UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE INVESTIGACIONES EN ECOSISTEMAS Y SUSTENTABILIDAD

EL COMPLEJO *QUERCUS LAETA* (FAGACEAE): DIVERSIDAD GENÉTICA Y DELIMITACIÓN DE ESPECIES MEDIANTE ANÁLISIS INTEGRATIVOS

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

SADDAN MORALES SALDAÑA

TUTOR PRINCIPAL DE TESIS: DR. ANTONIO GONZÁLEZ RODRÍGUEZ
INSTITUTO DE INVESTIGACIONES EN ECOSISTEMAS Y SUSTENTABILIDAD, UNAM
COMITÉ TUTOR: DR. EFRAÍN TOVAR SÁNCHEZ
INSTITUTO DE ECOLOGÍA, UNAM. CIByC, UAEM
COMITÉ TUTOR: DRA. SUSANA VALENCIA ÁVALOS
FACULTAD DE CIENCIAS, UNAM

MORELIA, MICHOACÁN, ABRIL. 2023

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE INVESTIGACIONES EN ECOSISTEMAS Y SUSTENTABILIDAD

EL COMPLEJO *QUERCUS LAETA* (FAGACEAE): DIVERSIDAD GENÉTICA Y DELIMITACIÓN DE ESPECIES MEDIANTE ANÁLISIS INTEGRATIVOS

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

SADDAN MORALES SALDAÑA

TUTOR PRINCIPAL DE TESIS: DR. ANTONIO GONZÁLEZ RODRÍGUEZ
INSTITUTO DE INVESTIGACIONES EN ECOSISTEMAS Y SUSTENTABILIDAD, UNAM
COMITÉ TUTOR: DR. EFRAÍN TOVAR SÁNCHEZ
INSTITUTO DE ECOLOGÍA, UNAM. CIByC, UAEM
COMITÉ TUTOR: DRA. SUSANA VALENCIA ÁVALOS
FACULTAD DE CIENCIAS, UNAM

MORELIA, MICHOACÁN, ABRIL. 2023

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS

ENTIDAD IIES-M

OFICIO CPCB/212/2023

ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
P r e s e n t e

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día **28 de noviembre de 2022,** se aprobó el siguiente jurado para el examen de grado de **DOCTOR EN CIENCIAS,** del estudiante **MORALES SALDAÑA SADDAN,** con número de cuenta **306290170** con la tesis titulada **""EL COMPLEJO *Quercus laeta* (FAGACEAE): DIVERSIDAD GENÉTICA Y DELIMITACIÓN DE ESPECIES MEDIANTE ANÁLISIS INTEGRATIVOS",** realizada bajo la dirección del **DR. ANTONIO GONZÁLEZ RODRÍGUEZ**, quedando integrado de la siguiente manera:

Presidente:	DR. ALBERTO KEN OYAMA NAKAGAWA
Vocal:	DRA. SILVANA MARTÉN RODRÍGUEZ
Vocal:	DR. DAVID SEBASTIAN GERNANDT
Vocal:	DRA. CAROLINA GRANADOS MENDOZA
Secretario:	DRA. SUSANA VALENCIA ÁVALOS

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

A T E N T A M E N T E
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Ciudad Universitaria, Cd. Mx., a 22 de febrero de 2023

COORDINADOR DEL PROGRAMA

DR. ADOLFO GERARDO NAVARRO SIGÜENZA

POSGRADO EN CIENCIAS BIOLOGICAS
COORDINACIÓN

Agradecimientos Institucionales

Al programa de Doctorado en Ciencias Biológicas de la Universidad Nacional Autónoma de México por el apoyo proporcionado durante la realización de mis estudios.

Al consejo Nacional de Ciencias y tecnología por la beca de Doctorado otorgada para realizar mis estudios de posgrado (CONACYT_BECA/CVU- 83720/624511)

Al Programa de Apoyo a Proyectos de Investigación e Innovación tecnológica PAPIIT DGAPA-UNAM por el financiamiento del proyecto 210020

Al programa de Apoyo a los estudios de Posgrado (PAEP-UNAM) por los apoyos recibidos.

Al Morton Arboretum por la beca otorgada para asistir a una estancia de investigación y al Pritzker Lab perteneciente al Field Museum que en conjunto brindaron sus instalaciones, laboratorios y equipos para desarrollar parte de esta tesis.

Al Laboratorio Nacional de Análisis y Síntesis Ecológica (LANASE-UNAM) por los recursos bioinformáticos brindados.

Al Dr. Antonio González Rodríguez, director de tesis y tutor principal de mis estudios de posgrado de Doctorado. A la Dra. Susana Valencia Ávalos y al Dr. Efraín Tovar Sánchez, integrantes de mi comité tutor, por su asesoría y consejos durante mi formación doctoral.

Agradecimientos a título personal

A mis compañeros de laboratorio de Genética de la Conservación del Instituto de Investigaciones en Ecosistemas y Sustentabilidad por su apoyo en el trabajo de campo ycolaboración durante estos años.

A mi jurado de tesis, Dra. Carolina Granados Mendoza, Dra. Silvana Martén Rodríguez yDres. David Sebastian Gernandt, Ken Oyama Nakagawa por sus comentarios y observaciones para enriquecer este trabajo.

Al Dr. Andrew L. Hipp por su activa colaboración en el proyecto y su ayuda brindada en elmarco teórico y metodológico. A Mira Garner, Marlen Hahn por la ayuda en la logística y el trabajo de laboratorio en mi estancia.

Al Dr. Antonio González Rodríguez por recibirme en su laboratorio, las oportunidades brindadas y por el apoyo recibido durante los últimos meses del proceso de graduación.

A las personas no mencionadas aquí, pero que he conocido y me han brindado su ayuda a lo largo de estos años. Sin duda alguna he aprendido de ustedes.

