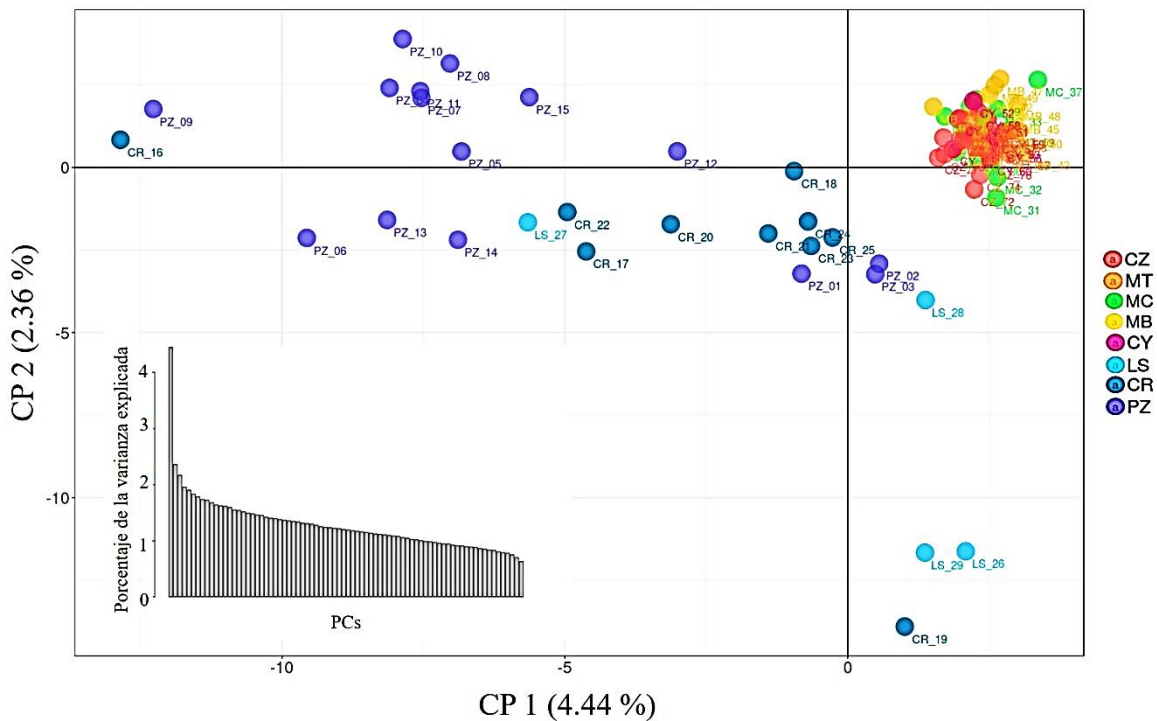


Figura 9.- Análisis de componentes principales (ACP), entre los ocho sitios de muestreo, considerando los dos primeros componentes principales.

Debido a que la varianza acumulada por los eigenvalores en el ACP no mostró un comportamiento de parábola, lo cual es necesario para seleccionar el número óptimo de CPs a utilizar, previo al Análisis Discriminante de Componentes Principales (ADCP) se llevó a cabo un análisis de validación por cruzamiento, donde se obtuvo 10 CPs como un valor óptimo para el ADCP (Anexo IV).

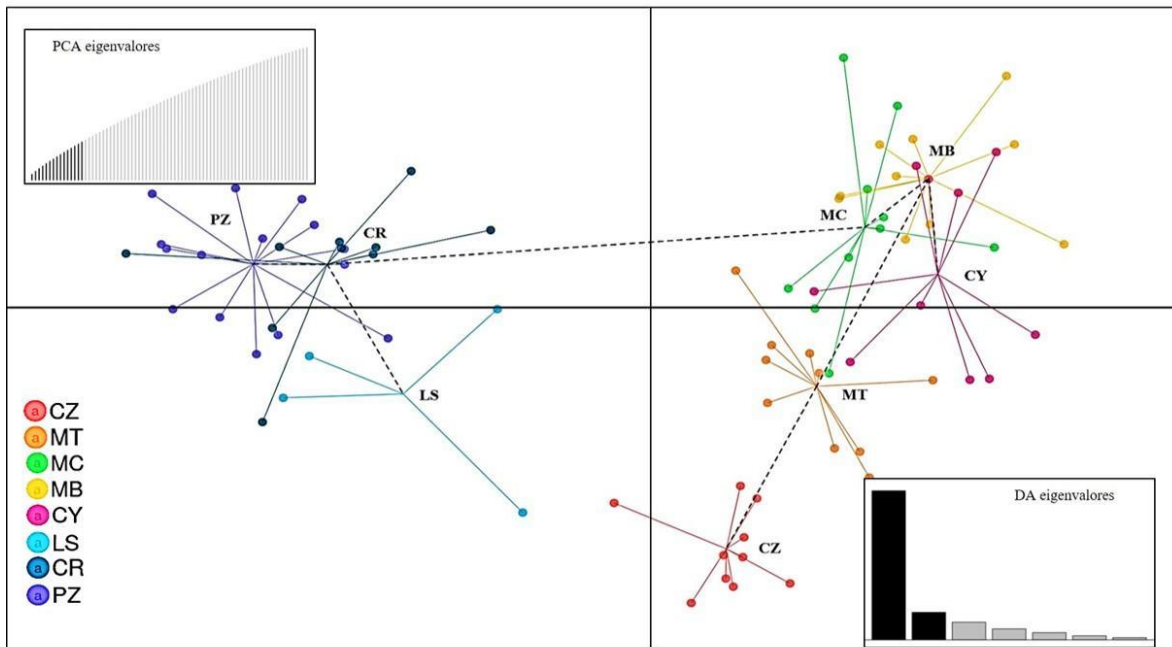


Figura 10.- Análisis Discriminante de Componentes Principales (ADCP), entre los ocho sitios, con 10 componentes principales.

En el ADCP (Fig. 10), se observa la separación entre los sitios de la zona Norte y Sur. Con la red de expansión mínima, la cual representa la proximidad entre los sitios, se observa que la conexión entre las zonas está dada por los sitios CR y MC. También, el sitio CR es un conector entre los otros dos sitios de la zona Sur, PZ y LS. Por otro lado, en la zona Norte, los sitios MC, MB y CY presentan una fuerte cercanía entre ellos, mientras que los sitios MT y CZ tienden hacia un extremo. Ambos análisis, el ACP y el ADCP, mostraron un patrón de agrupamiento similar.

Los dos enfoques evolutivos de agrupamiento considerados para el análisis de la estructura poblacional, mediante la inferencia aproximada del modelo bayesiano y las estimaciones de máxima probabilidad, se apoyan de los programas fastStructure y ADMIXTURE. Ambos enfoques mostraron un valor óptimo de grupos entre uno y dos.

Con fastStructure, el modelo simple indica que los individuos tienden a comportarse como un solo grupo, mientras que el logístico indica dos posibles grupos. Por otro lado, con ADMIXTURE, el error de validación por cruzamiento para $K=1$ y 2 , muestran los valores más bajos, 0.41 y 0.42 respectivamente (Anexo V).

Los valores obtenidos de cada análisis se graficaron considerando $K=2$ (Fig. 11). De acuerdo con el modelo simple de fastStructure, se observa que todos los individuos forman un solo grupo, sin embargo, en el modelo logístico y ADMIXTURE se observan dos grupos, donde la mayoría de los individuos de la zona Sur presentan altas probabilidades de pertenencia a un segundo agrupamiento.

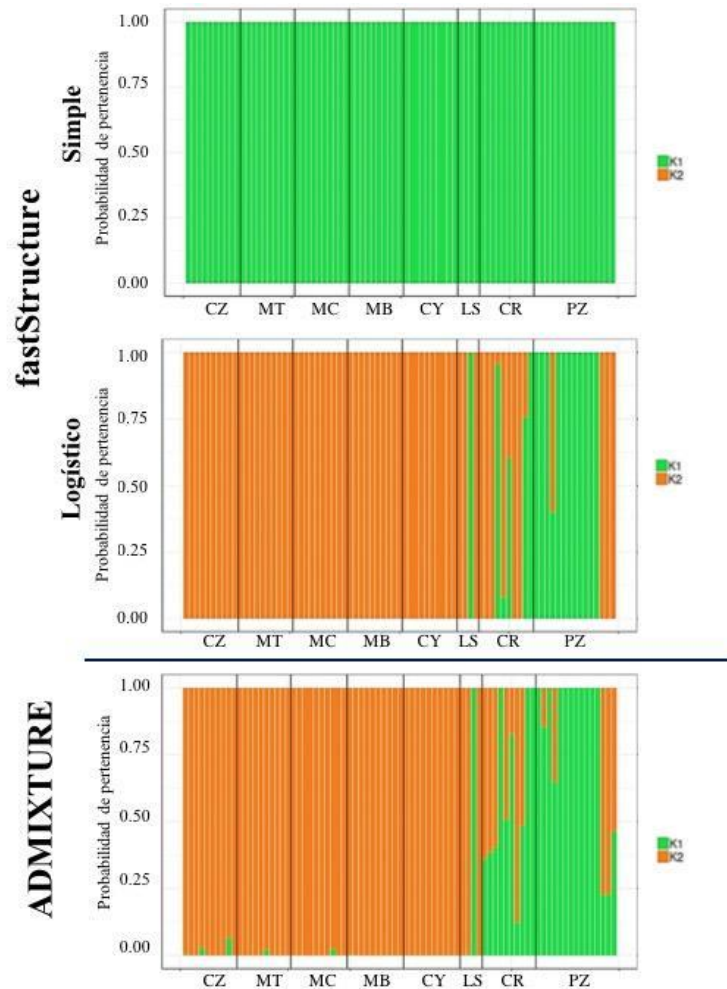


Figura 11.- Diagramas que representan la estructura poblacional asumiendo un agrupamiento de $K=2$.

Los promedios de probabilidad de pertenencia por sitios, de ADMIXTURE, se visualizaron sobre una imagen satelital del área de estudio (Fig. 12). En la zona Sur se presenta dos agrupamientos de individuos en cada sitio, y conforme disminuye la latitud (Norte a Sur) aumenta la probabilidad de pertenencia al segundo grupo.

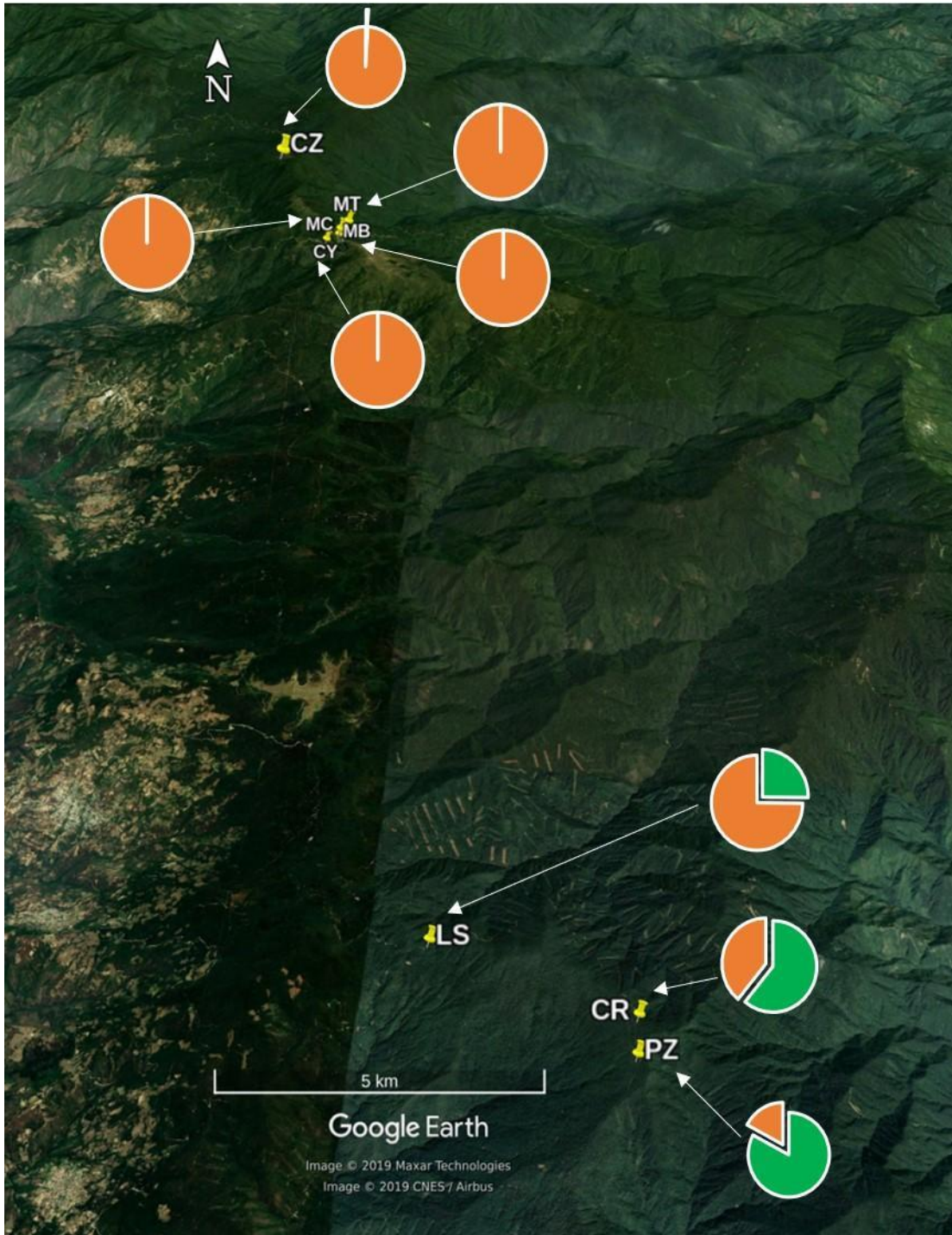


Figura 12.- Promedios de probabilidad de pertenencia por sitios a partir de los datos obtenidos con ADMIXTURE, sobre una imagen satelital del área de estudio.

En el Análisis de Variación Molecular (AMOVA), el mayor porcentaje de la varianza total fue debida a la diferencia entre los individuos. Entre los ocho sitios se observó un bajo porcentaje de varianza (2.041%). Los coeficientes de diferenciación y de endogamia fueron bajos, cercanos a cero ($F_{ST}=0.0204$, $F_{IS}= 0.0101$, $p<0.05$).

Los coeficientes F_{ST} entre pares de sitios fueron bajos (<0.046) (Fig. 13). Los valores entre los sitios de la zona Norte son bajos y similares entre ellos (0.0252-0.0295), por otro lado, los valores de F_{ST} entre los sitios de la zona Sur son contrastantes. Mientras que los sitios PZ y CR presentan el valor más bajo (0.0229), los valores de estos dos sitios con respecto a LS son mayores a los observados en la zona Norte (>0.035). El valor más alto de F_{ST} se observa entre LS y MB (0.0451).

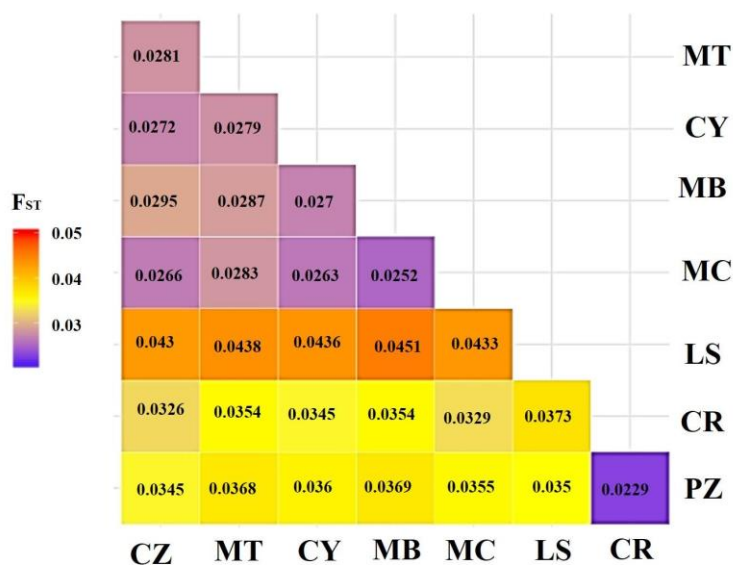


Figura 13.- Coeficientes de F_{ST} entre pares de sitios. El gradiente de color representa el rango de valores.

8.1.4 SNPs outliers

Los valores de F_{ST} para cada SNP entre las dos zonas, Norte (CZ, MT, MC, MB y CY) y Sur (LS, CR y PZ), se encontraron entre 0 y 0.2096 (media = 0.0112, DE = 0.0235). La mayoría con valores de F_{ST} fueron <0.05 , lo que indica una diferenciación muy baja.

Con la selección del 1% de los SNPs con los F_{ST} más altos ($F_{ST} \geq 0.1$), se identificaron 82 SNPs (Fig. 14). A partir de los cuales, se realizó una ACP para visualizar su agrupamiento (Fig. 15).

En el ACP, los dos primeros CPs explican el 40.17% de la varianza acumulada (Fig. 15a). El primer CP separa a la mayoría de los individuos entre las dos zonas, excepto LS_28, LS_29 y CR_18. Mientras que, con el segundo, se observan tres grupos en la zona Norte, los cuales son una mezcla de individuos entre los cinco sitios.