A esos amigos encontrados por casualidad pero que acaban siendo hermanos

A mis padres y mi familia por su apoyo incondicional…

Contenido

RESUMEN

La hibridación introgresiva de múltiples especies plantea un desafío para la comprensión taxonómica y filogenética de taxones con un gran número de especies intercruzables concurrentes. Los robles, género *Quercus*, ejemplifican esta situación. Son muy diversos en simpatría y se cruzan libremente, creando singameones de especies interfértiles. Además, son notoriamente variables en morfología y tienen grandes tamaños efectivos poblacionales y alta heterocigosidad. Como consecuencia, sus complicaciones taxonómicas y filogenéticas se derivan no solo de la introgresión, sino también de la plasticidad y el sorteo incompleto de linajes. Aunque se dispone de una filogenia fechada y bien resuelta para los robles americanos, la cual ha permitido delimitar los principales clados, quedan por definir las relaciones evolutivas dentro de muchos de esos clados, en particular para el relativamente joven y excepcionalmente diverso clado de los robles blancos mexicanos (subsección *Leucomexicana*). En este sentido, la asignación taxonómica y la inferencia filogenética de *Quercus laeta*, especie endémica de México y ampliamente distribuida en diferentes tipos de hábitat, ha sido motivo de controversia debido a que entre poblaciones presenta un polimorfismo morfológico excepcional que ha sido difícil de clasificar. Estas poblaciones constituyen un complejo en el que se han propuesto siete descripciones, aunque seis de ellas han sido sinonimizadas bajo *Q. laeta*: *Quercus bipedalis*, *Quercus centralis*, *Quercus obscura*, *Quercus pallescens*, *Quercus prinopsis* y *Quercus transmontana*.

En el capítulo I se caracterizó el grado y patrón de diferenciación morfológica y genética entre diferentes morfotipos en *Quercus laeta*. El objetivo fue evaluar si algunos de estos morfotipos pueden considerarse entidades específicas distintas o son más bien parte de un continuo de variación. Se analizaron nueve loci de microsatélites y dos regiones intergénicas de ADN de cloroplasto. Las diferencias morfológicas se evaluaron mediante morfometría geométrica. El ADN del cloroplasto mostró una baja diferenciación, lo que sugiere introgresión o intercambio de haplotipos ancestrales entre los morfotipos de *Q. laeta*. Los microsatélites nucleares indicaron diferenciación entre dos grupos genéticos principales, los cuales fueron congruentes con la diferenciación morfométrica. En conclusión, los marcadores nucleares y las variaciones morfométricas sugieren la existencia de al menos dos entidades diferentes dentro de *Q. laeta*.

En el capítulo II, se utilizaron secuencias de 155 genes nucleares de bajo número de copias para identificar potenciales linajes dentro del complejo *Q. laeta*. Asimismo, mediante análisis tanto de concatenación como coalescentes multiespecies, se evaluó la ubicación filogenética de estos linajes en relación con otras especies del clado *Leucomexicana*. Además, se utilizaron métodos de redes filogenéticas para evaluar la importancia genómica de la introgresión reciente o histórica entre linajes. Finalmente, se estimó la relevancia de la diferenciación ambiental en la diversificación de estos linajes de encinos. Los resultados obtenidos sustentan tres grandes conclusiones. Primero, nuestros datos apoyan la existencia de por lo menos cuatro especies en el complejo: *Q. centralis*, *Q. laeta*, *Q. prinopsis* y *Q. transmontana*. En segundo lugar, nuestro estudio demuestra que el complejo *Q. laeta* es polifilético. Finalmente, la existencia de una diferenciación de nicho ecológico significativa entre los linajes del complejo *Q. laeta* sugiere que la diferenciación ecológica ha sido parte de la divergencia evolutiva entre estos robles blancos, ya sea impulsando la especiación o como resultado de ella.

En el capítulo III, se utilizó microscopía electrónica de barrido para realizar un análisis cualitativo y cuantitativo tanto de tricomas como de ceras foliares para un subconjunto de 48 individuos pertenecientes a *Q. centralis*, *Q. laeta*, *Q. prinopsis* y *Q. transmontana*, con el objetivo de determinar el valor de tales rasgos micromorfológicos en la taxonomía del grupo. El tipo y la densidad de los tricomas, así como el número y la longitud de los brazos de los tricomas permitieron detectar diferencias significativas entre *Q. centralis*, *Q. laeta*, *Q. prinopsis* y *Q. transmontana*. Estos resultados proporcionaron información útil para identificar y discriminar correctamente entre diferentes especies presentes dentro del complejo *Quercus laeta.*

Considerando la evidencia obtenida a partir de múltiples disciplinas proponemos que el complejo *Quercus laeta* está compuesto al menos por cuatro diferentes especies: *Quercus centralis*, *Quercus prinopsis*, *Quercus laeta* y *Quercus transmontana*, cada una con sus propias características micromorfológicas, afinidades climáticas y tendencias evolutivas, aunque se requiere un estudio más detallado para aclarar el estatus de los linajes geográficamente estructurados detectados dentro de *Q. centralis* y *Q. laeta*. Aunque las designaciones de nombres de especies putativas

presentadas en este trabajo se basan en los nombres propuestos por Trelease, será necesario una investigación nomenclatural detallada para las especies aquí propuestas antes de que puedan ser redescritos formalmente. Complementariamente, este enfoque permitió aumentar el rigor en la delimitación de especies dentro del género *Quercus*, remarcando que el uso de marcadores de cloroplasto podría no ser de los más indicados al momento de establecer límites dentro de complejos de especies de encinos. En contraste, el uso de marcadores de microsatélites en conjunto con una evaluación morfométrica podría servir como un primer paso en el descubrimiento de especies de encinos a pesar de su propensión a la hibridación; sin embargo, es importante señalar que la evaluación de cualquier hipótesis de especie debe realizarse bajo un marco filogenético. En este contexto, el uso del enriquecimiento dirigido permitió detectar genes ortólogos útiles para la estimación multilocus, que en conjunto con evidencia de diferenciación de nicho climático soportan la hipótesis sobre la existencia de múltiples especies dentro del complejo *Quercus laeta*.