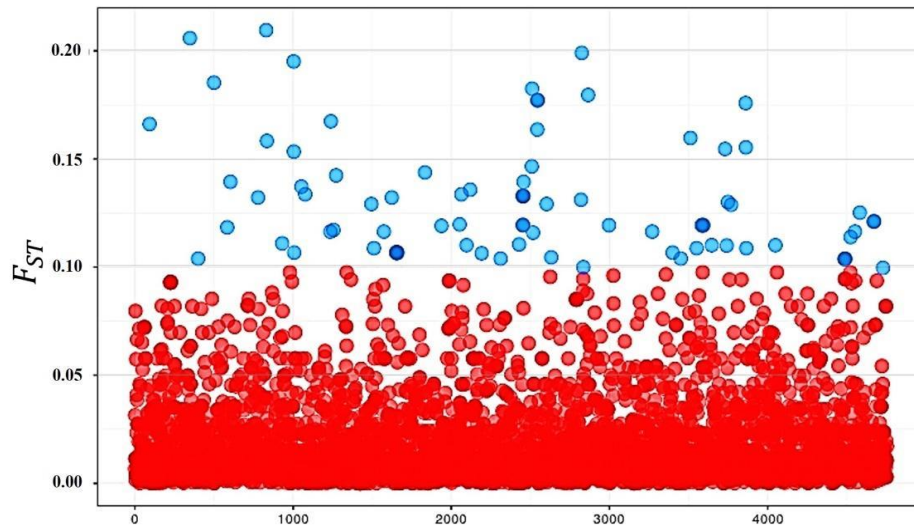


Figura 14.- Distribución de los SNPs según su valor F_{ST} entre los dos agrupamientos propuestos. En azul se presenta el 1% de los SNPs con los valores más altos de F_{ST} .

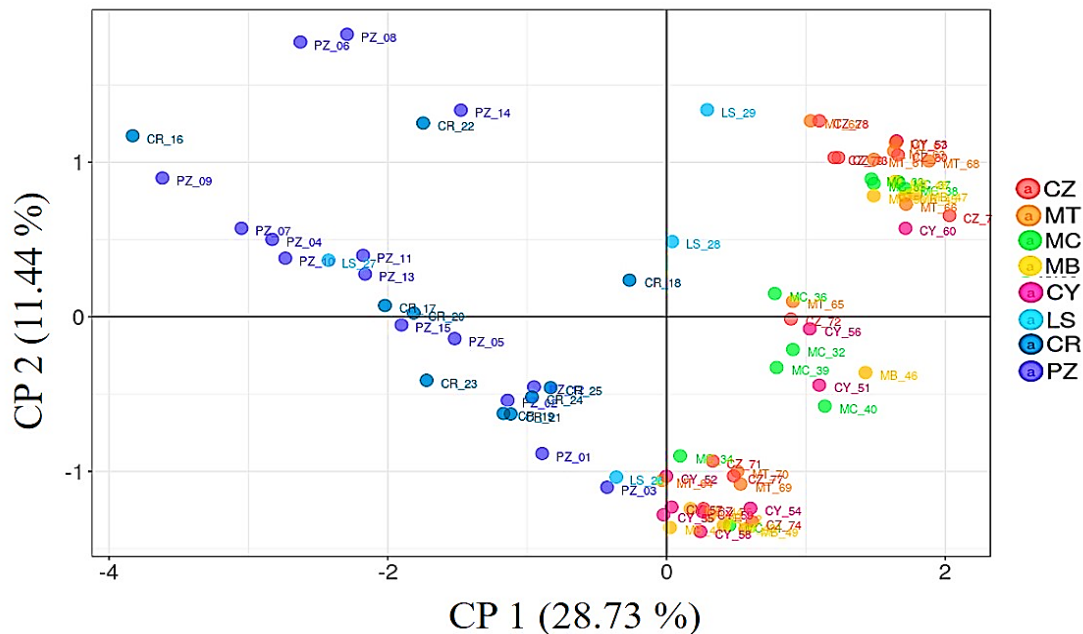


Figura 15.- Análisis de componentes principales (ACP), a partir del conjunto de los 82 SNPs *outliers*.

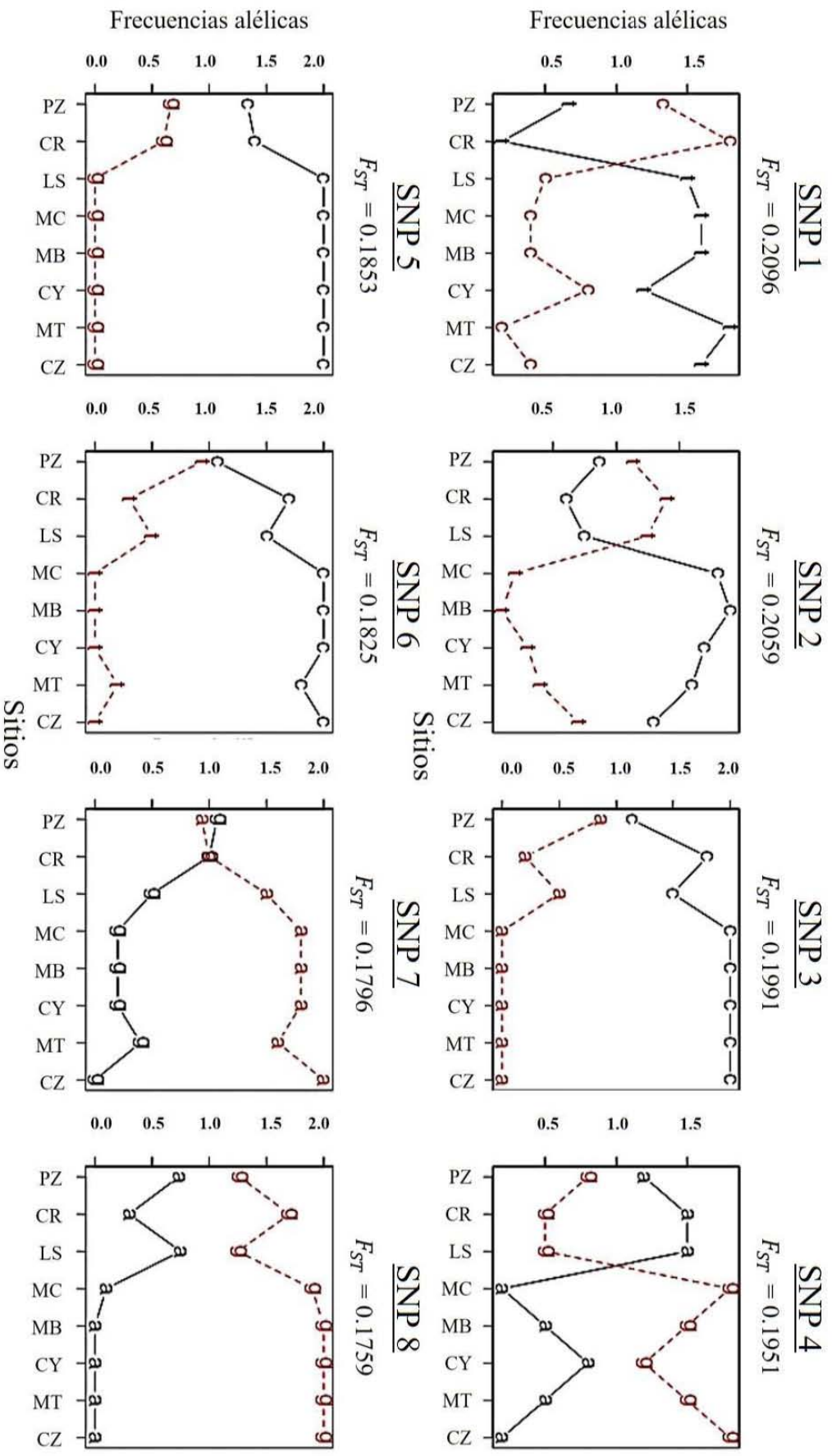


Figura 16.- Cambios de las frecuencias alélicas de los SNPs con los F_{ST} más altos, en los ocho sitios de muestreo.

En el SNP 3, se observa una fluctuación en las frecuencias alélicas entre los sitios de la zona Norte y Sur. Mientras que en la zona Norte solo se observa un alelo para el SNPs, la citosina, en la zona Sur se observa la presencia de otro alelo, adenina.

En el SNP 5, solo en los sitios CR y PZ, ambos de la zona Sur, se observan SNPs con cambios de dos alelos, guanina y citosina, mientras en que los demás sitios, este SNP solo se encuentra con el alelo citosina.

8.2 Distribución potencial

Variables ambientales

Para el conjunto de variables de CONABIO se seleccionaron 10 variables (Cuadro 4), mientras que para WorldClim se seleccionaron ocho (Cuadro 5). Esta selección, fue considerando el coeficiente de correlación entre las variables y sus contribuciones al modelo de distribución mediante una prueba de Jackknife.

Cuadro 4.- Capas obtenidas del Geoportal de CONABIO, usadas como variables para el modelo de distribución potencial.

Variable	Clave	Escala
Existencia de leña en México a nivel municipal	EXL	1:250000
Exposición	EXP	1:4000000
Precipitación media anual	PRE	1:4000000
Regímenes de humedad del suelo	RHS	1:4000000
Área de captación específica modificada (acumulación de agua por escurrimientos superficiales)	ACE	1:4000000
Pendiente	PDT	1:4000000
Edafología (tipo de suelo)	EDA	1:1000000
Modelo digital de elevaciones	ELE	1:4000000
Hidrogeología (capacidad de retención de agua, mediante la porosidad y permeabilidad del suelo y subsuelo)	HGE	1:4000000
Temperatura media anual	TEM	1:4000000

Para el conjunto de las variables de CONABIO (Fig. 17a), la variable de existencia de leña (EXL), la cual señala el servicio ecosistémico de provisión de leña, fue seleccionada como una primera exploración de la información que se tiene acerca del impacto antropogénico sobre el área de estudio. Por otra parte, se observa una alta correlación negativa entre la elevación (ELE), la edafología (EDA) y el área de captación específica

(ACE); y entre la pendiente (PDT) y la exposición (EXP). Por otro lado, se observa una alta correlación positiva entre la precipitación (PRE), los regímenes de humedad del suelo (RHS) y la existencia de leña a nivel municipio (EXL).

Cuadro 5.- Capas ambientales obtenidas de WorldClim, usadas como variables para el modelo de distribución potencial. En gris se señalan las ocho variables seleccionadas para la modelación.

Variable	Clave
Temperatura media anual	BIO 01
Rango diurno medio (promedio mensual de temperatura máxima - temperatura mínima)	BIO 02
Isotermalidad (BIO2 / BIO7)	BIO 03
Temperatura estacional (desviación estándar * 100)	BIO 04
Temperatura máxima del mes más cálido	BIO 05
Temperatura mínima del mes más frío	BIO 06
Rango anual de temperatura (BIO5-BIO6)	BIO 07
Temperatura media de cuatrimestre más húmedo	BIO 08
Temperatura media de cuatrimestre más seco	BIO 09
Temperatura media del cuatrimestre más cálido	BIO 10
Temperatura media del cuatrimestre más frío	BIO 11
Precipitación anual	BIO 12
Precipitación del mes más húmedo	BIO 13
Precipitación del mes más seco	BIO 14
Estacionalidad de la precipitación (coeficiente de variación)	BIO 15
Precipitación del cuatrimestre más húmedo	BIO 16
Precipitación del trimestre más seco	BIO 17
Precipitación del cuatrimestre más cálido	BIO 18
Precipitación del cuatrimestre más frío	BIO 19

Las variables de hidrogeología (HGE) y temperatura (TMP) presentan el mismo valor en todos los puntos de presencia de la especie y, por lo tanto, no se determinó su coeficiente de correlación, sin embargo, se mantuvieron en la modelación debido a que este comportamiento podría indicar alta especificidad de la especie hacia determinadas variables.

Mediante un primer modelo de distribución potencial con todas las variables, los resultados de la prueba de Jackknife, indicaron que todas las variables de CONABIO contribuyen altamente. Este modelo obtuvo un AUC de 0.999, lo que sugiere una buena capacidad discriminadora del modelo, por lo que, se seleccionaron todas las variables para el modelo de distribución potencial y su proyección al país.

Para el conjunto de las variables de WorldClim (Fig. 17b), las variables con los menores coeficientes de correlación fueron: BIO 02, BIO 04, BIO 07, BIO 13, BIO 16, BIO 18, BIO 19. Por otro lado, en el primer modelo de distribución con todas las variables, los resultados de la prueba de Jackknife indicaron que solo las variables: BIO 11, BIO 18 y BIO 19, mostraron altos valores contribución al modelo ($\geq 22\%$).

Por lo anterior, se seleccionaron las siguientes variables: la temperatura media anual (BIO 01); la isothermalidad (BIO 03); la temperatura media del cuatrimestre más húmedo (BIO 08); la temperatura media del cuatrimestre más seco (BIO 09); la temperatura promedio del cuatrimestre más frío (BIO 11); la precipitación del cuatrimestre más lluvioso (BIO 16); la precipitación del cuatrimestre más cálido (BIO 18); y la precipitación del cuatrimestre más frío (BIO 19).

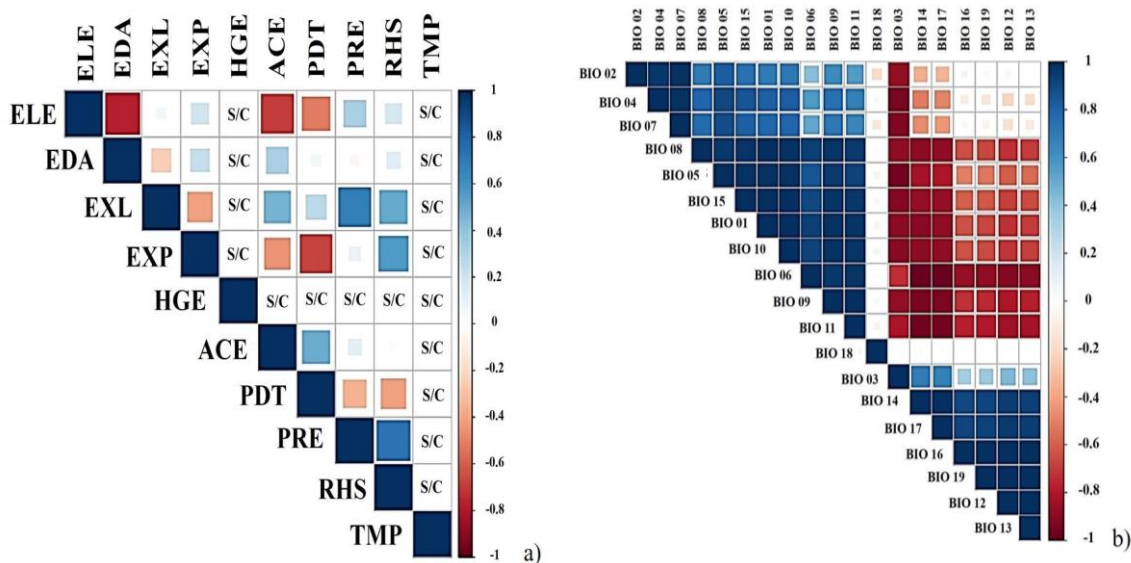


Figura 17.- Análisis de correlación entre los dos conjuntos de variables: a) entre el conjunto de variables de CONABIO (Cuadro 4); y b) entre las 19 variables de WorldClim (Cuadro 5). El tamaño del cuadrado indica que tan fuerte es la correlación entre las variables y el color indica si la correlación es positiva (azul) o negativa (rojo). S/C significa que la variable no mostró cambios en sus valores.

Modelo de distribución potencial actual

Los valores de AUC para evaluar el desempeño de los modelos, fueron de 0.999 para CONABIO y de 0.998 para WorldClim, ambos con una desviación estándar de ± 0.001 .

Los mapas de la distribución potencial a partir de los dos conjuntos de variables (Fig. 18), presentaron un patrón similar, indicando una distribución restringida al estado de Oaxaca y en especial a la zona montañosa que se desplaza a través de las provincias biogeográficas de la Sierra Madre del Sur y de la Provincia de Oaxaca.

En sus respectivas proyecciones al país, se observó que las probabilidades de ocurrencia de la especie fuera de su área de distribución conocida son nulas (CONABIO) o muy bajas (WorldClim).

Para el modelo con el conjunto de variables de CONABIO (Fig. 19a), la existencia de leña a nivel municipal (EXL) fue la variable que aportó la mayor ganancia al modelo cuando se consideró de forma aislada y la que disminuyó la ganancia al ser omitida. Lo que señala, que la información de esta variable por sí sola contribuye de manera significativa al modelo, además de contener una gran cantidad de información que no está presente en las otras variables. El propósito de esta variable fue representar el servicio ecosistémico de provisión de leña.

Las siguientes dos variables que más contribuyeron al modelo fueron: la elevación (ELE) y la edafología (EDA). Con la variable de edafología, solo se observaron dos valores para la distribución de la especie: el luvisol vértico en el sitio CZ (zona Norte) y el acrisol húmico en los otros sitios.

Por otro lado, la variable que menos contribuyó al modelo fue la exposición (EXP), debido a que los puntos de presencia de la especie no muestran alguna tendencia hacia determinado tipo de exposición: norte, sur, este u oeste.

Las variables de temperatura media anual (TMP) e hidrogeología (HGE), mantuvieron un solo valor en todos los puntos de registro de la especie. En la primera, fue la zona térmica templada y para la segunda fue un suelo con rocas metamórficas y con baja permeabilidad.