ABSTRACT

Multispecies introgressive hybridization poses a challenge to taxonomic and phylogenetic understanding of taxa with high numbers of co-occurring, inter-crossable species. The oaks, genus *Quercus*, exemplify this situation. Oaks are highly diverse in sympatry and cross freely, creating syngameons of interfertile species. Moreover, they are notoriously variable in morphology and have large effective population sizes and high heterozygosity. Their taxonomic and phylogenetic complications consequently derive not only from introgression, but also from plasticity and incomplete lineage sorting. Although a well-resolved, dated phylogeny is available for monophyletic American oaks, which has allowed to delimit major clades, evolutionary relationships within many of those clades remain to be defined, particularly for the relatively young and exceptionally diverse Mexican white oak clade (subsection *Leucomexicana*). In this sense, taxonomic assignment, and phylogenetic inference of *Quercus laeta*, an endemic species to Mexico and widely distributed in different habitat types, has been a matter of controversy because it shows exceptional morphological polymorphism among populations that has been challenging to sort. These populations constitute a complex from which seven descriptions have been proposed, although six of them have been synonymized under *Q. laeta*: *Quercus bipedalis*, *Quercus centralis*, *Quercus obscura*, *Quercus pallescens*, *Quercus prinopsis*, and *Quercus transmontana*.

In chapter I, the degree and patterns of morphological and genetic differentiation among different morphotypes in the white oak *Quercus laeta* were characterized. The aim was to evaluate if some of these can be considered as distinct specific entities or are rather part of a continuum of variation. Nine microsatellite loci and two intergenic regions of chloroplast DNA were analyzed. Morphological differences were evaluated using geometric morphometrics. Chloroplast DNA showed low differentiation, suggesting introgression, or sharing of ancestral haplotypes among the *Q. laeta* morphotypes. Nuclear microsatellites indicated differentiation into two distinct main genetic groups, which were congruent with morphometric differentiation. In conclusion, nuclear markers and morphological variations suggest the existence of at least two different entities within *Q. laeta*.

In chapter II, sequences of 155 low-copy nuclear genes were used to identify potential lineages within the *Quercus laeta* complex. Likewise, by both concatenation and multispecies coalescent approaches, the phylogenetic placement of these lineages relative to other oak species in the Mexican white oak clade was evaluated. Also, phylogenetic network methods were used to evaluate the genomic significance of recent or historical introgression among lineages. Finally, using climatic niche comparisons, the importance of environmental differentiation in the diversification of these oak lineages was estimated. The results obtained support three major conclusions. First, our data supports the existence of at least four species in the complex: *Q. centralis*, *Q. laeta*, *Q. prinopsis* and *Q. transmontana*. Second, our study demonstrates that the *Q. laeta* complex is polyphyletic. Finally, the existence of significant ecological niche differentiation among the *Q. laeta* complex lineages suggests that ecological differentiation has been a part of evolutionary divergence among these white oaks, either driving speciation or resulting from it.

In chapter III, scanning electron microscopy was used to perform a qualitative and quantitative analysis of trichomes and foliar waxes for a subset of 48 individuals belonging to *Q. centralis*, *Q. laeta*, *Q. prinopsis* and *Q. transmontana*, with the goal of determining the value of such micromorphological traits in the taxonomy of the group. Type and density of trichomes, and trichome arm number and length significantly differentiated among *Q. centralis*, *Q. laeta*, *Q. prinopsis* and *Q. transmontana*. Our results provided useful information to correctly identify and discriminate among different species present within the *Q. laeta* complex.

Considering the evidence obtained from multiple disciplines, we propose that the *Quercus laeta* complex is actually composed at least of four different species: *Quercus centralis*, *Quercus prinopsis*, *Quercus laeta*, and *Quercus transmontana*, each with their own micromorphological characteristics, climatic affinities, and evolutionary trends, although a more detailed study is required to clarify the status of the geographically structured lineages detected within *Q. centralis* and *Q. laeta*. Although the putative species name designations presented in this study are based on the names proposed by Trelease, a detailed nomenclatural investigation will be necessary for species proposed here before they can be formally

redescribed. Complementarily, this approach allowed increasing the rigor in the delimitation of species within the *Quercus* genus, noting that the use of chloroplast markers might not be the most appropriate when establishing limits within complexes of oak species. In contrast, the use of nuclear microsatellites in conjunction with morphometric evaluation could serve as a first step in oak species discovery despite their propensity for hybridization. However, it is important to note that evaluation of any species hypothesis must be carried out under a phylogenetic framework. In this context, the use of directed enrichment allowed the detection of orthologous genes useful for multilocus estimation, which together with evidence of climatic niche differentiation support the hypothesis of the existence of multiple species within the *Quercus laeta* complex.