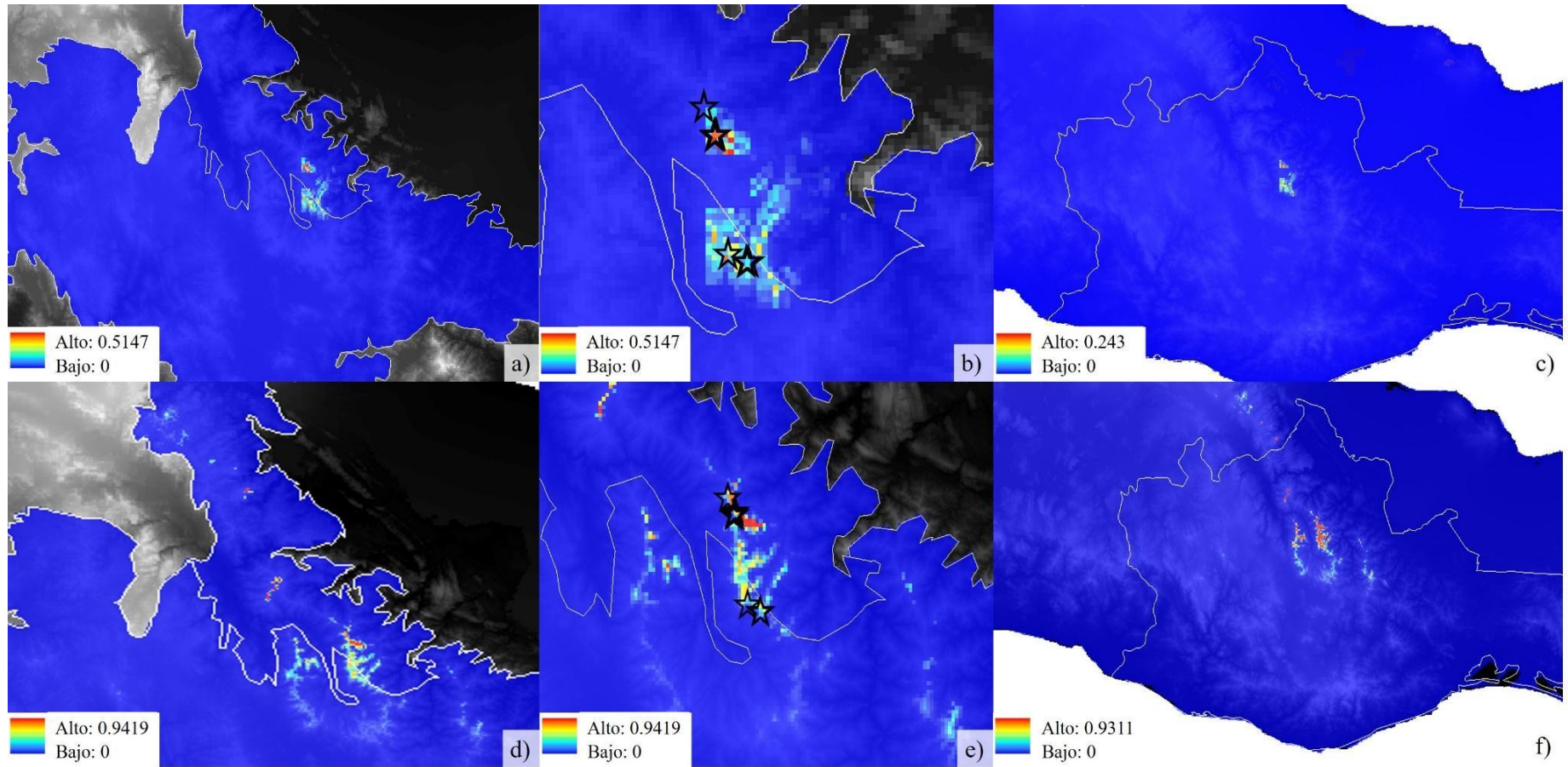


Figura 18.- Mapa de la distribución potencial de *Q. macdougallii*: a-b) Modelo con las variables de WorldClim, en el área de las provincias biogeográficas SMS y PO; y c) a nivel nacional, señalando el límite del estado de Oaxaca; d-e) Modelo con las variables de CONABIO en el área de las provincias SMS y PO; y f) a nivel nacional, señalando el límite del estado de Oaxaca. La probabilidad de ocurrencia va de alta (rojo) a nula (azul) Las estrellas representan los sitios de muestreo para este estudio. El mapa de elevaciones en el fondo presenta su gradiente de altitud de blanco (mayor altitud) a negro (menor altitud).

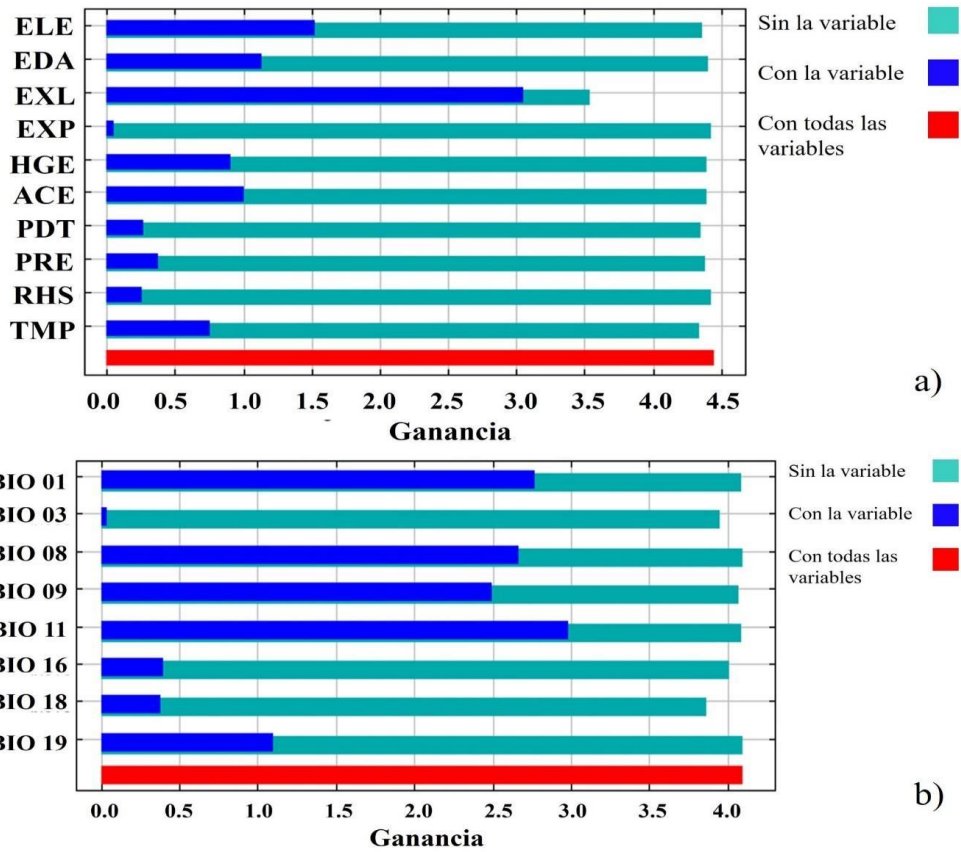


Figura 19.- La prueba de Jackknife mostró la contribución al modelo de cada variable (azul), sin la variable (turquesa) y con todas las variables (rojo): a) CONABIO; y b) WorldClim. Los valores mostrados son promedio de las 100 réplicas ejecutadas.

Para el modelo con el conjunto de variables de WorldClim (Fig. 19b), la temperatura promedio del cuatrimestre más frío (BIO 11), aportó la mayor ganancia al modelo cuando se consideró de forma aislada. Las otras tres variables que más contribuyeron al modelo fueron: la temperatura media anual (BIO 01); la temperatura media del cuatrimestre más húmedo (BIO 08); y la precipitación del cuatrimestre más frío (BIO 09), por otro lado, la variable que menos contribuyó al modelo fue la isotermalidad (BIO 03).

La precipitación del cuatrimestre más cálido (BIO 18) fue la variable que al ser omitida disminuyó la ganancia en el modelo, lo que indica que la información que aporta esta variable no está presente en las demás.

En ambos modelos, CONABIO y WorldClim, se observa un patrón similar de distribución restringida, así como la presencia de características específicas para la presencia de *Q. macdougallii*.

Modelos de distribución potencial futura

Las proyecciones al futuro de la distribución potencial de la especie, hacia el 2050 y 2070, se basaron en las ocho variables seleccionadas de WorldClim (Fig. 20). En ambos escenarios (RCP4.5 y RCP8.5), se observa que la zona donde se ha descrito actualmente la presencia de *Q. macdougallii*, la Sierra Norte de Oaxaca, seguirá siendo la que presente la mayor probabilidad de ocurrencia. Sin embargo, esta probabilidad gradualmente irá disminuyendo, desde la actual (0.94), las de 2050 (RCP4.5= 0.82; RCP8.5= 0.78), hasta las de 2070 (RCP4.5= 0.31; RCP8.5=0.22), de igual manera, sus áreas con los hábitats idóneos para su establecimiento (Cuadro 6).

Cuadro 6.- Áreas de distribución potencial en Km². Actual, con los modelos de CONABIO y WorldClim. Proyección futura, con los modelos de WorldClim bajo los escenarios de cambio climático RCP4.5 y RCP8.5, para los años 2050 y 2070. Se señala el valor de probabilidad de ocurrencia; el mínimo se estableció como 0.2 para cada modelo y el máximo fue el valor más alto reportado para cada modelo.

Probabilidad de ocurrencia					
Distribución	Modelo		Mín.	Máx.	Área (Km ²)
ACTUAL	WorldClim		0.2	0.94	230.87
2050	RCP4.5	(aumento de 0.9 a 2°C)	0.2	0.82	13.09
	RCP8.5	(aumento de 1.4 a 2.6°C)	0.2	0.78	96.23
2070	RCP4.5	(aumento de 1.1 a 2.6°C)	0.2	0.31	2.45
	RCP8.5	(aumento de 2.6 a 4.8°C)	0.2	0.22	0.82

Podemos observar un gran decremento en el área (Km²) tan solo desde la distribución potencial actual (230.87) hacia el año 2050 (RCP4.5=13.09 y RCP8.5=96.23). Este decremento se mantiene en las proyecciones al 2070, donde se observa una considerable reducción (RCP4.5=2.45 y RCP8.5=0.82).

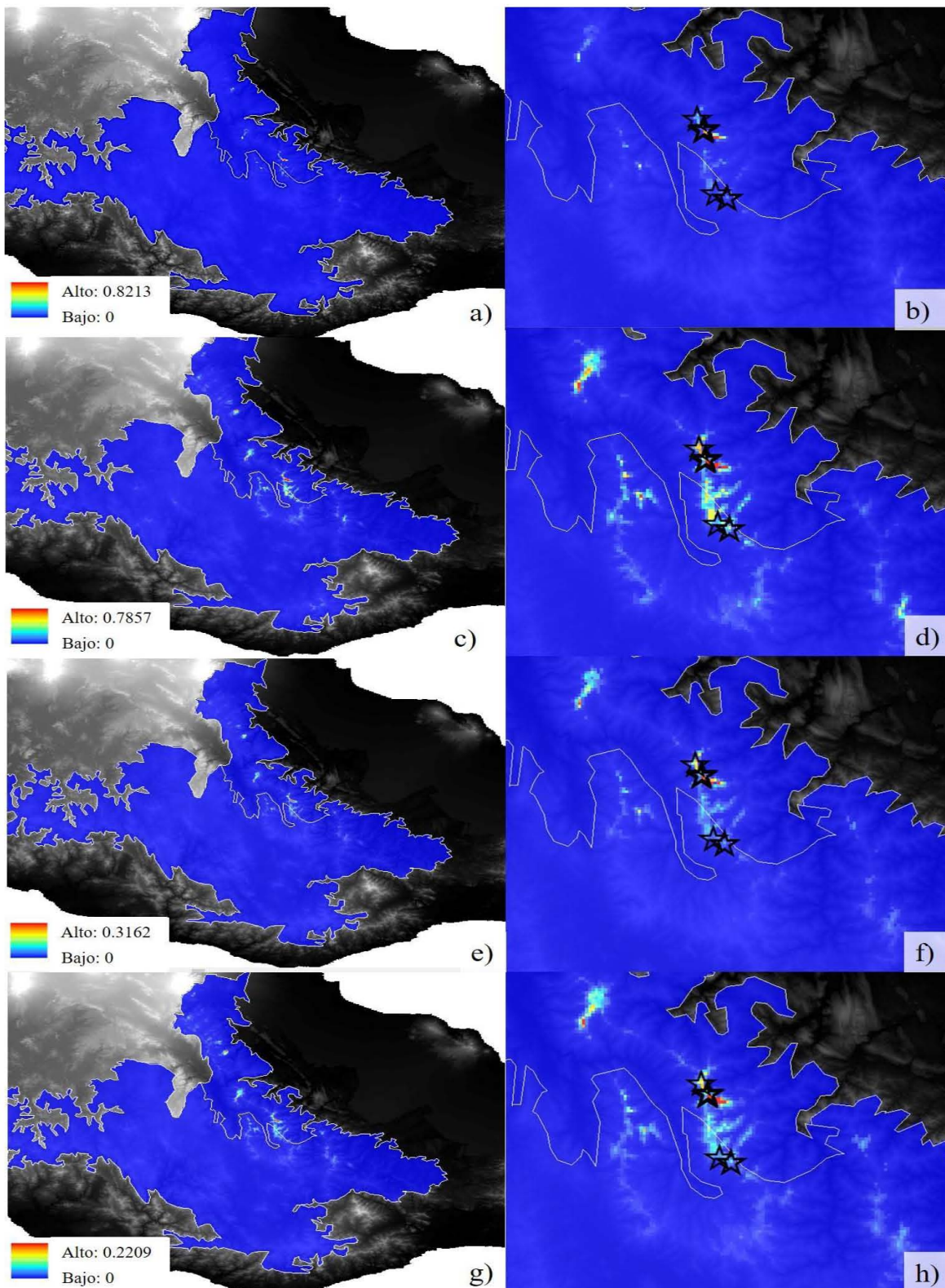


Figura 20.- Mapas de distribución potencial proyectadas al futuro: a-b) escenario RCP4.5 y c-d) escenario RCP8.5, en el 2050; e-f) escenario RCP4.5 y g-h) escenario RCP8.5 en el 2070. En forma de estrella se muestran los sitios de muestreo.

8.3 La relación entre la variación genómica, la geografía y el ambiente

En la prueba Mantel se observaron valores altos de correlación ($r= 0.7604$, $p\text{-valor}= 0.002$), entre la matriz de distancias geográficas y la matriz de distancias genéticas, la cual, fue generada a partir de los valores F_{ST} entre los sitios y convertidos a valores de flujo génico, $Nm=F_{ST}/(1-F_{ST})$ (Fig. 21).

Una alta correlación positiva indica que a mayor distancia geográfica mayor distancia genética entre los sitios de muestreo. Este patrón se observa entre la mayoría de los sitios, sin embargo, la relación entre PZ y LS se aleja de esta tendencia, lo que indica que, aunque se encuentran geográficamente cercanos sus distancias genéticas entre ellos son más diferentes de lo esperado.

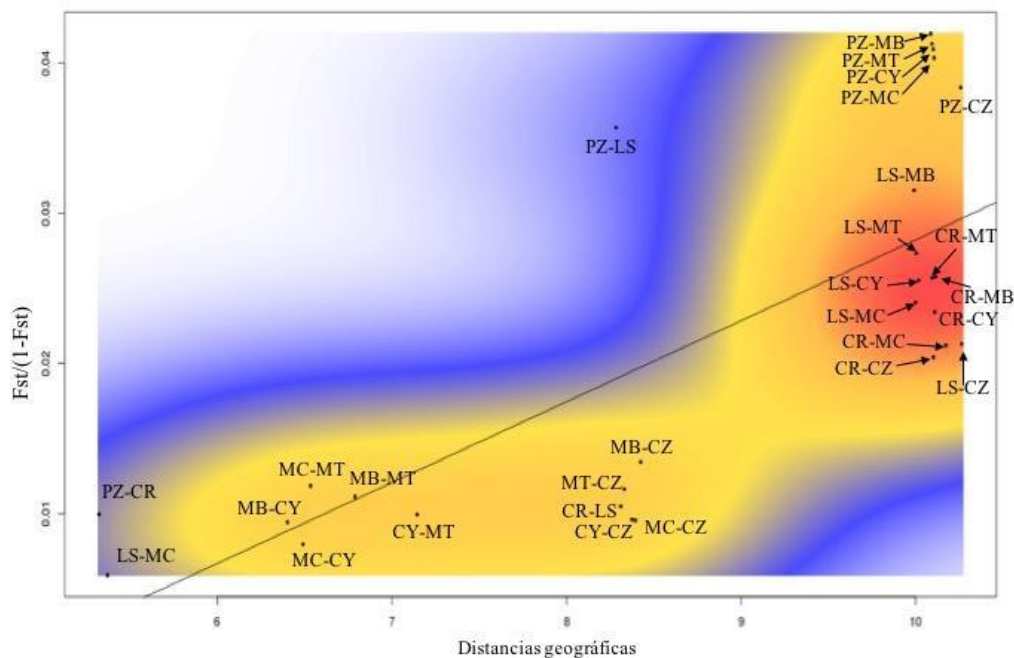


Figura 21.- Diagrama de la correlación de Mantel entre una matriz de distancias geográficas (Km) y una de distancias genéticas representadas por flujo génico (Nm), los colores representan donde se acumulan la mayoría de los valores, de azul a rojo.

Para la prueba parcial de Mantel, la tercera matriz de distancias se obtuvo a partir de las variables utilizadas en los modelos de distribución potencial; CONABIO y WorldClim, con 10 y 8 variables respectivamente.

Con la prueba parcial de Mantel (Cuadro 7), se vuelve a observar una alta correlación significativa entre el flujo génico y las distancias geográficas. Por otro lado, al analizar las matrices de variables ambientales con el flujo génico entre los sitios, se observaron valores muy bajos de correlación (~ 0.03 , $p\text{-valor} > 0.1$). Se observa un patrón similar entre las matrices de distancias geográficas y las ambientales.

También, se realizó una prueba parcial de Mantel a partir de las variables de WorldClim que se utilizaron para el modelo de distribución futura (escenarios RCP4.5 y RCP8.5 para el 2050 y 2070) (Anexo VI), donde se observaron bajos valores de correlación ($p\text{-valor} > 0.08$).

Cuadro 7.- Coeficientes de correlación entre las matrices, obtenidos por medio de la prueba parcial de Mantel. *Gen*=matriz de flujo génico, *Geo*=matriz de distancias geográficas, *Amb*=matriz de distancias euclidianas entre las variables ambientales: CONABIO y WorldClim, con 10 y 8 variables respectivamente.

	CONABIO		WorldClim	
	Parcial de Mantel (<i>r</i>)	<i>p</i> -valor	Parcial de Mantel (<i>r</i>)	<i>p</i> -valor
<i>Gen~Geo</i>	0.7604	0.002	0.7604	0.002
<i>Gen~Amb</i>	0.033	0.367	0.0382	0.383
<i>Geo~Amb</i>	0.1353	0.203	0.0565	0.335

9. DISCUSIÓN

La especie *Q. macdougallii*, es el único encino endémico de la sección *Quercus* reportado para el estado de Oaxaca (Valencia, 2004; Romero-Rangel *et al.*, 2015). Los pocos trabajos acerca de este encino, en su mayoría, se han enfocado en el aspecto ecológico y en la zona Norte de su distribución, en donde se ha observado un bajo número poblacional (Anacleto-Carmona, 2015), además de una baja tasa de germinación y de supervivencia, en especial en ambientes con disturbio (Pariona *et al.*, 2017; Clark-Tapia *et al.*, 2018), lo que aumenta la vulnerabilidad de la especie ante cambios drásticos en su ambiente.

Este estudio, además de reportar otros sitios de presencia de la especie (en la zona Sur), presenta por primera vez un análisis de variación genética a nivel genoma; obtenido mediante la genotipificación basada en secuenciación (SBG), y un análisis de distribución potencial actual y futura para *Q. macdougallii*. Se espera que este estudio, contribuya al conocimiento de la especie, y sea un material de apoyo en la toma de decisiones para la conservación de la especie.

La variación genómica ¿cuándo es alta o baja?

Los resultados del análisis de variación genómica, muestran valores cercanos a 0.2. La diversidad nucleotídica (π) y poblacional (θ) fueron similares (~ 0.2), y lo mismo se observó en los valores de la heterocigosidad observada (H_o) para cada sitio (~ 0.21), los cuales, en general, fueron mayores a los de la heterocigosidad esperada (H_e). Estos valores son más bajos que los reportados para la especie en un estudio previo, en el cual se analizaron tres microsatélites de la serie *quru-GA*, se identificaron 26 alelos y se observaron niveles bajos de variación genética, $H_o = 0.4026$ y $H_e = 0.5572$ (Molina-Garay, 2011).

Cabe resaltar que, los marcadores analizados en el estudio previo de *Q. macdougallii*, de la serie *quru-GA*, son nucleares y neutrales. Los cuales, han sido ampliamente utilizados en otros estudios de genética de poblaciones en encinos, reportando una variación baja, como en *Q. stellata*, encino endémico del sureste de Estados Unidos ($H_o = 0.417$; Chatwin *et al.*, 2014), una variación moderada, como en *Q. mullerii*, encino endémico de Oaxaca ($H_o = 0.55$; Pingarroni *et al.*, 2019, en prensa), y una variación alta, como en *Q. sideroxylla*, encino endémico del centro de México ($H_o = 0.73$; Alfonso-Corrado *et al.*, 2014).

Por otro lado, en el reciente estudio acerca de la variación genómica de *Q. rugosa*, encino de amplia distribución en México, donde se identificaron 5,136 SNPs, se observó una $H_e=0.364$, y cual sugiere una alta diversidad genética (Martins *et al.*, 2018). En otros estudios de variación genómica en especies arbóreas, se han reportado niveles de variación bajos, como en el encino *Q. championii*, donde mediante el análisis de 5,207 SNPs, se observó a nivel poblacional una H_o entre 0.044 y 0.078, y una H_e entre 0.062 y 0.092 (Jiang *et al.*, 2019). Mientras que, en otras especies, se han observado mayores niveles de variación, por ejemplo, para *A. koa*, una acacia endémica de las islas de Hawaii, mediante el análisis de 11,000 SNPs se ha reportado una diversidad genética alta, con una H entre 0.29 y 0.33 (Gugger *et al.*, 2018).

Las diferencias entre el tipo y cantidad de marcadores analizados, resaltan cuando se intenta establecer a partir de qué valor de heterocigosidad, podemos considerar una diversidad genética baja o alta. En el presente estudio, a partir del análisis de 8,186 SNPs, se observaron valores de heterocigosidad poblacional cercanos a 0.2. Estos marcadores, además de presentar tasas de mutación relativamente bajas (10^{-8} a 10^{-9} ; Brumfield *et al.*, 2003), al ser bialélicos, establecen un máximo valor teórico de heterocigosidad de 0.5 (Singh *et al.*, 2013).

Por otro lado, estos SNPs presentan una alta confiabilidad, ya que, al ser obtenidos a partir del enfoque de reducción del genoma, permiten analizar una mayor cobertura del genoma y, por consiguiente, obtener una mejor estimación de la diversidad genética total de la especie (McKinney *et al.*, 2017). Debido a lo anterior, los resultados sugieren que *Q. macdougalii* presenta una diversidad genética moderada.

¿Expansión o reducción de la especie?

Los valores observados en la prueba D de Tajima se alejaron del modelo de neutralidad con valores negativos, aunque muy cercanos a cero, entre -0.42 en el sitio PZ y -2.88 en LS.

Estos valores negativos en la prueba D de Tajima, podría indicar una selección purificadora o direccional, especialmente en el sitio LS. Valores similares han sido reportado en otros encinos, como en *Q. lobata*, encino endémico de California, en donde al analizar SNPs de 40 genes candidatos, se obtuvo un valor bajo y negativo (-0.44), el cual, se relaciona con una expansión poblacional y/o alguna forma de selección ligada, sugiriendo que esos SNPs, podrían estar en genes bajo una fuerte selección purificadora.

Asimismo, valores negativos en esta prueba, pueden indicar que las poblaciones recientemente han pasado por un cuello de botella o por un barrido selectivo (Brookfield, 2001). El cambio en el tamaño poblacional de la especie, puede reflejarse en los alelos únicos, ya que un evento de cuello de botella tiende a eliminarlos, lo cual, podría tener importantes implicaciones desde una perspectiva evolutiva local, ya que esta diversidad alélica puede influir en la capacidad de adaptación de la especie y, por consiguiente, en su viabilidad poblacional a largo plazo (Weiser *et al.*, 2013).

Estudios genéticos de encinos endémicos han encontrado bajos números de alelos (Molina-Garay, 2011; Pingarroni *et al.*, 2019, en prensa), en comparación con encinos de amplia distribución (Romero-Rangel *et al.*, 2015). Estos bajos valores, pueden estar fuertemente influenciados por el tamaño poblacional, por ejemplo, en *Q. sideroxyla*, se observó que el número de alelos únicos es sensible y tiende a aumentar conforme al tamaño poblacional (Alfonso-Corrado *et al.*, 2014).

En el sitio LS, de la zona Sur, se observó la D de Tajima más baja, así como el menor número de alelos únicos (78), asimismo, presenta el número poblacional más bajo, ya que en el sitio de muestreo solo se localizaron cinco individuos. Esto podría indicar que ha pasado recientemente por un cuello de botella o que su recuperación y/o expansión después del evento ha sido más lenta, en comparación con los demás sitios. Por otra parte, los dos sitios con mayor número de alelos únicos, PZ (370), en la zona Sur, y CZ (178), en la zona Norte, se encuentran en extremos opuestos de la distribución conocida de la especie.

Una alta diversidad alélica, en especial, de alelos únicos, puede conferir a la especie cierta ventaja al enfrentar los cambios en su ambiente. En las especies arbóreas forestales, las poblaciones periféricas tienden a ser afectadas significativamente por el cambio climático, por lo cual, su capacidad de adaptación dependerá de la interacción entre los procesos de flujo génico, las presiones selectivas, la deriva génica y su tasa de crecimiento poblacional (Fady *et al.*, 2016; Müller & Gailing, 2019).

Los altos valores de alelos únicos observados en estos dos sitios, podría indicar que mantienen un tamaño efectivo poblacional mayor que los otros sitios. Las observaciones en campo indican que ambos sitios presentan bajos niveles de perturbación, y que en el sitio CZ, posiblemente se encuentre el mayor número de individuos de la especie.

Por otra parte, esa gran cantidad de alelos únicos en el sitio PZ, puede indicar bajos niveles de flujo génico, lo cual, podría favorecer el aislamiento y por consiguiente aumentar su diferenciación con respecto a los otros sitios (Slatkin, 1985; Chakraborty, 1993). Esto, podría ser posible, considerando que PZ se encuentran en la zona más alta, de la distribución conocida de la especie, lo que complicaría la llegada del flujo génico de los sitios vecinos que se encuentran a menor altitud.

Los resultados sugieren que la especie podría encontrarse en proceso de expansión, en especial, en la zona Sur. Por otra parte, es posible que el sitio LS haya pasado recientemente por un evento de cuello de botella, confiriendo una mayor vulnerabilidad a sus individuos frente a los efectos de la deriva génica.

Una sola población, pero ¿con dos grupos genéticos?

En este estudio, se observaron valores bajos de diferenciación. Se obtuvo una $G_{ST}= 0.063$, un valor mayor al observado en poblaciones de *Q. lobata*, encino endémico de California (Grivet *et al.*, 2008), y en poblaciones de *Q. rugosa*, encino de amplia distribución en México (Martins *et al.*, 2018), ambos con una $G_{ST}= 0.05$.

En estudios genéticos de especies maderables de vida larga, se ha reportado una alta variación dentro de sus poblaciones, sin embargo, la diversidad entre ellas, de acuerdo con la estimación de G_{ST} es significativamente menor (~ 0.08), que la reportada para especies herbáceas de ciclo de vida anual ($G_{ST} > 0.25$) (Porth & El-Kassaby, 2014).

Los análisis de distancias genéticas, entre los sitios y entre los individuos, mostraron en general, valores bajos y similares, con excepción del sitio LS, el cual, presentó el valor más alto de sustituciones nucleotídicas (7.13). Asimismo, los análisis ACP y ADCP, mostraron una separación entre las dos zonas, en donde las menores distancias genéticas entre los sitios CR (Sur) y MC (Norte), sugieren una conexión entre las dos zonas de estudio. De igual manera, se observó un agrupamiento estrecho entre los individuos del Norte y un agrupamiento más disperso entre los del Sur.

Este patrón fue similar al que se obtuvo mediante los dos enfoques evolutivos de agrupamiento, fastStructure (Raj *et al.*, 2014), y ADMIXTURE (Alexander *et al.*, 2009). ADMIXTURE y fastStructure con el modelo simple, indicaron un solo agrupamiento,

señalando que la especie se comporta como una sola población, sin embargo, fastStructure con el modelo logístico, indicó dos posibles grupos genéticos. Raj *et al.* (2014), mencionan que, cuando la estructura es más difícil de resolver, el modelo logístico genera estimaciones más precisas de ascendencia, por lo cual, se sugiere la presencia de dos grupos genéticos para la especie.

El coeficiente promedio de F_{ST} fue 0.0204, este valor fue menor al reportado para especies de amplia distribución, como en *Q. suber* ($F_{ST}= 0.0541$; Pina-Martins *et al.*, 2019), en *Q. rugosa* ($F_{ST}= 0.056$; Martins *et al.*, 2018) y en otras especies arbóreas estructuradas, como en el eucalipto ($F_{ST}=0.095$; Cappa *et al.*, 2013). Sin embargo, fue mayor a lo observado en encinos europeos no estructurados, como *Q. robur* ($F_{ST}= 0.012$) y *Q. petraea* ($F_{ST}= 0.013$) (Guichoux *et al.*, 2013).

Por otra parte, entre los sitios de la zona Norte se observaron valores similares (~ 0.02), y menores que los observados entre los sitios de la zona Sur (~ 0.03), en donde, LS fue el sitio que mostró los mayores valores de diferenciación con respecto a todos los demás sitios, entre 0.035 y 0.0451. Esto indica, la presencia de altos niveles de flujo génico entre los sitios del Norte, mientras que, en la zona Sur, se observa una estructura incipiente, en donde, posiblemente el aislamiento, el bajo número poblacional y los efectos de un cuello de botella, están generando presiones de diferenciación sobre la especie.

Los bajos valores de diferenciación poblacional observada y la moderada diversidad genética, pueden indicar altas tasas de flujo génico histórico (Loveless & Hamrick, 1984), similar a lo observado en las poblaciones de *Q. lobata* (Grivet *et al.*, 2008), *Q. sideroxylla* (Alfonso-Corrado *et al.*, 2014) y *Q. rugosa* (Martins *et al.*, 2018). Esto es característico de poblaciones remanentes de una especie que en el pasado tuvo una distribución más amplia y continua (Alfonso-Corrado *et al.*, 2004; Gorgonio-Ramírez, 2012; Alfonso-Corrado *et al.*, 2014).

Por otro lado, en este estudio, se observó un coeficiente de endogamia cercano a cero, $F_{IS}=0.0101$, siendo ligeramente mayor al observado en encinos europeos de mayor distribución, como *Q. robur* ($F_{IS}= -0.004$) y *Q. petraea* ($F_{IS}= 0.001$) (Guichoux *et al.*, 2013). El bajo coeficiente de endogamia, así como una distribución que tiende a ser agregada a cortas y medianas distancias en la zona Norte (Anacleto-Carmona, 2015), sugieren que *Q.*

macdougallii, como en otros encinos, podría presentar un entrecruzamiento autoincompatible (Ellstrand, 2014), previniendo así, altos niveles de endogamia a pesar de que sus individuos se encuentren cerca (Alfonso-Corrado *et al.*, 2014).

En el análisis de red de expansión mínima, se observó la conexión entre los individuos a partir de sus distancias genéticas, asimismo, en este análisis no se identificaron individuos clonales para la especie. Por otra parte, el individuo CR_16 mostró distancias genéticas similares con individuos del sitio PZ y CR, representando una conexión entre ambos sitios, lo cual, indica la presencia de flujo génico entre ambos.

También se observó que siete de los 29 individuos de la zona Sur (PZ_01, PZ_02, PZ_03, CR_19, LS_26, LS_28 y LS_29), se separaron de los demás individuos de la zona Sur y mostraron una mayor similitud con los de la zona Norte. Esto podría apoyar que, la especie se comporta como una sola población, con el genotipo que se observa en la zona Norte, mientras que en los otros 22 individuos de la zona Sur se ha identificado la presencia de un grupo genético con estructuración incipiente.

Los SNPs *outliers*, ¿revelan señales de divergencia?

Para este estudio, los SNPs *outliers* se identificaron, al analizar las diferencias a nivel de variación y de diferenciación genética mediante el valor F_{ST} , para cada SNP (Vasemägi & Primmer, 2005). Estos SNPs *outliers*, se encuentran fuera del patrón de neutralidad esperado, por lo cual, pueden ser candidatos para identificar señales de selección divergente (Narum & Hess, 2011). Este enfoque, ha sido utilizado en diversos estudios de genómica de poblaciones en encinos, con el fin de identificar patrones de adaptación local y gradientes de diferenciación, por ejemplo, entre el Norte y Sur en las poblaciones de *Q. rugosa* (Martins *et al.*, 2018), y entre el Este y Oeste, en las de *Q. suber* (Pina-Martins *et al.*, 2019).

En este estudio, se identificaron 82 SNPs *outliers*, entre las dos zonas de estudio, Norte y Sur, los cuales, mostraron valores bajos de diferenciación, entre 0.1 y 0.2. Estos SNPs, fueron el 1% del total de los marcadores identificados (8,186 SNPs), un porcentaje cercano a lo reportado en *Q. suber* (0.9%; Pina-Martins *et al.*, 2019) y *Q. rugosa* (1.4%; Martins *et al.*, 2018).

En el ACP se observó la separación de los individuos de la zona Sur, con excepción de tres individuos, LS_28, LS_29 y CR_18, los cuales, mostraron menores distancias genéticas con los individuos del Norte, asimismo, se observó la presencia de tres grupos en la zona Norte. Lo anterior, podría indicar un cierto grado de divergencia entre ambas zonas, lo cual, podría estar asociado con ciertas diferencias ambientales, como el tipo de suelo y la precipitación, además de ser una zona de transición entre dos provincias biogeográficas, al norte la provincia de Oaxaca y al sur, la Sierra Madre del Sur.

Por otro lado, los tres grupos observados en la zona Norte, indican que aunque los individuos se encuentran geográficamente cercanos y, con bajas diferencias ambientales entre ellos, pueden existir otros factores que están promoviendo una selección divergente, por ejemplo, las relaciones bióticas, la pendiente, la elevación, o el suelo (Sork *et al.*, 1993; Cuervo-Alarcon *et al.*, 2018).

La separación entre las dos zonas, así como los valores bajos de diferenciación (F_{ST}) y endogamia (F_{IS}), podría sugerir un equilibrio entre la adaptación local y el flujo génico, ya que mientras la primera promueve diferencias en los individuos entre distintas áreas, la segunda tiende a mantener a la especie como una sola población. Esto se ha reportado también en *Q. suber*, en donde se observó un bajo gradiente de diferenciación entre la zona Este y Oeste (Pina-Martins *et al.*, 2019). De igual forma, la presencia de estos gradientes puede ayudar a la especie a responder mejor frente a fuertes presiones selectivas, como las ocasionadas por el cambio climático (Kremer *et al.*, 2012; De Kort *et al.*, 2014).

El análisis de las frecuencias alélicas de los ocho SNPs *outliers*, con los valores más altos de diferenciación, mostraron una proporción 3:1 en las mutaciones puntuales de tipo transición y con respecto a las de tipo transversión. Diversos autores han reportado que las primeras tienden a presentarse en mayor proporción que las segundas, y que los cambios entre purinas y pirimidinas (transversiones), pueden afectar la expresión génica (Wakeley, 1996; Lachagari *et al.*, 2019).

En los ocho SNPs *outliers* se observó una tendencia de diferenciación entre las dos zonas, siendo muy clara en las dos transversiones. Debido a ello y considerando que en la zona Sur es posible que se encuentra un grupo genético incipiente para la especie, podríamos señalar que, en los sitios de la zona Sur, en especial en PZ y CR, es donde se han presentado

los nuevos alelos para estos SNPs *outliers*, los cuales, sólo presentan un alelo en los sitios de la zona Norte. Por consiguiente, estos SNPs *outliers* podrían ser interesantes candidatos para estudios de genómica funcional (Homolka *et al.*, 2013; Petit *et al.*, 2013; Müller & Gailing, 2019), e incluso para estudios comparativos con otras especies de encinos endémicos, como *Q. lobata* (Sork *et al.*, 2016b) y de amplia distribución, como *Q. rugosa* (Martins *et al.*, 2018).