I. INTRODUCCIÓN GENERAL

1.1 El problema de las especies

El problema de definir que es una especie y delimitar a las especies de una manera objetiva, consistente y biológicamente significativa para todos los seres vivos es lo que se conoce como el problema de las especies (Hey 2001; Zachos, 2016; 2018). Es innegable la importancia que siempre se les ha atribuido a las especies, refiriéndose a estas como las unidades fundamentales en la biología (Mayr, 1982; De Queiroz, 2005; Hausdorf, 2011). Sin embargo, esta relevancia viene acompañada de una disparidad en la conceptualización que las especies pueden adquirir bajo diferentes contextos, lo cual a menudo causa una confusión innecesaria. Algunos ejemplos de la constante polémica que ha surgido sobre la idealización y la operatividad del concepto de especie incluyen debates clásicos sobre si las especies existen o no (realismo *versus* nominalismo), así como discusiones contemporáneas sobre sí deberíamos considerar a las especies como categoría o como taxon, o si las especies taxonómicas (T species) son sinónimo de las especies evolutivas (E species), inclusive si debemos considerar a las especies como entidades diacrónicas o sincrónicas (Mayden, 1997; Hey, 2001; Lee, 2003; Hey, 2006; De Queiroz, 2007; Hart, 2010; Hausdorf, 2011; Zachos, 2016; 2018).

Si a las discusiones anteriores añadimos el hecho de que se han propuesto más de 20 conceptos diferentes de especie (Mayden, 1997; Hausdorf, 2011; Zachos 2018; Hong 2020) y que muchos de los conceptos y sus definiciones asociadas son al menos parcialmente incompatibles entre sí, el obtener inferencias disímiles sobre los límites entre taxones y en consecuencia sobre la cantidad de especies reconocidas son resultados habituales en el campo de la delimitación de especies. Dicha incompatibilidad entre los distintos conceptos de especie tiene que ver con las diferentes propiedades biológicas sobre las cuales se basan (De Queiroz, 1998; 2007). En este contexto, Mayden (1997) y De Queiroz (1998, 2007) han propuesto que la clave para conciliar y estandarizar los diferentes conceptos de especie es identificar un núcleo común, de modo que las hipótesis durante el proceso de delimitación podrían ser construidas bajo un concepto primario y unificado de especie, el cual carece de propiedades definidas. De tal modo que uno de los más importantes consensos sobre el

concepto de especie es que pueden ser consideradas como segmentos de linajes de poblaciones o meta poblaciones que evolucionan de manera conjunta (De Queiroz, 1998; 2007). En consecuencia, si estos criterios son considerados como el elemento común para ser la propiedad primaria de la especie, las propiedades que definen a los conceptos alternativos deberán ser reinterpretadas como propiedades contingentes, es decir, propiedades que las especies pueden adquirir o no durante el curso de su existencia (De Queiroz, 2007). De modo que la diversidad de los diferentes conceptos de especie puede explicarse postulando que las propiedades contingentes de las especies serían consideradas como propiedades definitorias secundarias, es decir, evidencia utilizada para reconocer a los diferentes linajes (De Queiroz, 2007; Wiens, 2007), las cuales surgen en diferentes momentos durante el proceso de especiación.

1.2 Especies y Especiación

En el tiempo de la síntesis evolutiva moderna, se consideraba que la especiación requería un aislamiento reproductivo completo de los taxones nacientes, a menudo a través del aislamiento geográfico. Por lo tanto, entender cómo evolucionan estas barreras reproductivas era clave en la comprensión de la especiación (Dobzhansky, 1937; Mayr, 1982). Bajo este modelo, la ausencia de flujo génico es considerada un requisito fundamental en la divergencia de linajes, ya que se considera que el intercambio genético puede revertir la divergencia (Wang et al., 2020). En este contexto, los genomas de las especies serían interpretados como unidades coadaptadas que están separadas de otras unidades similares por barreras reproductivas, lo que implica que la divergencia entre especies únicamente ocurre a través del aislamiento del genoma completo y por lo tanto las especies que se forman por hibridación no podrían ser consideradas verdaderas especies (Wu, 2001; De la Torre et al., 2014; Wang et al., 2020).

Ahora, si consideramos que lo fundamental de la especiación es la divergencia de linajes, independientemente de las propiedades contingentes, y que además la especiación es un proceso continuo a través del tiempo, la idea de especiación en presencia de flujo génico, que en primera instancia parecería paradójica, tendría bastante sentido ya que los agentes separadores que conducen a la divergencia superan la cohesión promovida a través del flujo

génico (Zachos, 2016; Wang et al., 2020). Bajo este modelo, las especies están definidas por un conjunto de loci que gobiernan los caracteres morfológicos, reproductivos, etológicos y/o ecológicos, de modo que el gen es la unidad básica en la diferenciación de las especies (Wu, 2001; Wu y Ting, 2004; Wang et al., 2020). Bajo esta visión, los límites entre las especies son semi-permeables, por lo que algunas regiones genómicas comparten genes introgresados entre especies, mientras que otras regiones acumulan divergencias en respuesta a la selección natural, también llamadas "islas genómicas", que son, en teoría, más divergentes que el resto del genoma (Wu y Ting, 2004, Osada y Wu, 2005; Cruickshank y Hanh, 2014).

El modelo de “islas genómicas de especiación” ha promovido la visión génica de la especiación (Turner et al., 2005; Martin et al., 2013; Renaut et al., 2013; Brandvain et al., 2014; Duranton et al., 2018). En este contexto, Wu (2001) fue uno de los primeros en declarar que la especiación refleja un proceso de distinción genealógica emergente, en lugar de una discontinuidad que afecta a todos los genes simultáneamente, lo cual sugiere que los eventos de hibridación introgresiva entre poblaciones sustancialmente divergentes serían bastante comunes (Sankararaman et al., 2014; Coyner et al., 2015; Morii et al., 2015; Schumer et al., 2018). Además, aparentemente, las especies incipientes pueden acumular rápidamente una divergencia sustancial incluso en presencia de flujo genético (De la Torre et al., 2014). Durante la hibridación, en especies que muestran adaptaciones a diferentes ambientes, la selección divergente actúa sobre un subconjunto de genes, contrarrestando la homogenización ocasionada por el efecto del flujo génico y previniendo la introgresión en regiones genómicas circundantes (Chapman et al., 2013). Como resultado, los límites entre especies se mantienen a pesar de la hibridación e introgresión (Andrew y Rieseberg, 2013).