Debido a lo anterior, el análisis de los SNPs *outliers* no mostró evidencia sustentable de genotipos adaptados a nivel local, sin embargo, a nivel regional los resultados sugieren la presencia de alelos asociados a cada una de las zonas de estudio, Norte y Sur. Por otra parte, se sugiere realizar análisis de genómica funcional para identificar SNPs de genes candidatos adaptativos (Homolka *et al.*, 2013; Sork *et al.*, 2016b).

La distribución presente y futura de la especie: ¿reducción o expansión?

El análisis de distribución potencial de *Q. macdougalii*, indicó una distribución restringida en ambos modelos; CONABIO y WorldClim, aun utilizando diferentes escalas en sus variables, nacionales e internacionales, lo cual, respalda esta distribución potencial restringida a la Sierra Norte de Oaxaca. Esta distribución potencial, coincide con la descrita en trabajos previos (Molina-Garay, 2011; Anacleto-Carmona, 2015), lo cual, resalta el estado de riesgo en que se puede encontrar la especie debido a su limitada área potencial de ocurrencia.

Por otra parte, Anacleto-Carmona (2015), menciona que en la zona Norte de la distribución de *Q. macdougalii*, hay una dominancia de elementos arbóreos del género *Pinus* y *Quercus*, así como un aprovechamiento forestal comunitario para la venta y el autoconsumo de recursos maderables obtenidos de los pinos y encinos. Debido a ello, para el modelo con las variables de CONABIO se consideró la existencia y consumo de leña a nivel municipal, para representar el servicio ecosistémico de provisión de leña (Ghilardi, 2006). Esta variable fue altamente significativa para el modelo de distribución, lo cual, sugiere que la especie se encuentra en una zona donde existe un consumo de leña de encino.

Actualmente, *Q. macdougalii* no se debe utilizar como fuente de leña o carbón, debido principalmente su inclusión en la UICN (2010). Por otra parte, la zona de la Sierra Norte ha sido descrita como una de las más conservadas en el país, con un manejo sustentable de los recursos forestales del género *Pinus* (Gómez-Mendoza *et al.*, 2006; Anta-Fonseca *et*

al., 2010; Álvarez & Rubio, 2013). Sin embargo, es necesario ahondar en los impactos de las actividades humanas sobre los recursos forestales de los encinos, para prevenir escenarios de sobreexplotación y llegar a un manejo sustentable de sus recursos (Ramírez-Marcial *et al.*, 2001; Abrams, 2003; Gorgonio-Ramírez *et al.*, 2017).

Otra variable que contribuyó significativamente al modelo de distribución fue la elevación, la cual, se ha reportado como un factor importante en el establecimiento de los encinos (Firmat *et al.*, 2017; Leroy *et al.*, 2019). Por ejemplo, Valencia (2004) menciona que, *Q. macdougallii* presenta una distribución fuertemente restringida a zonas altas, ya que se encuentra a partir de los 2,600 hasta los 3,000 msnm. Por otra parte, en este estudio, se reportaron individuos cerca de los 3,100 msnm en la zona Sur de su distribución, lo cual, puede estar relacionado con las diferencias genéticas encontradas entre las zonas.

La variable de edafología, además de contribuir significativamente al modelo, mostró ser una variable altamente específica para la presencia de *Q. macdougallii*, ya que la especie solo se distribuye en dos tipos de suelo, el luvisol vértico, presente solo en el sitio CZ, en el extremo de la zona Norte, y el acrisol húmico para los demás sitios. Es importante señalar que, de acuerdo con INEGI (1998), los suelos del tipo acrisol solo están presente en el 0.5% del territorio mexicano, mientras que los del tipo luvisol se encuentran en el 2.4%. Esto resalta la especificidad y singularidad edafológica del área en que se encuentra la especie.

En el modelo a partir del conjunto de variables de WorldClim, la temperatura promedio del cuatrimestre más frío (BIO 11) contribuyó significativamente al modelo de la distribución potencial de *Q. macdougallii*. Esto coincide con lo reportado por Antúnez *et al.* (2017), quienes, mediante el uso de modelos aditivos generalizados simples, mencionan que la temperatura promedio del mes más frío es una variable que puede influir considerablemente sobre la abundancia de la especie.

Otra de las variables que mostró una gran influencia en el modelo de distribución potencial, evidenciando parte de los requerimientos ecológicos de la especie, fue la precipitación del cuatrimestre más cálido (BIO 18), lo cual, sugiere que eventos prolongados de sequía y alteraciones en la distribución temporal de las precipitaciones durante el verano, podrían afectar considerablemente a la especie, como se observó en *Q. oleoides* (Ramírez-Valiente *et al.*, 2018).

En especies arbóreas de vida larga se sigue investigando su adaptación frente a situaciones de estrés, como cambios drásticos de temperatura y precipitación (Lee *et al.*, 2005). En ciertos encinos se ha reportado una gran resistencia a las sequías, debido a la profundidad de sus raíces y su arquitectura hidráulica (Cavender-Bares *et al.*, 2007; Limousin *et al.*, 2010; Skelton *et al.*, 2018). Por otro lado, sus interacciones bióticas también pueden influir considerablemente en su permanencia, por ejemplo, Moricca & Ragazzi (2008), estudiaron la interacción entre el encino *Q. ilex* y *Discula quercina*, un hongo endófito, y encontraron que, dependiendo de los cambios drásticos en la temperatura y la disponibilidad de agua, su relación mutualismo-neutralismo, puede cambiar el rol del hongo hacia un patógeno débil o un invasor oportunista. Debido a ello, se sugiere estudiar las interacciones de *Q. macdougalii* con otros organismos, para obtener un panorama más completo y un mejor entendimiento de la situación en que se encuentra.

Por otro parte, frente a los efectos del cambio climático, es necesario conocer las posibles características futuras del ambiente de que se encuentra actualmente la especie. En este estudio, las proyecciones al futuro de la distribución potencial de la especie muestran que la zona donde se distribuye *Q. macdougalii*, en la Sierra Norte de Oaxaca, seguirá presentando la única probabilidad de ocurrencia de la especie, sin embargo, irá disminuyendo conforme pase el tiempo y la temperatura aumente, especialmente en la zona Sur. Los resultados sugieren una reducción drástica del área con las condiciones ambientales idóneas para la especie; entre el área potencial actual, con 230 Km², hasta el área potencial para el 2070, con una estimación de 0.82 a 2.45 Km².

Debido a que los bosques pueden desaparecer en ciertas áreas a un ritmo más rápido del que pueden migrar o volver a crecer en nuevas áreas (Parmesan, 2006), es necesario conocer cómo reaccionan especies arbóreas que son clave para mantener el funcionamiento de los ecosistemas y sus servicios ambientales. Por ejemplo, los bosques de pino y encino en los Alpes suizos (Rebetez & Dobbertin, 2004) y en el norte de Arizona (Mueller *et al.*, 2005), ya se han visto afectados por aumentos en la temperatura, donde la mortalidad y los patógenos han aumentado y la productividad forestal ha disminuido.

En este estudio, se identificó una distribución restringida de la especie, así como variables que confieren altos grados de especificidad a la especie, como la edafología, la temperatura promedio del cuatrimestre más frío y la precipitación del cuatrimestre más

cálido. Por otra parte, los modelos de distribución futura bajo dos escenarios de cambio climático, sugieren una considerable reducción en el área de las condiciones idóneas para la especie, siendo la zona Sur la más afectada.

La relación entre la variación genómica, la geografía y el ambiente.

Se identificó un patrón significativo de aislamiento por distancia para *Q. macdougallii* ($r=0.7604$, $p\text{-valor}=0.002$), mayor al encontrado en *Q. rugosa* ($r=0.475$, $p\text{-valor}=0.015$; Martins *et al.*, 2018), encino de amplia distribución en México. Este patrón de aislamiento por distancia, concuerda con el gradiente de diferenciación genética observado entre la zona Norte y Sur. Lo cual, señala que, a pesar de los niveles bajos de diferenciación, estos están pueden estar altamente relacionados con las distancias geográficas.

Por otra parte, la relación entre el sitio LS y PZ, ambos de la zona Sur, no mantiene este patrón de aislamiento, lo que indica que, aunque se encuentran geográficamente cercanos, sus distancias genéticas son mayores a lo esperado. Considerando los análisis de variación genómica del sitio LS, además de las observaciones en campo, es posible que este comportamiento esté relacionado con un evento de cuello de botella, y con el efecto de barreras orográficas sobre el intercambio de flujo génico, los cuales, han podido ejercer una presión de aislamiento y diferenciación sobre el sitio LS. Por ejemplo, en el encino endémico de California, *Q. lobata*, se ha observado el intercambio de flujo génico diferencial entre sitios, lo cual, se ha relacionado con la dirección y velocidad del viento (Pluess *et al.*, 2007).

Por otro lado, en las pruebas parcial de Mantel, en donde se consideraron los dos conjuntos de variables previamente identificadas en el análisis de distribución potencial; CONABIO y WorldClim, no se identificó una correlación significativa entre las distancias genéticas y el ambiente ($p\text{-valor}>0.1$). Asimismo, es interesante señalar que tampoco se observó una correlación entre las distancias geográficas y ambientales. Esto puede indicar que, las características ambientales no varían notablemente entre los sitios de muestreo de *Q. macdougallii*, lo cual, resalta el estrecho margen de variación que las condiciones ambientales deben mantener para la permanencia de la especie, lo cual, se vio reflejado en la reducción de su distribución potencial futura.

Debido a que este análisis solo se basó en promedios por sitio, se sugiere profundizar en el análisis de las asociaciones ambientales con la variación genómica, a nivel de individuo

y por loci, mediante enfoques con mayor capacidad y confiabilidad, como los reportados en *Q. rugosa* (Martins *et al.*, 2018) y en *Q. championii* (Jiang *et al.*, 2019).

Lo anterior, sugiere que en *Q. macdougallii*, el bajo gradiente de diferenciación genética, observado entre los sitios de la zona Norte y Sur, podría estar más relacionado con un patrón de aislamiento por distancias geográficas, que por las diferencias ambientales entre los sitios de estudio. No obstante, es necesario realizar un estudio más detallado de las asociaciones entre el ambiente y la variación observada, a nivel individuo y por loci, para confirmar lo anterior (Gugger *et al.*, 2018; Martins *et al.*, 2018; Jiang *et al.*, 2019).

***Q. macdougallii*, un encino endémico y su conservación.**

Los análisis de distribución potencial, mostraron que *Q. macdougallii* presenta una alta especificidad hacia determinadas variables ambientales, así como una distribución altamente restringida, la cual, abarca una zona de transición entre dos de las provincias biogeográficas más complejas del país; la provincia de Oaxaca, en la zona Norte, y la Sierra Madre del Sur, en la zona Sur (Morrone, 2005). Esta zona presenta una gran riqueza de especies y altos niveles de endemismo (Villaseñor *et al.*, 2005).

Estas características son comunes en una especie endémica (Anderson, 1994), las cuales, pueden complicar la supervivencia de la especie frente a cambios drásticos en su ambiente. Por ejemplo, Enquist (2002), investigó los cambios de bioma proyectados en Costa Rica y descubrió que varios de los que presentan el mayor número de especies endémicas, son especialmente vulnerables al calentamiento global y tienden a experimentar las mayores reducciones de su área o en algunos escenarios de cambio climático pueden llegar a desaparecer por completo.

Los análisis de distribución futura de *Q. macdougallii*, sugieren una pérdida drástica en su potencial área de permanencia. La pérdida del área de bosques templados en México por efectos del cambio climático se ha estimado en un 13% (Villers & Trejo, 1998). Asimismo, se espera que las modificaciones en la temperatura y la precipitación reduzcan significativamente los rangos actuales de distribución de aproximadamente 34 especies arbóreas de los géneros *Pinus* y *Quercus* (Gómez-Mendoza & Arriaga, 2007).

Ponce-Reyes *et al.* (2012), sugieren que, en la Sierra Norte de Oaxaca, a pesar del rápido cambio climático, será posible lograr mantener fragmentos relativamente grandes de bosque templado y bosque de niebla. Esto señala, una posibilidad de supervivencia para la especie, por lo cual es necesario impulsar estrategias de conservación para *Q. macdougallii*.

Por otra parte, los análisis de variación genómica mostraron que la especie presenta una moderada diversidad genética, lo cual, puede proporcionar cierto grado de adaptabilidad a la especie para asegurar la supervivencia de al menos algunos individuos durante eventos extremos (Meireles *et al.*, 2017). Sin embargo, es importante resaltar que, en las especies de vida larga, como los encinos, sus tiempos de respuesta y adaptación suelen complicar su permanencia (Neale & Kremer, 2011; Kremer *et al.*, 2012), por ello, estos estudios genéticos y ecológicos representan una oportunidad para proponer líneas de conservación.

Los resultados de este estudio, sugieren la implementación de medidas de conservación y monitoreo de *Q. macdougallii* en toda su distribución, debido a todas las características ecológicas y genéticas que se conocen hasta ahora de este encino endémico, el cual, se distribuye en una zona que además de presentar altos niveles de diversidad y endemismos, y una compleja historia biogeográfica (Morrone, 2005; Villaseñor *et al.*, 2005; Espinosa *et al.*, 2008), ha sido propuesta como área de conservación prioritaria para el género *Quercus* en Oaxaca (Ramírez-Toro *et al.*, 2017), el estado con el mayor número de especies en el país (Valencia, 2004; Romero-Rangel *et al.*, 2015).

Se sugiere priorizar la conservación de la zona Sur, ya que, de acuerdo con los análisis de distribución futura, será el área más afectada, y considerando los análisis de variación genómica, es en esta zona donde podemos encontrar individuos con los dos grupos genéticos identificados en este estudio. Asimismo, se sugiere un monitoreo de la dinámica poblacional de la especie y de su variación genética, para evaluar sus respuestas ante el estrés ocasionado por los cambios drásticos en su ambiente, en especial, en el sitio LS, donde solo se encontraron cinco individuos (Mimura *et al.*, 2017).

10. CONCLUSIONES

- Se observan niveles moderados de diversidad genética para *Q. macdougallii*.
- La especie se comporta como una sola población, sin embargo, se observa un gradiente de variación a lo largo de su distribución de Norte a Sur, así como un grupo genético con una incipiente estructuración en la zona Sur.
- Los SNPs *outliers* no muestran evidencia sustentable de genotipos adaptados a nivel local, sin embargo, a nivel regional, los resultados sugieren la presencia de alelos asociados a las zonas de estudio, Norte y Sur.
- La especie presenta una distribución restringida, con alta especificidad hacia las variables ambientales: edafología, elevación, temperatura promedio del cuatrimestre más frío y precipitación del cuatrimestre más cálido.
- Las proyecciones al futuro bajo escenarios del cambio climático, sugieren una reducción drástica en el área con las condiciones favorables para su permanencia, siendo la zona Sur la más afectada.
- Los resultados sugieren que el bajo gradiente de diferenciación genética observado entre los sitios de la zona Norte y Sur, podría estar más relacionado con un patrón de aislamiento por distancias geográficas, que por las diferencias ambientales entre los sitios de estudio. No obstante, es necesario realizar un estudio más detallado de las asociaciones entre el ambiente y la variación observada, a nivel individuo y por loci.
- Se sugiere priorizar la conservación y monitoreo de la especie en todo su rango de distribución, particularmente en la zona Sur, donde se observaron los dos grupos genéticos y donde los efectos del cambio climático posiblemente serán los más drásticos.

PERSPECTIVAS Y APLICACIONES

Este estudio, es el primero en explorar la variación genómica de un encino endémico de México, mediante el uso de una técnica de secuenciación masiva (NGS). Esto no solo proporciona una nueva perspectiva en la genómica de poblaciones de *Q. macdougalii*, también constituye un aporte significativo al estudio a nivel genoma de los encinos, en especial para un país que alberga el mayor número de especies del género *Quercus* en el mundo.

Por otro lado, a partir de la integración entre los análisis de variación genómica y los modelos de distribución, se sugiere una conjetura inicial sobre el futuro de la especie, sin embargo, actualmente no hay un conjunto estándar de parámetros bioinformáticos para la identificación de SNPs en especies arbóreas, ya que estos análisis dependen del organismo estudiado, por lo cual, los resultados deben ser considerados con cautela, ya que aún falta profundizar en el uso de las herramientas bioinformáticas y sus parámetros (Chen *et al.*, 2018; Pina-Martins *et al.*, 2019).

El individuo LS_30, del sitio LS, de la zona Sur, presentó diferencias que sugieren podría pertenecer a otra especie, ser un híbrido o que sus diferencias son por un error durante la secuenciación. Por lo cual, se sugiere realizar un análisis más detallado de este individuo, por ejemplo, un análisis de caracteres morfológicos, y la comparación de sus secuencias obtenidas en este estudio, con otras bases de datos genéticos (Cannon *et al.*, 2018).

Los datos genómicos obtenidos en este estudio pueden dar paso a los análisis de genómica funcional, por ejemplo, caracterizando genes funcionales que son críticos para la resiliencia de la especie frente a los drásticos cambios ambientales (Petit *et al.*, 2013; Lesur *et al.*, 2015; Gugger *et al.*, 2016; Konar *et al.*, 2016; Sork *et al.*, 2016b; Plomion *et al.*, 2018). Por otra parte, estos datos también permitirían investigar SNPs asociados a interacciones bióticas con otras especies, como hongos e insectos (Hersch-Green *et al.*, 2011).

Otro posible estudio sería acerca de sus relaciones filogenéticas con otros encinos, lo cual podría brindar más información acerca del pasado evolutivo del género, al integrar datos de una especie endémica en México (Cavender-Bares *et al.*, 2018; Hipp *et al.*, 2018).

Con la integración de los datos generados en este estudio y otros enfoques, como el ecológico, funcional y social, se podrían conocer y predecir, con mayor certeza, las respuestas de la especie frente a las presiones ambientales futuras, y así promover estrategias adecuadas para la conservación de *Q. macdougalii*.

Por ejemplo, se podrían identificar ciertos genotipos que respondan favorablemente a determinadas condiciones ambientales y promover su migración asistida hacia zonas que, por el cambio climático, se espera que cambien hacia esas condiciones (Neale & Kremer, 2011).

Los scripts utilizados en este estudio se pueden encontrar en el repositorio de GitHub: <https://github.com/NellyJazminPC/Population-Genomics>.

LITERATURA CITADA

- Abrams, M.D. (2003). Where has all the white oak gone? *BioScience* 53(10): 927-939.
- Agarwal, M., Shrivastava, N., & Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant cell reports* 27(4): 617-631.
- Alberto, F.J., Derory, J., Boury, C., Frigerio, J.M., Zimmermann, N.E., & Kremer, A. (2013). Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. *Genetics* 195(2): 495-512.
- Alexander, D.H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC bioinformatics* 12(1): 246.
- Alexander, D.H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 19(9): 1655-1664.
- Alfonso-Corrado, C., Clark-Tapia, R., Monsalvo-Reyes, A., Rosas-Osorio, C., González-Adame, G., *et al.* (2014). Ecological-Genetic Studies and Conservation of Endemic *Quercus sideroxyla* (Trel.) in Central Mexico. *Natural Resources* 5: 442-453.
- Alfonso-Corrado, C., Esteban-Jiménez, R., Clark-Tapia, R., Piñero, D., Campos, J.E., & Mendoza, A. (2004). Clonal and Genetic Structure of Two Mexican Oaks: *Quercus eduardii* and *Q. potosina* (Fagaceae). *Evolutionary Ecology* 18: 585-599.
- Alfonso-Corrado, C., Naranjo-Luna, F., Clark-Tapia, R., Campos, J., Rojas-Soto, O., *et al.* (2017). Effects of environmental changes on the occurrence of *Oreomunnea mexicana* (Juglandaceae) in a biodiversity hotspot Cloud Forest. *Forests* 8(8): 261.
- Allendorf, F.W., Hohenlohe, P.A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature reviews genetics* 11(10): 697.
- Álvarez, S., & Rubio, A. (2013). Línea base de carbono en bosque mixto de pino-encino de la Sierra Juárez (Oaxaca, México): Aplicación del modelo CO2FIX v. 3.2. *Revista Chapingo. Serie ciencias forestales y del ambiente* 19(1): 125-137.
- Anacleto-Carmona, E. (2015). Abundancia y distribución de *Quercus macdougalii* (Fagaceae) especie endémica de la Sierra Juárez, Oaxaca. Tesis de Licenciatura. Universidad de la Sierra Juárez, Ixtlán de Juárez, Oaxaca, México.
- Anderson, S. (1994). Area and endemism. *The Quarterly Review of Biology* 69: 451-471.
- Anta-Fonseca, S., Galindo-Leal, C., González-Medrano, F., Koleff-Osorio, P., Meave del Castillo, J., *et al.* (2010). Sierra Norte de Oaxaca. En: El bosque mesófilo de montaña en México; Amenazas y Oportunidades Para su Conservación y su Manejo Sostenible. (1ª ed.). CONABIO: México D.F., México. pp. 108-115.
- Antúnez, P., Hernández-Díaz, J., Wehenkel, C., & Clark-Tapia, R. (2017). Generalized models: an application to identify environmental variables that significantly affect the abundance of three tree species. *Forests* 8(3): 59.

Appleby, N., Edwards, D., & Batley, J. (2009). New technologies for ultra-high throughput genotyping in plants. En: Somers DJ, Langridge P, Gustafson JP (Eds.) *Plant genomics*. Humana Press, Kentucky. pp. 19-40.

Araujo, M.B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of biogeography* 33(10): 1677-1688.

Axelrod, D. (1983). Biogeography of oaks in the Arcto-tertiary providence. *Annals Missouri Botanical Garden* 70: 629-657.

Bacon, J.R. (1998). La calidad genética de los encinos (*Quercus* spp.) en la Sierra Madre Occidental de México e implicaciones para su manejo forestal. *Ubamari* 15: 122-140.

Bamshad, M., Wooding, S., Salisbury, B.A., & Stephens, J.C. (2004). Deconstructing the relationship between genetics and race. *Nature reviews genetics* 5(8): 598.

Banda, K., Delgado-Salinas, A., Dexter, K.G., Linares-Palomino, R., Oliveira-Filho, A., *et al.* (2016). Plant diversity patterns in neotropical dry forests and their conservation implications. *Science* 353: 1383-1387.

Barrick, J.E., & Lenski, R.E. (2013). Genome dynamics during experimental evolution. *Nature Reviews Genetics* 14(12): 827.

Brookfield, J.F.Y. (2001). Population genetics: the signature of selection. *Current Biology* 11(10): R388-R390.

Brumfield, R.T., Beerli, P., Nickerson, D.A., & Edwards, S.V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution* 18.5: 249-256.

Caldu-Primo, J.L., Mastretta-Yanes, A., Wegier, A., & Piñero, D. (2017). Finding a needle in a haystack: Distinguishing Mexican maize landraces using a small number of SNPs. *Frontiers in genetics* 8: 45.

Caldu-Primo, J.L., Mastretta-Yanes, A., Wegier, A., & Piñero, D. (2017) Data from: Finding a needle in a haystack: distinguishing Mexican maize landraces using a small number of SNPs. Dryad Digital Repository. <https://doi.org/10.5061/dryad.j2n8q>.

Cannon, C.H., Brendel, O., Deng, M., Hipp, A.L., Kremer, A., *et al.* (2018). Gaining a global perspective on Fagaceae genomic diversification and adaptation. *New Phytologist* 218: 894-897.

Cappa, E.P., El-Kassaby, Y.A., Garcia, M.N., Acuña, C., Borralho, N.M., *et al.* (2013). Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. *PLoS One* 8(11): e81267.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology* 22(11): 3124-3140.

Cavender-Bares, J. (2007). Chilling and freezing stress in live oaks (*Quercus* section *Virentes*): intra and interspecific variation in PS II sensitivity corresponds to latitude of origin. *Photosynthesis Research* 94: 437-453.

Cavender-Bares, J. (2016). Diversity, distribution and ecosystem services of the North American oaks. *International Oaks* 27: 37-49.

Cavender-Bares, J., Kothari, S., Meireles, J.E., Kaproth, M.A., Manos, P.S., & Hipp, A.L. (2018). The role of diversification in community assembly of the oaks (*Quercus* L.) across the continental U.S. *American Journal of Botany* 105: 565-586.

Cavender-Bares, J., Sack, L. & Savage, J. (2007). Atmospheric and soil drought reduce nocturnal conductance in live oaks. *Tree Physiology* 27: 611-620.

Chakraborty, R. (1993). Analysis of genetic structure of populations: meaning, methods, and implications. En: *Human population genetics*. Springer, Boston, MA. pp. 189-206.

Charlesworth, D., & Wright, S. (2001). Breeding systems and genome evolution. *Current opinion in genetics and development* 11: 685-690.

Chatwin, W.B., Carpenter, K.K., Jimenez, F.R., Elzinga, D.B., Johnson, L.A., & Maughan, P.J. (2014). Microsatellite primer development for post oak, *Quercus stellata* (Fagaceae). *Applications in plant sciences* 2(10): 1400070.

Cheeke, T.E., Phillips, R.P., Brzostek, E.R., Rosling, A., Bever, J.D., & Fransson, P. (2017). Dominant mycorrhizal association of trees alters carbon and nutrient cycling by selecting for microbial groups with distinct enzyme function. *New Phytologist* 214: 432-442.

Chen, F., Dong, W., Zhang, J., Guo, X., Chen, J., *et al.* (2018). The sequenced angiosperm genomes and genome databases. *Frontiers in plant science* 9: 418.

Clark-Tapia, R., Mendoza-Ochoa, A., Aguirre-Hidalgo, V., Antúnez, P., Campos-Contreras, J.E., *et al.* (2018). Reproducción sexual de *Quercus macdougalii*, un encino endémico de la Sierra Juárez, Oaxaca. *Madera y Bosques* 24(2).

Clark-Tapia, R., Suárez-Mota, M.E., Matías-Nolasco, M., & Velasco-Hipólito, F. (2016). Clima: pasado, presente y futuro. En: Clark-Tapia, R., Ramos-Morales, M.F., Alfonso-Corradó, C., Mendoza-Díaz, M.M. & Fuente-Carrasco, M. (Eds.), *Recursos hídricos de la Sierra Norte de Oaxaca: caracterización, diagnóstico y gestión*. México. Universidad de la Sierra Juárez. pp. 38-47.

Clayton, D. (2015). Imputed SNP analyses and meta-analysis with snpStats.

Collins, S. (2011). Many possible worlds: expanding the ecological scenarios in experimental evolution. *Evolutionary Biology* 38(1): 3-14.

Comisión Nacional para el Conocimiento y Uso de la Biodiversidad [CONABIO]. (1997). 'Provincias biogeográficas de México'. Escala 1:4 000 000. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, México, D. F.

Contreras-Medina, R., Luna-Vega, I., & Ríos-Muñoz, C. (2010). Distribución de *Taxus globosa* (Taxaceae) en México: Modelos Ecológicos de Nicho, Efectos del Cambio de Uso de Suelo y Conservación. *Revista chilena de historia natural* 83(3): 421-433.

Cuervo-Alarcon, L., Arend, M., Müller, M., Sperisen, C., Finkeldey, R., & Krutovsky, K.V. (2018). Genetic variation and signatures of natural selection in populations of European beech (*Fagus sylvatica* L.) along precipitation gradients. *Tree Genetics & Genomes* 14(6): 84.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., *et al.* (2011). The variant call format and VCFtools. *Bioinformatics* 27(15): 2156-2158.

De Kort, H., Vandepitte, K., Bruun, H.H., Closset-Kopp, D., Honnay, O., & Mergeay, J. (2014). Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Molecular Ecology* 23(19): 4709-4721.

De Luna-Bonilla, O. A. (2018). Análisis de la diversidad genética de tres especies del género *Quercus* sección *Lobatae*, del grupo *Acutifoliae* (Fagaceae). Tesis de Maestría. Universidad Nacional Autónoma de México, Cd. Mx.

Denk, T., Grimm, G.W., Manos, P.S., Deng, M., & Hipp, A.L. (2017). An updated infrageneric classification of the oaks: review of previous taxonomic schemes and synthesis of evolutionary patterns. En: Gil-Pelegrín, E., Peguero-Pina, J.J., & Sancho-Knapik, D. (Eds.), *Oaks physiological ecology. Exploring the functional diversity of genus Quercus* L. Cham, Switzerland: Springer International Publishing. pp. 13-38.

Diario Oficial de la Federación. (2018). Proyecto de Modificación del Anexo Normativo III, Lista de especies en riesgo de la Norma Oficial Mexicana NOM-059-SEMARNAT-2010, Protección ambiental-Especies nativas de México de flora y fauna silvestres-Categorías de riesgo y especificaciones para su inclusión, exclusión o cambio-Lista de especies en riesgo, publicada el 30 de diciembre de 2010. Recuperado de http://diariooficial.gob.mx/nota_detalle.php?codigo=5534594 &fecha=13/08/2018.

Dickie, I.A., Schnitzer, S.A., Reich, P.B., & Hobbie, S.E. (2007). Is oak establishment in old-fields and savanna openings context dependent? *Journal of Ecology* 95: 309-320.

Dobzhansky, T., & Pavlovsky, O. (1957). An Experimental Study of Interaction between Genetic Drift and Natural Selection. *Evolution* 11(3): 311-319.

Dray, S., & Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software* 22(4): 1-20.

Edwards, D., Forster, J., Chagné, D., & Batley, J. (2007). What are SNPs? En: Oraguzie, N., Rikkerink, E., Gardiner, S. & Silva, H. (Eds.) *Association mapping in plants*. Springer, New York. pp. 41-52.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., *et al.* (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29(2): 129-151.

Ellstrand, N. C. (2014). Is gene flow the most important evolutionary force in plants? *American journal of botany* 101(5): 737-753.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., *et al.* (2011). A robust, simple Genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS one* 6(5):1-9.

Enquist, C.A.F. 2002. Predicted regional impacts of climate change on the geographical distribution and diversity of tropical forests in Costa Rica. *Journal of Biogeography* 29:519-534.

Eriksson, G. (1998). Evolutionary forces influencing variation among populations of *Pinus sylvestris*. *Silva Fennica* 32: 173-184.

Espinosa, D., Ocegueda, S., Aguilar, C., Flores, Ó., Llorente-Bousquets, J., & Vázquez, B. (2008). El conocimiento biogeográfico de las especies y su regionalización natural. En: Sarukhán, J. (Coord. Gral.). *Capital natural de México. Vol. I: Conocimiento actual de la biodiversidad*. Comisión Nacional para el conocimiento y Uso de la Biodiversidad. México, D.F., México. pp. 33-65.

Fady, B., Cottrell, J., Ackzell, L., Alía, R., Muys, B., *et al.* (2016). Forests and global change: what can genetics contribute to the major forest management and policy challenges of the twenty-first century? *Regional Environmental Change* 16(4): 927-939.

Fick, S.E., & Hijmans, R.J. (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*.

Firmat, C., Delzon, S., Louvet, J.M., Parmentier, J., & Kremer, A. (2017). Evolutionary dynamics of the leaf phenological cycle in an oak metapopulation along an elevation gradient. *Journal of Evolutionary Biology* 30(12): 2116-2131.

Frich, P., Alexander, L.V., Della-Marta, P., Gleason, B., Haylock, M., Klein Tank, A.M.G., & Peterson, T. (2002). Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Research* 19(3): 193-212.

Fu, Y.B., & Yang, M.H. (2017). Genotyping-by-sequencing and its application to oat genomic research. En: Gasparis, S. (Ed), *Oat-methods and protocols*. Springer, New York. pp. 169-187.

Ghilardi, A. (2006). 'Existencia de leña en México a nivel municipal', escala: 1:250000. Centro de Investigaciones en Ecosistemas (CIEco), Universidad Nacional Autónoma de México Campus Morelia. Proyecto: FQ003, Extraído del proyecto FQ003: Servicios Ecosistémicos de México: patrones, tendencias y prioridades de investigación. Financiado por la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) México, DF. Michoacán, México.

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., *et al.* 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* 9(2): e90346.

Gómez-Mendoza, L., & Arriaga, L. (2007). Modeling the effect of climate change on the distribution of oak and pine species of Mexico. *Conservation Biology* 21(6): 1545-1555.

Gómez-Mendoza, L., Vega-Peña, E., Ramírez, M. I., Palacio-Prieto, J. L., & Galicia, L. (2006). Projecting land-use change processes in the Sierra Norte of Oaxaca, Mexico. *Applied Geography* 26(3-4): 276-290.

González-Martínez, S.C., Krutovsky, K.V., & Neale, D.B. (2006). Forest-tree population genomics and adaptive evolution. *New Phytologist* 170(2): 227-238.

Gorgonio-Ramírez, M. (2012). Variabilidad y Estructura Genética de *Quercus eduardii* (Fagaceae) en Sierra Fría, Aguascalientes, Oaxaca, México. Tesis de Licenciatura. Universidad de la Sierra Juárez, Ixtlán de Juárez, Oaxaca, México.

Gorgonio-Ramírez, M. (2015). Ecología y genética de *Quercus crassifolia* (Fagaceae) en sitios de manejo forestal y de uso local en Sierra Juárez, Oaxaca. Tesis de Maestría. Universidad de la Sierra Juárez, Ixtlán de Juárez, Oaxaca, México.

Gorgonio-Ramírez, M., Clark Tapia, R., Campos, J.E., Monsalvo-Reyes, A., & Alfonso-Corrado, C. (2017). Diversidad y estructura genética de *Quercus crassifolia* en sitios de manejo forestal y uso local en Sierra Juárez, Oaxaca. *Madera y bosques* 23(2): 85-98.

Grivet, D., Sork, V.L., Westfall, R.D., & Davis, F.W. (2008). Conserving the evolutionary potential of California valley oak (*Quercus lobata* Née): A multivariate genetic approach to conservation planning. *Molecular Ecology* 17(1): 139-156.

Guerrero-Sánchez, V.M., Maldonado-Alconada, A.M., Amil-Ruiz, F., Verardi, A., Jorrín-Novio, J.V., & Rey, M.D. (2019). Ion Torrent and Illumina, two complementary RNA-seq platforms for constructing the holm oak (*Quercus ilex*) transcriptome. *PloS one* 14(1): e0210356.

Gugger, P.F., Cokus, S.J., & Sork, V.L. (2016). Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*). *Tree Genetics and Genomes* 12(2): 15.

Gugger, P.F., Peñaloza-Ramírez, J.M., Wright, J.W., & Sork, V.L. (2017). Whole-transcriptome response to water stress in a California endemic oak, *Quercus lobata*. *Tree Physiology* 37(5): 632-644.

Gugger, P.F., Liang, C.T., Sork, V.L., Hodgskiss, P., & Wright, J.W. (2018). Applying landscape genomic tools to forest management and restoration of Hawaiian koa (*Acacia koa*) in a changing environment. *Evolutionary applications* 11(2): 231-242.

Guichoux, E., Garnier-Géré, P., Lagache, L., Lang, T., Boury, C., & Petit, R.J. (2013). Outlier loci highlight the direction of introgression in oaks. *Molecular Ecology* 22(2): 450-462.

Guisan, A., & Zimmermann, N.E. (2000). Predictive habitat distribution model in ecology. *Ecological Modelling* 135: 147-186.

Hedrick, P.W. (2005). A standardized genetic differentiation measure. *Evolution* 59(8): 1633-1638.

Hedrick, P.W. (2011). Genetics of populations: Forth edition. Sudbury, MA: Jones & Bartlett Learning.

Hedrick, P.W., & Hurt, C.R. (2012). Conservation genetics and evolution in an endangered species: research in Sonoran topminnows. *Evolutionary applications* 5(8): 806-819.

Hernández-Ruíz, J., Herrera-Cabrera, B.E., Delgado-Alvarado, A., Salazar-Rojas, V.M., Bustamante-Gonzalez, Á., *et al.* (2016). Distribución potencial y características geográficas de poblaciones silvestres de *Vanilla planifolia* (Orchidaceae) en Oaxaca, México. *Revista de Biología tropical* 64(1): 235-246.