1.3 Enfoques multi evidencia para delimitar especies

De Queiroz (2005) distingue tres componentes relacionados al problema de las especies: (1) la definición correcta de especies (¿qué es una especie?); (2) ¿cuáles son los procesos responsables de la existencia de especies? y, (3) ¿cómo deben delimitarse las especies? Bajo la premisa de que los dos primeros problemas son conceptuales, mientras que el tercero es metodológico, un enfoque que se ha propuesto como posible solución teórica al problema de las especies se basa en una visión jerárquica, la cual distingue entre un concepto ontológico

de especie y un concepto de especie operativo (Zachos, 2016). En este contexto, la identificación y delimitación de taxones de especies debería llevarse a cabo con rigor empírico utilizando métodos robustos y altamente replicables que permitan la identificación de distintos linajes evolutivos (Dayrat, 2005; Padial et al., 2010; Schlick- Steiner et al., 2010).

Históricamente los enfoques morfológicos han sobresalido en el campo de la delimitación de especies. Sin embargo, las investigaciones deberían ser conducidas con particular atención a la historia de vida, distribución geográfica, morfología, genética y comportamiento (cuando es aplicable) del sistema focal, conduciendo así el campo de la delimitación de especies a un cruce interesante entre distintas disciplinas y fuentes de evidencia donde diversas metodologías y enfoques filosóficos convergen (Rannala y Yang, 2003; Dayrat, 2005; De Queiroz, 2007; Carstens et al., 2013). Si bien esta convergencia entre diferentes metodologías puede reflejar un importante dinamismo en el campo de la delimitación de especies, también puede impedir la estabilización de una taxonomía alfa, la cual puede diferir significativamente dependiendo del criterio de aplicación para delimitar especies (Padial et al., 2010; Schlick- Steiner et al., 2010). Una taxonomía alfa inestable podría tener inmensas ramificaciones prácticas, como dividir y agrupar continuamente taxones basados en metodologías no replicables. Otra de las grandes dificultades que enfrentan la delimitación de especies y la biología evolutiva es el desacuerdo sobre en qué momento a lo largo del continuo de la especiación estos linajes que evolucionan de manera separada pueden considerarse como diferentes especies (Padial et al., 2010).

Schlick- Steiner et al. (2010) plantean dos grandes enfoques iniciales a partir de los cuales la delimitación de especies se puede desarrollar: i) enfoques de descubrimiento y ii) enfoques basados en hipótesis. Bajo los enfoques de descubrimiento los ejemplares son analizados sin hipótesis previas acerca de las especies presentes, de tal manera que las hipótesis de especies se derivan de los datos analizados. Mientras que, en el enfoque basado en hipótesis, el conjunto de ejemplares se utiliza para probar hipótesis previas sobre la existencia de especies, como cuando una delimitación de especies hipotetizada a partir de datos morfológicos se prueba usando genética de poblaciones.

Complementariamente, Padial et al. (2010) integran dos grandes marcos conceptuales bajo los cuales delimitar especies: i) la integración por congruencia y ii) la integración por acumulación. La mayor ventaja del enfoque por congruencia es que promueve la estabilidad taxonómica soportando la validez de una especie mediante varios conjuntos de caracteres siempre y cuando no estén vinculados, mientras que la mayor limitación al exigir congruencia entre caracteres taxonómicos es el riesgo de subestimar el número de especies debido a que los procesos de especiación no siempre están acompañados por cambios en los caracteres a todos los niveles (Padial et al., 2010), además de que las tasas relativas de cambio de caracteres durante la diversificación de linajes son heterogéneas. Varios estudios han soportado la idea de que la carencia de congruencia entre caracteres es una situación frecuente resultado de diferentes modos y circunstancias de especiación (Padial et al., 2010). Un riesgo adicional de aplicar la integración por congruencia podría ocasionar un sesgo hacia descubrir especies viejas, es decir, especies que divergieron en un pasado distante por lo cual incrementan la probabilidad de mostrar un sorteo completo de linajes, monofilia reciproca para muchos loci, presentar diferencias en los caracteres debido a nuevas mutaciones y la fijación a través de procesos adaptativos o neutrales (Roe y Sperling, 2007; Shaffer y Thomson, 2007; Degnan y Rosenberg, 2009).

En contraparte, el enfoque de integración por acumulación está basado en el supuesto de que la divergencia en cualquiera de los atributos del organismo que constituyen caracteres taxonómicos puede proporcionar evidencia para la existencia de una especie, por lo que la congruencia es deseada pero no considerada necesaria (Padial et al., 2010). Debido a que todos los caracteres taxonómicos son contingentes en existencia, orden de aparición y magnitud de divergencia durante el continuo de la especiación, bajo este enfoque, las evidencias de todos los conjuntos de caracteres se ensamblan acumulativamente, las concordancias y discordancias son explicadas a partir de la perspectiva evolutiva de las poblaciones bajo estudio y la decisión es hecha basada en la información disponible, la cual puede conducir al reconocimiento de una especie sobre la base de un solo conjunto de caracteres si estos caracteres se consideran buenos indicadores de divergencia de linajes (Padial et al., 2010). Una ventaja importante de este enfoque es que no vincula la delimitación de especies a la identificación de ninguna propiedad biológica particular. Por lo que los

taxónomos pueden seleccionar y concentrase sobre el conjunto de caracteres taxonómicos más apropiados para cada grupo de organismos. La acumulación probablemente es el enfoque más adecuado para descubrir especies recientemente divergentes en radiaciones adaptativas debido al proceso escalonado de especiación en gradientes ecológicos (Shaffer y Thomson, 2007). Sin embargo, la principal limitante de la integración por acumulación es que el uso acrítico de una sola línea de evidencia puede conducir a la sobreestimación en el número de especies (Padial et al., 2010).

1.4 El género *Quercus* como modelo en estudios de especiación

Un sistema modelo, en general, es aquel que tiene características que lo hacen particularmente bueno para el avance de la ciencia en cierta dirección (Cavender-Bares, 2019). Dentro de esta definición, algunos autores han resaltado la diferencia entre organismos modelo y clados modelo. El primer término hace referencia a aquellos organismos que se han estudiado ampliamente debido a que son fáciles de mantener y reproducir en un entorno de laboratorio o que tienen ventajas para examinar aspectos particulares de forma experimental. En contraste, los clados modelo, son definidos como linajes de organismos que permiten avanzar en el entendimiento de conceptos particulares (Knapp et al., 2004; Buell, 2009; O'Grady y DeSalle, 2018; Conner et al., 2021). Bajo esta premisa, y desde una perspectiva evolutiva, el género *Quercus* resulta ser un interesante clado modelo para el estudio de ciertos procesos como la especiación, debido a que permiten entender el papel de fenómenos como la introgresión, la hibridación y la dinámica evolutiva del flujo genético en poblaciones de organismos de vida larga, todo esto con relación a los límites entre especies y sus procesos adaptativos (Curtu et al., 2007; Pennisi, 2016; Yang et al., 2016; Cavender-Bares, 2019; Crowl et al., 2020).

Durante las últimas décadas diversos estudios han abordado el papel de la hibridación e introgresión sobre procesos de adaptación y especiación dentro del género *Quercus* (González-Rodríguez et al., 2004a; Scotti-Saintagne et al., 2004; Mir et al., 2006; Curtu et al., 2007; Peñaloza-Ramírez et al., 2010; Gailing y Curtu, 2014; Ortego et al., 2014; Cavender-Bares et al., 2015; Eaton et al., 2015; Goicoechea et al., 2015; Backs y Ashley, 2016; Hauser et al., 2017; Yang et al., 2016; Hipp et al., 2018; Kim et al., 2018; Crowl et al.,

2020; Hipp et al., 2020), encontrando reiteradamente que los encinos mantienen la coherencia de las especies, a pesar de la introgresión. Recientemente, las herramientas genómicas han facilitado la capacidad de detectar introgresión antigua proporcionando un nuevo contexto para comprender como la introgresión pudo haber contribuido a las relaciones evolutivas entre las especies (Eaton et al., 2015; Hauser et al., 2017; McVay et al., 2017a; McVay et al., 2017b; Kim et al., 2018; Crowl et al., 2020). Además, el desarrollo de herramientas estadísticas ha permitido identificar y cuantificar el flujo de genes entre linajes de encinos evolutivamente distantes, demostrando que la introgresión ancestral ha influido en los genomas de especies que han entrado en contacto secundario. Sin embargo, también se sugiere que estos procesos no son omnipresentes en todas las especies quedando aspectos aún sin conocer (McVay et al., 2017a; McVay et al., 2017b; Kim et al., 2018).

1.5 Diversidad y sistemática del género *Quercus*

Dentro de la familia Fagaceae, el género *Quercus* presenta la mayor distribución geográfica a nivel mundial, abarcando regiones templadas y tropicales del hemisferio norte, por lo que constituye uno de los géneros más dominantes de plantas leñosas en esta región, tanto en biomasa como en riqueza de especies (Nixon, 1993a; Valencia-Á., 2004; Nixon, 2006; Hipp et al., 2018, Cavender-Bares, 2006; Cavender-Bares, 2019). Las estimaciones para la riqueza de especies a nivel mundial son heterogéneas; estimaciones moderadas han propuesto entre 300 y 400 especies (Nixon et al., 1997), mientras que los menos conservadores estiman entre 500 y 600 especies en todo el mundo (Soepadmo, 1972 en Jones 1986; Govaerts y Frondin, 1998). Recientemente se ha propuesto que existen más de 435 especies a nivel mundial (Denk et al., 2017, Manos y Hipp, 2021).

El sistema de clasificación más reciente para el género *Quercus* (Fagaceae) reconoce dos subgéneros (*Cerris* con alrededor de 140 spp., y *Quercus* con estimaciones de 295 spp.), los cuales albergan ocho secciones (Denk et al., 2017; Manos y Hipp, 2021). Se ha estimado que la división entre los subgéneros de *Quercus* data del Eoceno temprano (56 – 47.8 m.a.) y ha sido reconocida desde los primeros trabajos moleculares del género (Manos et al., 1999; Oh y Manos, 2008) hasta estudios recientes apoyados por datos RAD-seq (Hipp et al., 2020; Manos y Hipp, 2021). Además, los dos subgéneros representan una profunda división

biogeográfica entre los taxones modernos: el subgénero *Cerris* está restringido a Europa, Asia, y el norte de África, mientras que el subgénero *Quercus* se limita en gran medida al continente americano con excepción de dos dispersiones de regreso a Eurasia (Denk et al., 2017; Manos y Hipp, 2021). Particularmente, el subgénero *Quercus* presenta cinco secciones: *Virentes* (ca. 7 spp.), *Lobatae* (ca. 120 spp.), *Protobalanus* (ca. 5 spp.), las cuales son exclusivas al continente americano, y *Ponticae* (ca. 2 spp.) y *Quercus* (ca. 150 spp.) que se distribuyen tanto en el viejo como en el nuevo mundo (Denk et al., 2017; Manos y Hipp, 2021).