Hersch-Green, E.I., Turley, N.E., & Johnson, M.T. (2011). Community genetics: what have we accomplished and where should we be going? *Philosophical Transactions of the Royal Society B: Biological Sciences* 366(1569): 1453-1460.

Hipp, A.L., Manos, P.S., González-Rodríguez, A., Hahn, M., Kaproth, M., *et al.* (2018). Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. *New Phytologist* 217(1): 439-452.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., & Jarvis, A. (2005). Very high-resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society* 25(15): 1965-1978.

Holliday, J.A., Aitken, S.N., Cooke, J.E., Fady, B., González-Martínez, S.C., *et al.* (2017). Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. *Molecular ecology* 26(3): 706-717.

Homolka, A., Schueler, S., Burg, K., Fluch, S., & Kremer, A. (2013). Insights into drought adaptation of two European oak species revealed by nucleotide diversity of candidate genes. *Tree genetics & genomes* 9(5): 1179-1192.

Hortal, J., Lobo, J.M., & Jiménez-Valverde, A. (2012). Basic questions in biogeography and the (lack of) simplicity of species distributions: putting species distribution models in the right place. *Natureza & Conservação* 10:108-118.

INEGI (1998) Condiciones generales del medio ambiente: Suelos. En: Estadísticas del Medio Ambiente, 1997. Informe de la situación general en materia de equilibrio ecológico y protección al ambiente, 1995-1996. México. pp. 50-56.

INEGI. (2000). Instituto Nacional de Estadística Geografía e Informática. Modelo digital del terreno. Escala 1:250,000. México: INEGI.

IPCC (2007). Intergovernmental Panel on Climate Change. Climate change: synthesis report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change core writing team (IPCC: Geneva, Switzerland).

Işık, K. (2011). Rare and endemic species: why are they prone to extinction? *Turkish Journal of Botany* 35(4): 411-417.

Jarvis, A., Yeaman, S., Guarino, L., & Tohme, J. (2005). The role of geographic analysis in locating, understanding and using plant genetic diversity. *Methods in Enzymology* 395: 279-298.

Jiang, X.L., Gardner, E.M., Meng, H.H., Deng, M., & Xu, G.B. (2019). Land bridges in the Pleistocene contributed to flora assembly on the continental islands of South China: Insights from the evolutionary history of *Quercus championii*. *Molecular phylogenetics and evolution* 132: 36-45.

Jombart, T., & Ahmed, I. (2011). Adegnet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27(21): 3070-3071.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics* 11(1): 94.

Kamvar, Z.N., Tabima, J.F., & Grünwald, N.J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281.

Karmalkar, A.V., Bradley, R.S., & Diaz, H.F. (2011). Climate change in Central America and Mexico: Regional climate model validation and climate change projections. *Climate Dynamics* 37(3-4): 605-629.

Kawecki, T.J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology letters* 7(12): 1225-1241.

Kisdi, É. (2002). Dispersal: risk spreading versus local adaptation. *The American Naturalist* 159(6): 579-596.

Klug, W.S., & Cummings, M.R. (2013). *Conceptos de genética: Décima edición*. Pearson.

Knaus, B.J., & Grünwald, N.J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources* 17(1): 44-53.

Konar, A., Choudury, O., Gailing, O., Coggeshall, M.V., Staton, M.E., Emrich, S., Carlson, J., & Romero-Severson, J. (2016). A genetic map for the *Lobatae*. *International Oaks* 27: 181-189.

Kremer, A., Abbott, A.G., Carlson, J.E., Manos, P.S., Plomion, C., *et al.* (2012). Genomics of Fagaceae. *Tree Genetics & Genomes* 8(3): 583-610.

Kroeger, T., Escobedo, F.J., Hernandez, J.L., Varela, S., Delphin, S., *et al.* (2014). Reforestation as a novel abatement and compliance measure for ground-level ozone. *Proceedings of the National Academy of Sciences USA* 111(40): E4204-E4213.

Lachagari, V.B., Gupta, R., Lekkala, S.P., Mahadevan, L., Kuriakose, *et al.* (2019). Whole genome sequencing and comparative genomic analysis reveal allelic variations unique to a purple colored rice landrace (*Oryza sativa* ssp. *indica* cv. Purpleputtu). *Frontiers in plant science* 10: 513.

Langmead, B., & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods* 9(4): 357.

Larrañaga, N., & Hormaza, J.I. (2016). Advances in genetic diversity analysis in fruit tree crops. En: Lüttge, U., Cánovas, F.M., Matyssek, R. (Eds.) *Progress in botany* 77. Springer International Publishing. Switzerland. pp. 245-264.

Laurent, F., & Vilá, M. (2003). Diversity patterns of plant functional types in relation to fire regime and previous land use in Mediterranean woodlands. *Journal of Vegetation Science* 14(3): 389-398.

Lee, T.D., Reich, P.B., & Bolstad, P.V. (2005). Acclimation of leaf respiration to temperature is rapid and related to specific leaf area, soluble sugars and leaf nitrogen across three temperate deciduous tree species. *Functional Ecology* 19(4): 640-647.

Lehtomäki, J., & Moilanen, A. (2013). Methods and workflow for spatial conservation prioritization using Zonation. *Environmental Modelling and Software* 47: 128-137.

Leroy, T., Louvet, J.M., Lalanne, C., Le Provost, G., Labadie, K., *et al.* (2019). Adaptive introgression as a driver of local adaptation to climate in European white oaks. *bioRxiv* 584847.

Lesur, I., Le Provost, G., Bento, P., Da Silva, C., Leplé, J.C., *et al.* (2015). The oak gene expression atlas: insights into Fagaceae genome evolution and the discovery of genes regulated during bud dormancy release. *BMC genomics* 16(1): 112.

Letelier, A. M. (2007). Tamaño efectivo de la población. En: Eguiarte, L.E., Souza, V. & Aguirre, X. (Eds.) *Ecología molecular*. INE, SEMARNAT, CONABIO, UNAM. México, D.F. pp. 63-86.

Li, J., Li, H., Jakobsson, M., Li, S.E.N., Sjödin, P.E.R., & Lascoux, M. (2012). Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Molecular ecology* 21(1): 28-44.

Limousin, J.M., Longepierre, D., Huc, R., & Rambal, S. (2010). Change in hydraulic traits of Mediterranean *Quercus ilex* subjected to long-term throughfall exclusion. *Tree Physiology* 30(8): 1026-1036.

Loveless, M.D., & Hamrick, J.L. (1984). Ecological determinants of genetic structure in plant populations. *Annual review of ecology and systematics* 15(1): 65-95.

Luikart, G., England, P.R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature reviews genetics* 4(12): 981-994.

Manos, P.S., Doyle, J.J., & Nixon, K.C. (1999). Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Molecular phylogenetics and evolution* 12(3): 333-349.

Manos, P.S., & Stanford, A.M. (2001). The historical biogeography of Fagaceae: tracking the tertiary history of temperate and subtropical forests of the Northern Hemisphere. *International Journal of Plant Sciences* 162(S6): S77-S93.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209-220.

Mariette, S., Le Corre, V., Austerlitz, F., & Kremer, A. (2002). Sampling within the genome for measuring within-population diversity: trade-offs between markers. *Molecular Ecology* 11(7): 1145-1156.

Martínez-Meyer, E., Peterson A.T., & Hargrove W. (2004). Ecological Niches as Stable Distributional Constraints on Mammal Species, with Implications for Pleistocene Extinctions and Climate Projections for Biodiversity. *Global Ecology and Biogeography* 13(4):305-314.

Martins, K., Gugger, P.F., Llanderal-Mendoza, J., González-Rodríguez, A., Fitz-Gibbon, S. T., *et al.* (2018). Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. *Evolutionary Applications* 11(10): 1842-1858.

Mastretta-Yanes, A., Arrigo, N., Álvarez, N., Jorgensen, T.H., Piñero, D., & Emerson, B.C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular ecology resources* 15(1): 28-41.

Mateo, G.R., Felicísimo, A.M., & Muñoz, J. (2011). Modelos de distribución de las especies: una revisión sintética. *Revista Chilena de Historia Natural* 84(2): 217-240.

McKinney, G.J., Larson, W.A., Seeb, L.W., & Seeb, J.E. (2017). RAD seq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry *et al.* (2016). *Molecular ecology resources* 17(3): 356-361.

McVay, J.D., Hauser, D., Hipp, A.L., & Manos, P.S. (2017). Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in western North America. *Genome* 60(9): 733-742.

Meireles, J.E., Beulke, A., Borkowski, D.S., Romero-Severson, J., & Cavender-Bares, J. (2017). Balancing selection maintains diversity in a cold tolerance gene in broadly distributed live oaks. *Genome* 60: 762-769.

Mejías, J.A., Arroyo, J., & Marañón, T. (2007). Ecology and biogeography of plant communities associated with the post Plio-Pleistocene relict *Rhododendron ponticum* subsp. *baeticum* in southern Spain. *Journal of biogeography* 34(3): 456-472.

Mimura, M., Yahara, T., Faith, D.P., Vázquez-Domínguez, E., Colautti, R.I., *et al.* (2017). Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary applications* 10(2): 121-139.

Molina-Garay, C. (2011). Diversidad genética y estructura poblacional de *Quercus macdougalii*, encino endémico de Oaxaca. Tesis de Licenciatura. Universidad Nacional Autónoma de México, Tlalnepantla, Estado de México, México.

Morales, N. (2012). Modelos de distribución de especies: Software Maxent y sus aplicaciones en Conservación. *Revista Conservación Ambiental* 2(1): 1-3.

Moricca, S., & Ragazzi, A. (2008). Fungal endophytes in Mediterranean oak forests: a lesson from *Discula quercina*. *Phytopathology* 98(4): 380-386.

Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology* 51(2): 238-254.

Morrone, J.J. (2005). Hacia una síntesis biogeográfica de México. *Revista mexicana de biodiversidad* 76(2): 207-252.

Mueller, R.C., Scudder, C.M., Porter, M.E., Trotter, R.T., Gehring, C.A., & Whitham, T.G. (2005). Differential tree mortality in response to severe drought: evidence for long-term vegetation shifts. *Journal of Ecology* 93:1085-1093.

Müller, M., & Gailing, O. (2019). Abiotic genetic adaptation in Fagaceae. *Plant Biology* 20(5): 783-795.

Namkoong, G. (2001). Forest genetics: pattern and complexity. *Canadian Journal of Forest Research* 31(4): 623-632.

Naranjo-Luna, F.J. (2014). Ecología y genética de *Oreomunnea mexicana* (Standl.) J.F. Leroy (Juglandaceae), especie relicto del bosque de niebla de la Sierra Juárez, Oaxaca. Tesis de Maestría. Universidad de la Sierra Juárez.

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R., & Hohenlohe, P.A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular ecology* 22(11): 2841-2847.

Narum, S.R., & Hess, J.E. (2011). Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular ecology resources* 11: 184-194.

Navarro, L., & Guitian, J. (2003). Seed germination and seedling survival of two threatened endemic species of the northwest Iberian Peninsula. *Biological Conservation* 109(3): 313-320.

Neale, D.B., & Kremer, A. (2011). Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* 12(2): 111.

Nei, M. (1987). Molecular evolutionary genetics. Columbia University Press, Nueva York.

Nguyen, N.H., Williams, L.J., Vincent, J.B., Stefanski, A., Cavender-Bares, J., *et al.* (2016). Ectomycorrhizal fungal diversity and saprotrophic fungal diversity are linked to different tree community attributes in a field-based tree experiment. *Molecular ecology* 25(16): 4032-4046.

Nixon, K.C. (1998). El género *Quercus* en México. En: Ramamoorthy, T.R., Bye, R., Lot, A., & Fa, J. (Eds.), *Diversidad Biológica de México: Orígenes y Distribución*. Instituto de Biología. Universidad Nacional Autónoma de México. México. pp. 435-447.

Nixon, K.C. (2006) Global and Neotropical Distribution and Diversity of Oak (genus *Quercus*) and Oak Forests. En: Kappelle M. (Ed.) *Ecology and Conservation of Neotropical Montane Oak Forests (ecological studies)*. Springer. Berlin, Heidelberg. pp. 1-13.

Nixon, K., & Muller, C. (1997). *Quercus* Linnaeus sect. *Quercus* white oaks. En: Flora of North America Editorial Committee (Ed.), *Flora of North America North of Mexico*. Oxford University Press. New York. pp. 436-506.

Oyama, K., Ramírez-Toro, W., Peñaloza-Ramírez, J. M., Pérez Pedraza, A. E., Torres-Miranda, C. A., Ruiz-Sánchez, E., & González-Rodríguez, A. (2018). High genetic diversity and connectivity among populations of *Quercus candicans*, *Quercus crassifolia*, and *Quercus castanea* in a heterogeneous landscape in Mexico. *Tropical Conservation Science* 11: 1940082918766195.

Pariona, N., Martínez, A.I., Hernandez-Flores, H., & Clark-Tapia, R. (2017). Effect of magnetite nanoparticles on the germination and early growth of *Quercus macdougalii*. *Science of the Total Environment* 575: 869-875.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. *Annual Reviews of Ecology, Evolution, and Systematics* 37:637-669.

Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological modelling* 133(3): 225-245.

Peñaloza-Ramírez, J. M., González-Rodríguez, A., Mendoza-Cuenca, L., Caron, H., Kremer, A., & Oyama, K. (2010). Interspecific gene flow in a multispecies oak hybrid zone in the Sierra Tarahumara of Mexico. *Annals of botany* 105(3): 389-399.

Peres-Neto, P.R., Jackson, D.A., & Somers, K.M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis* 49(4): 974-997.

Peterson, A.T. (2001). Predicting species geographic distributions based on ecological niche modeling. *The Condor* 103(3): 599-605.

Petit, R.J., Carlson, J., Curtu, A.L., Loustau, M.L., Plomion, C., *et al.* (2013). Fagaceae trees as models to integrate ecology, evolution and genomics. *New Phytologist* 197(2): 369-371.

Phillips, S.J., Anderson, R.P., & Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling* 190(3-4): 231-259.

Pina-Martins, F., Baptista, J., Pappas J.G., & Paulo, O.S. (2019). New insights into adaptation and population structure of cork oak using genotyping by sequencing. *Global change biology* 25(1): 337-350.

Pingarroni, A., Molina-Garay, C., Rosas-Osorio, C., Alfonso-Corrado, C., Clark-Tapia, R., *et al.* (EN PRENSA) Abundancia y diversidad genética de *Quercus mulleri*, especie microendémica amenazada de Oaxaca. *Madera y Bosques*.

Piñero, D., Barahona, A., Eguiarte, L., Rocha, A., & Salas, R. (2008). La variabilidad genética de las especies: aspectos conceptuales y sus aplicaciones y perspectivas en México. *Capital natural de México*. pp. 415-435.

Plomion, C., Aury, J.M., Amselem, J., Alaeitabar, T., Barbe, V., *et al.* (2016a). Decoding the oak genome: public release of sequence data, assembly, annotation and publication strategies. *Molecular Ecology Resources* 16(1): 254-265.

Plomion, C., Aury, J.M., Amselem, J., Leroy, T., Murat, F., *et al.* (2018). Oak genome reveals facets of long lifespan. *Nature Plants* 4(7): 440-452.

Plomion, C., Bastien, C., Bogeat-Triboulot, M.B., Bouffier, L., Déjardin, A., *et al.* (2016b). Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science* 73(1): 77-103.

Pluess, A.R., Sork, V.L., Dolan, B., Davis, F.W., Grivet, D., *et al.* (2009). Short distance pollen movement in a wind-pollinated tree, *Quercus lobata* (Fagaceae). *Forest Ecology and Management* 258(5): 735-744.

Ponce-Reyes, R., Reynoso-Rosales, V.H., Watson, J.E., VanDerWal, J., Fuller, R.A., *et al.* (2012). Vulnerability of cloud forest reserves in Mexico to climate change. *Nature climate change* 2(6): 448.

Porth, I., & El-Kassaby, Y. (2014). Assessment of the genetic diversity in forest tree populations using molecular markers. *Diversity* 6(2): 283-295.

Primack, R.B. (2012). *A Primer of Conservation Biology*. (6th ed). Sinauer Associates. Sunderland, Massachusetts.

Pritchard, J.K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155(2): 945-959.

Raj, A., Stephens, M., & Pritchard, J.K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197(2): 573-589.

Ramírez-Marcial, N., González-Espinosa, M., & Williams-Linera, G. (2001). Anthropogenic disturbance and tree diversity in montane rain forests in Chiapas, Mexico. *Forest ecology and management* 154(1-2): 311-326.

Ramírez-Toro, W., Torres-Miranda, A., González-Rodríguez, A., Ruiz-Sanchez, E., Luna-Vega, I., & Oyama, K. (2017). A multicriteria analysis for prioritizing areas for

conservation of oaks (Fagaceae: *Quercus*) in Oaxaca, southern Mexico. *Tropical Conservation Science* 10: 1940082917714227.

Ramírez-Valiente, J.A., Deacon, N.J., Etterson, J., Center, A., Sparks, J.P., *et al.* (2018). Natural selection and neutral evolutionary processes contribute to genetic divergence in leaf traits across a precipitation gradient in the tropical oak *Quercus oleoides*. *Molecular ecology* 27(9): 2176-2192.

Rebetez, M., & Dobbertin, M. (2004). Climate change may already threaten Scots pine stands in the Swiss Alps. *Theoretical and Applied Climatology* 79:1-9.

Ringnér, M. (2008). What is principal component analysis? *Nature biotechnology* 26(3): 303.

Robson, J. (2008). Plan de manejo de la biodiversidad, Comunidad Santiago Comaltepec, Distrito de Ixtlán de Juárez, Oaxaca. p. 61.

Romero-Rangel, S., Rojas-Centeno, E.C., & Rubio-Licona, L.E. (2015). Encinos de México. (*Quercus*, Fagaceae) 100 especies (1ª ed.). Universidad Nacional Autónoma de México. México.

Rosas-Osorio, J.C., Alfonso-Corrado, C., Monsalvo-Reyes, A., Clark-Tapia, R., Lira-Saade, R., & Campos-Contreras, J. (2010). The Genetic Variability of *Quercus grisea* Liebm. En: The Sierra Fría of Aguascalientes, México. *International Oaks* 21: 64-72.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145(4): 1219-1228.