Similar a lo que ocurre con la riqueza de especies a nivel mundial, la diversidad exacta de encinos americanos se desconoce, pero se calcula que comprende el 65 % de las especies estimadas para el género en todo el mundo (Nixon, 1997; Manos et al., 1999; Valencia-Á., 2004; Nixon, 2006; De Beaulieu y Lamant, 2010). Nixon (2006) estima aproximadamente 220 especies de encinos en todo el continente, distribuidas desde el sur de Canadá hasta el norte de los andes colombianos, incluyendo Cuba, abarcando ambientes como los humedales en la parte alta del medio Oeste de Estados Unidos, las zonas áridas de México y Estados Unidos y hasta selvas tropicales en el sur de México y Centroamérica. Sin embargo, la mayor riqueza de especies de encinos se localiza en México, aunque la riqueza específica de encinos mexicanos permanece poco entendida. Las cifras más conservadoras estiman entre 125 y 150 especies (Nixon, 1993b, Chávez, 1998; Govaers y Frondin, 1998) mientras que los menos conservadores calculan entre 170 y 250 especies de encinos mexicanos (Camus, 1936-1954; Trelease, 1924; González, 1993). Valencia-Á. (2004) proporciona una de las cifras más actualizadas con alrededor de 161 especies válidas para México (109 endémicas), de las cuáles cuatro corresponden a la sección *Protobalanus*, tres a *Virentes*, 76 a la sección *Lobatae* y 78 a la sección *Quercus* (Valencia-Á., 2004; Cavender-Bares et al., 2015), siendo esta sección la que presenta una mayor distribución geográfica debido a que pueden tolerar condiciones más áridas y secas, siendo particularmente más diversos en áreas xéricas en comparación con la sección *Lobatae*, cuyos miembros son más diversos en ambientes húmedos (González, 1993; Nixon, 1993b; Chávez, 1998; Valencia-Á., 2004).

1.6 Diversificación del género *Quercus* en América

El clado de los encinos americanos (*Quercus* subg. *Quercus*) comprende tres linajes endémicos: *Lobatae, Protobalanus* y *Virentes*, y dos linajes transcontinentales, las secciones *Ponticae* y *Quercus* (Denk et al., 2017; Manos y Hipp, 2021). Análisis filogenómicos calibrados en el tiempo en combinación con fósiles, apoyan una división inicial en el clado de los encinos americanos entre la sect. *Lobatae*, y el resto del clado entre ca. 54 y 48 m. a. (Hipp et al., 2018; Manos y Hipp, 2021). Además, se estima que hace ca. 33 m. a. ya habían surgido las secciones *Protobalanus*, *Ponticae*, *Virentes* y *Quercus* (Hipp et al., 2018).

Complementariamente, registros polínicos del Eoceno medio provenientes de la isla Axel Heiberg indican la presencia de linajes de robles modernos en latitudes altas de América del Norte hace aproximadamente 45 m. a. (McIntyre, 1991). En ambos casos, a medida que las temperaturas en las latitudes altas disminuyeron en 3-5 °C durante la transición climática Eoceno-Oligoceno (34 m. a.) (Liu et al., 2009), los encinos se dispersaron hacia el sur ocupando sus distribuciones actuales, ocurriendo un patrón de vicarianza paralela entre encinos rojos y encinos blancos, con cada sección divergiendo en un clado occidental ubicado en la provincia florística de California (CA-FP) hermano de un clado oriental de América del Norte (Hipp et al., 2018; Manos y Hipp, 2021). Sin embargo, existe una importante distinción ecológica entre estos dos grandes clados de roble: *Lobatae* muestra niveles de diversificación más bajos en comparación a la sección *Quercus* en el oeste de América del Norte, especialmente en la CA-FP y en los bosques xerófilos del suroeste americano (Manos y Hipp, 2021). En conjunto, estos estudios sugieren que los encinos americanos tienen un origen templado en el norte y solo más tarde colonizaron México y Centroamérica (Hipp et al., 2018; Hipp et al., 2020; Manos y Hipp, 2021).

Las reconstrucciones realizadas por Hipp et al. (2018) sugieren que los tiempos de vicarianza entre el clado del este de Norteamérica y el clado mexicano son anteriores a la incursión del desierto en Texas y en el noreste de México (aproximadamente hace 10 m. a.), que se asocia con la vicarianza del este de México con este de América del Norte en muchos clados de angiospermas (Manos y Meireles, 2015). Desde la llegada de los encinos a México (hace aproximadamente 14.5-20.8 m. a.), estos se han diversificado en aproximadamente 161

especies (Valencia-Á., 2014; Hipp et al., 2018) desarrollándose desde en chaparrales xéricos hasta en bosques mésicos y desde en climas templados hasta subtropicales, con una fuerte diversificación en las zonas montañosas de México (Nixon, 2006). Bajo este contexto, la correlación entre las tasas de especiación y diversificación ecológica sugiere que la alta diversidad de los encinos en América se vio moldeada en gran medida por la convergencia simpátrica en nichos edáficos y climáticos entre encinos rojos y blancos. Sin embargo, en México, estos procesos se intensificaron como resultado de una mayor oportunidad ecológica y/o una mayor alopatría en las áreas montañosas ecológicamente heterogéneas de México (Hipp et al., 2018).