Roy, J., Mooney, H.A., & Saugier, B. (2001). Terrestrial global productivity. Elsevier.

Sáenz-Romero, C., Rehfeldt, G.E., Crookston, N.L., Duval, P., St-Amant, R., Beaulieu, J., & Richardson, B.A. (2009). Spline models of contemporary, 2030, 2060 and 2090 climates for Mexico and their use in understanding climate-change impacts on the vegetation. *Climatic Change* 102(3-4): 595-623.

Singh, N., Choudhury, D.R., Singh, A.K., Kumar, S., Srinivasan, K., Tyagi, R.K., & Singh, R. (2013). Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS One* 8(12): e84136.

Skelton, R.P., Dawson, T.E., Thompson, S.E., Shen, Y., Weitz, A.P., & Ackerly, D. (2018). Low vulnerability to xylem embolism in leaves and stems of North American oaks. *Plant physiology* 177(3): 1066-1077.

Slatkin, M. (1985). Rare alleles as indicators of gene flow. *Evolution* 39(1): 53-65.

Slatkin, M. 1994. Gene flow and population structure. En: Real, L. (Ed.) *Ecological Genetics*. Princeton. NJ: Princeton University Press. pp. 3-17.

Soberón, J., Osorio-Olvera, L., & Peterson, T. (2017). Diferencias entre modelación de nichos y de áreas de distribución. *Revista Mexicana de Biodiversidad* 88(2): 437-441.

Sork, V.L., Fitz-Gibbon, S.T., Puiu, D., Crepeau, M., Gugger, P.F., Sherman, R., Stevens, K., Langley, C.H., Pellegrini, M., & Salzberg, S.L. (2016a). First draft assembly and annotation of the genome of a California endemic oak *Quercus lobata* Née (Fagaceae). *G3: Genes, Genomes, Genetics* 6(11): 3485-3495.

Sork, V.L., Squire, K., Gugger, P.F., Steele, S.E., Levy, E.D., & Eckert, A.J. (2016b). Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *American Journal of Botany* 103(1): 33-46.

Sork, V.L., Stowe, K.A., & Hochwender, C. (1993). Evidence for local adaptation in closely adjacent subpopulations of northern red oak (*Quercus rubra* L.) expressed as resistance to leaf herbivores. *The American Naturalist* 142(6): 928-936.

Stapley, J., Reger, J., Feulner, P. G., Smadja, C., Galindo, J., *et al.* (2010). Adaptation genomics: the next generation. *Trends in ecology & evolution* 25(12): 705-712.

Stocker, T., Qin, D., Plattner, G., Tignor, M., Allen, S., *et al.* (2013). IPCC, 2013. Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change.

Truong, H.T., Ramos, A.M., Yalcin, F., de Ruiter, M., van der Poel, H.J., *et al.* (2012). Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. *PloS one* 7(5): e37565.

Unión Internacional para la Conservación de la Naturaleza [UICN] (2010). Lista Roja. Recuperado de <http://www.iucn.org>

Valencia, S.A. (2004). Diversidad del género *Quercus* (Fagaceae) en México. *Boletín de la sociedad Botánica de México* (75): 33-53.

Vasemägi, A., & Primmer, C.R. (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology* 14: 3623-3642.

Varela, S., Terribile, L.C., de Oliveira, G., Diniz-Filho, J.A.F., González-Hernández, J., & Lima-Ribeiro, M.S. (2015). ecoClimate vs. Worldclim: variables climáticas SIG para trabajar en biogeografía. *Revista Ecosistemas* 24(3): 88-92.

Villers, L., & Trejo, I. (1998). El impacto del cambio climático en los bosques y áreas naturales protegidas de México. *Interciencia* 23:10-19.

Villaseñor, J.L., Maeda, P., López, J.J. C., & Ortiz, E. (2005). Estimación de la riqueza de especies de Asteraceae mediante extrapolación a partir de datos de presencia-ausencia. *Boletín de la Sociedad Botánica de México* (76): 518.

Villaseñor, J.L., & Téllez-Valdés, O. (2004). Distribución potencial de las especies del género *Jefea* (Asteraceae) en México. *Anales del Instituto de Biología. Universidad Nacional Autónoma de México. Serie Botánica* 75(2): 205-220.

Wakeley, J. (1996). The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends in ecology & evolution* 11(4): 158-162.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., *et al.* (2011). MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development* 4(4): 845.

Weigel, D., & Nordborg, M. (2015). Population genomics for understanding adaptation in wild plant species. *Annual review of genetics* 49: 315-338.

Weiser, E.L., Grueber, C.E., & Jamieson, I.G. (2013). Simulating Retention of Rare Alleles in Small Populations to Assess Management Options for Species with Different Life Histories. *Conservation Biology* 27: 335-344.

Wright, S. (1943). Isolation by distance. *Genetics* 28(2): 114-138.

Wright, S. (1951). The Genetical Structure of Populations. *Annals Eugenetic* 15: 323-354.

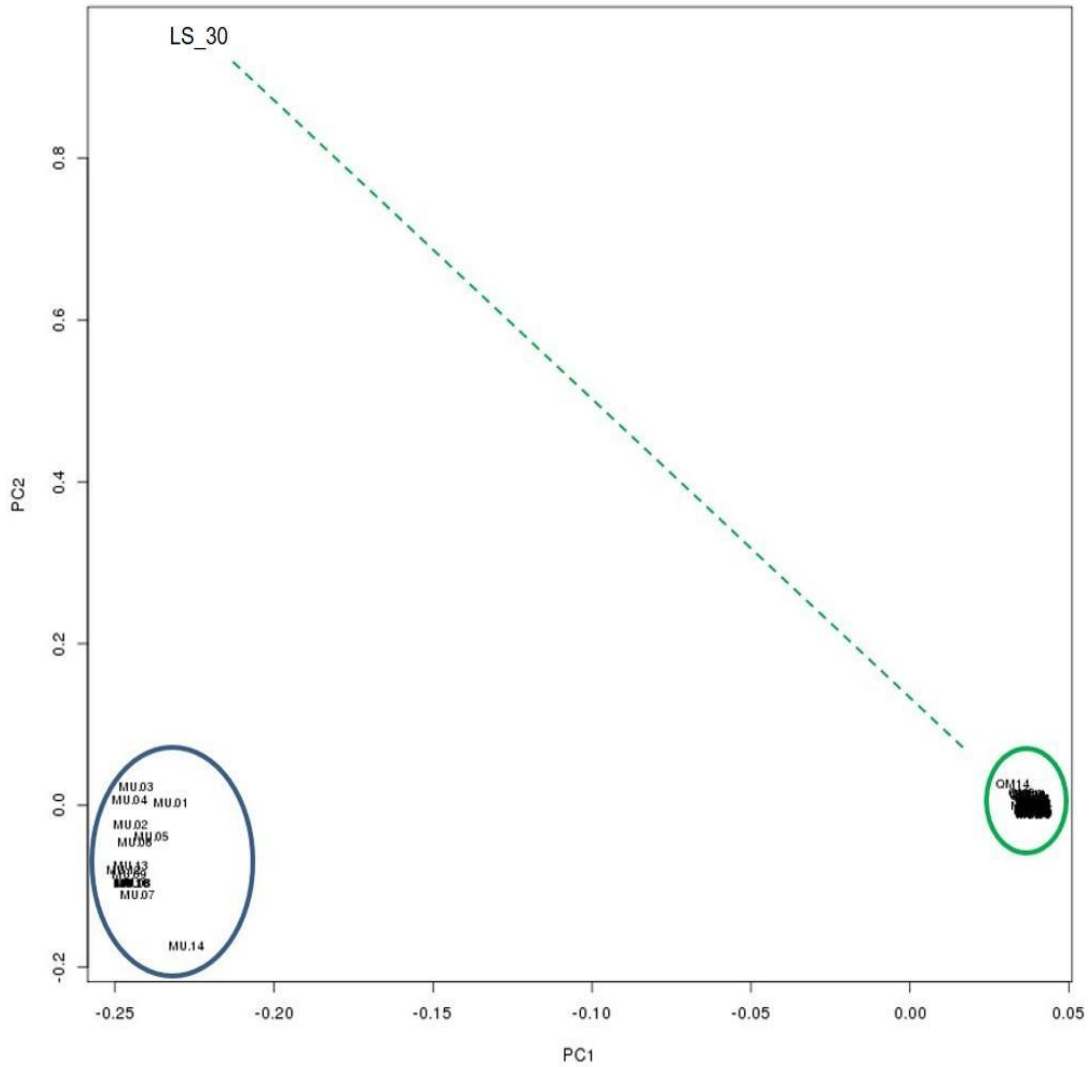
Yanga, X.Q., Kushwaha, S.P.S., Saran, S., Xu, J., & Roy, P.S. 2013. Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in Lesser Himalayan foothills. *Ecological engineering* 51: 83-87.

Zhang, Q., Tyler-Smith, C., & Long, Q. (2015). An extended Tajima's D neutrality test incorporating SNP calling and imputation uncertainties. *Statistics and its interface* 8(4): 447.

Zeigler, D. (2014). Natural selection. En: Zeigler, D. (Ed.), *Evolution: components and mechanisms*. Academic Press. Elsevier. pp. 9-22.

Anexo I

Análisis de Componentes Principales entre los 80 individuos de *Q. macdougalii* (sección *Quercus*) y 14 individuos de *Q. mullerii* (sección *Lobatae*). El primero en verde y el segundo en azul.



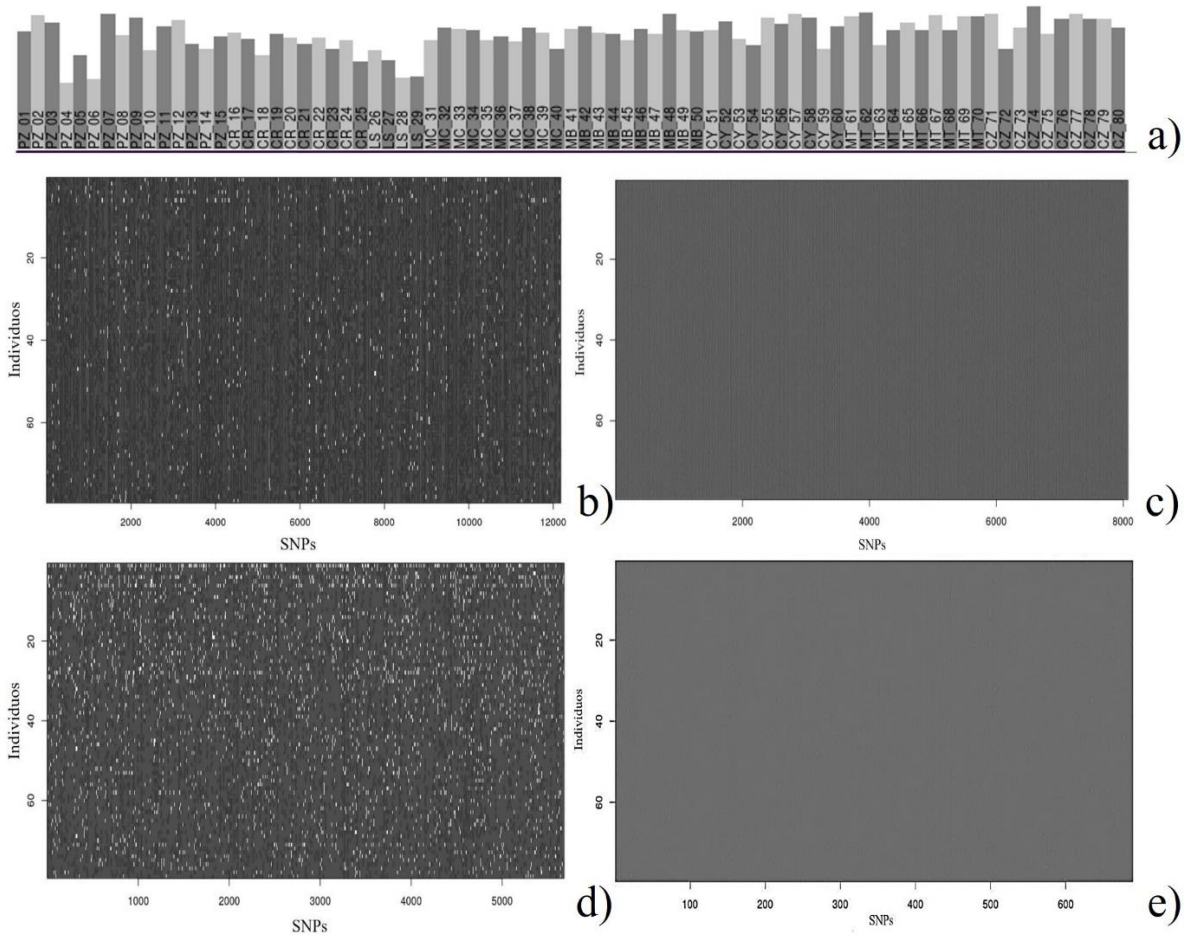
Anexo II

En el siguiente cuadro se muestran los parámetros utilizados y el número de SNPs obtenidos. Se seleccionaron las matrices de SNPs CGUM 1, CGUM 2, Stacks 1 y Stacks 2.

		maf	max-missing	SNPs
CGUM		0.008	0.9	12,352
		0.008	1	8,186
	CGUM 1	0.01	0.9	12,352
	CGUM 2	0.01	1	8,186
Stacks		0.008	0.9	5,682
		0.008	1	689
	Stacks 1	0.01	0.9	5,682
	Stacks 2	0.01	1	689

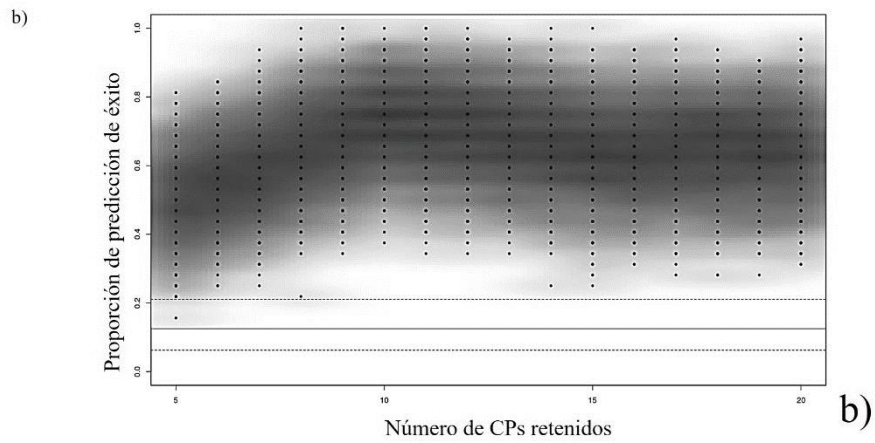
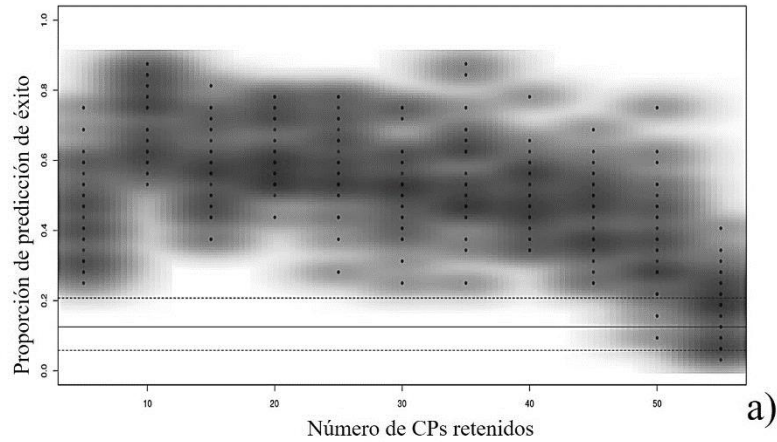
Anexo III

Visualización de los datos: a) cantidad de secuencias por individuo, en promedio se obtuvieron 700 mil lecturas por individuo, con un promedio en puntuaciones de calidad $\geq Q30$. Conjuntos de SNPs, con y sin datos faltantes: b) CGUM 1; c) CGUM 2; d) Stacks 1; e) Stacks 2.



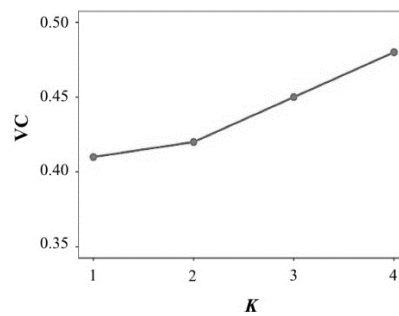
Anexo IV

Diagrama de dispersión de la validación cruzada para el número óptimo de CPs retenidos para el ADCP, con respecto a la proporción de la predicción de éxito: a) análisis hasta 55 CPs; b) de 5 a 20 CPs. El número óptimo de CPs fue de 10.



Anexo V

Validación por cruzamiento (VC) obtenido con ADMIXTURE. Valores de agrupamiento (K) de 1 a 4.



Anexo VI

Coefficientes de correlación entre las matrices, obtenidos por medio de la prueba parcial de Mantel. Gen=matriz de flujo génico, Geo=matriz de distancias geográficas, Amb=matriz de distancias euclidianas entre las variables ambientales (CONABIO con 10 variables y WorldClim con 8).

SITIOS	CONABIO		WorldClim 8v		45_50		85_50		45_70		85_70	
	Mantel	P-valor	Mantel	P-valor	Mantel	P-valor	Mantel	P-valor	Mantel	P-valor	Mantel	P-valor
Gen~Geo	0.7604	0.001	0.7604	0.002	0.7604	0.002	0.7604	0.0019	0.7604	0.0018	0.7604	0.0019
Gen~Amb	0.033	0.367	0.0382	0.383	0.1878	0.121	0.125	0.1931	0.125	0.1891	-0.0408	0.4993
Geo~Amb	0.1353	0.203	0.0565	0.335	0.268	0.086	0.1703	0.1421	0.1703	0.1382	-0.1353	0.2409