1.7 Problemas en la delimitación de especies dentro del género *Quercus*

A pesar de que ciertos aspectos sobre la diversificación y biogeografía del clado de encinos americanos han sido dilucidados (Hipp et al., 2018; Hipp et al., 2020; Manos y Hipp, 2021), temas como establecer relaciones filogenéticas a niveles profundos entre ciertos complejos de especies taxonómicamente complicados siguen pendientes, lo que refleja el reto que sigue siendo definir límites entre taxones dentro del género *Quercus*. Desde la publicación de las tres grandes monografías de este género realizadas por Trelease (1924), Müller (1942) y Camus (1936-1954), se han identificado diversos problemas taxonómicos asociados a la identidad específica, abarcando varios complejos taxonómicos como los estudiados por Nixon y Muller (1993), Spellenberg y Bacon (1996), Howard et al. (1997), Spellenberg (1998), Romero (2001), González-Rodríguez et al. (2004b), Vázquez y Nixon (2013), Valencia-A. et al. (2015, 2016), Sabás-Rosales et al. (2017), Nixon y Barrie (2017), McCauley et al. (2019) y McCauley y Oyama (2020).

Tradicionalmente, la delimitación de especies en el género se ha desarrollado mediante rasgos macro y micromorfológicos (González-Villarreal, 1986; Bello y Labat, 1987; Valencia, 1995; González-Villarreal, 2003a,2003b; Valencia y Lozada, 2003; y Scareli-Santos et al., 2013) respaldados por un concepto taxonómico de especie, que en esencia es la revisión de claves dicotómicas e identificación de morfología tanto en especímenes frescos como de herbario (Valencia-A., 2020). No obstante, debido a que las estructuras florales no son consideradas como caracteres de importancia en la diagnosis de

las especies, la determinación de ejemplares se realiza en gran medida con base en rasgos morfológicos de las hojas. Sin embargo, los encinos pueden tener una considerable variación en este tipo de caracteres, como resultado de plasticidad fenotípica, convergencia morfológica, variación genética intraespecífica y eventos de hibridación e introgresión (Hardin, 1975; Rushton, 1993; González-Rodríguez et al., 2004b; Tovar-Sánchez y Oyama, 2004; Valencia-Cuevas et al., 2015; Kusi y Karsai, 2020; Maya-García et al., 2020; Valencia-Á., 2020), lo que en conjunto con la escasez e inaccesibilidad del material tipo y de la literatura original (Valencia-Á., 2004) aumenta la problemática en el campo de la delimitación de especies dentro del género *Quercus*.

Actualmente, existe una tendencia a utilizar diversas evidencias, como la morfometría (Gailing et al., 2012; Ishida et al., 2003), filogeografía (McCauley et al., 2019), filogenómica (Hipp et al., 2018; 2020), genética de poblaciones (Albarrán-Lara et al., 2019) y nicho ecológico (Albarrán-Lara et al., 2019) para establecer límites entre especies de encinos de manera objetiva, revelando que muchas especies nominales en realidad están compuestas por complejos de especies, que morfológicamente pueden ser muy similares, pero genéticamente y/o ecológicamente divergentes. Sin embargo, particularmente para los encinos mexicanos, estos enfoques han sido poco explorados (Albarrán-Lara et al., 2019; McCauley et al., 2019), por lo que la delimitación de especies dentro del género *Quercus* sigue siendo un desafío en muchos complejos como *Acutifoliae*, *Microphyllae*, *Crassifoliae* y *Rugosae*, por mencionar algunos (Valencia-Á., 2020).

1.8 El complejo *Quercus laeta* como modelo para evaluar patrones de divergencia a través del continuo de especiación.

La asignación taxonómica de ejemplares a *Quercus laeta* ha sido motivo de controversia debido al alto grado de polimorfismo que presenta esta especie a lo largo de su distribución geográfica. *Quercus laeta* es endémica a México, pero ampliamente distribuida; se ha reportado en los estados de Aguascalientes, Coahuila, Ciudad de México, Durango, Guanajuato, Hidalgo, Jalisco, México, Michoacán, Nayarit, Nuevo león, San Luis Potosí, Sinaloa, Oaxaca y Zacatecas (Valencia-Á., 2004). Estas poblaciones constituyen un complejo para el cual se han propuesto siete nombres (Liebmann, 1854; Trelease, 1924; Valencia-Á.,

2004) aunque seis de ellos han sido posteriormente sinonimizados bajo *Quercus laeta*: *Quercus bipedalis* Trel., *Q. centralis* Trel., *Q. obscura* Trel., *Q. pallescens* Trel., *Q. prinopsis* Trel. y *Q. transmontana* Trel. (McVaugh, 1974; Rangel et al., 2002). La revisión de ejemplares de herbario ha llevado a descartar ciertos ejemplares determinados previamente como *Quercus laeta* que no cumplen con los caracteres diagnósticos descritos inicialmente para la especie, como son la presencia de papilas en el envés de la hoja, envés eglandular y tricomas sésiles ligeramente contortos (Liebmnann, 1854). Sin embargo, aun excluyendo los ejemplares incorrectamente determinados persiste una gran variación morfológica dentro de *Quercus laeta* lo que podría suponer la existencia de un complejo de especies dentro de esta entidad taxonómica. Recientemente, trabajos filogenéticos han encontrado que *Quercus laeta* no es un grupo monofilético, aportando aún más evidencia de que *Quercus laeta* en realidad es un complejo de especies (Hipp et al., 2018; Hipp et al., 2020). Sin embargo, el muestreo de diferentes poblaciones de *Quercus laeta* para estos trabajos ha sido escaso debido a que no se han contemplado toda la distribución geográfica que presenta este taxon.

Considerando lo anterior, y como punto de partida para este trabajo, se decidió categorizar la gran variación morfológica presente en las poblaciones de *Quercus laeta* bajo hipótesis de especies para contrastarlas utilizando diferentes líneas de evidencia. Con base en trabajo de herbario y en la revisión de holotipos disponibles, se decidió excluir las sinonimias *Q. obscura*, *Q. pallescens* y *Q. bipedalis*. Lo anterior, considerando que el holotipo de *Q. obscura* no presenta diferencias morfológicas visibles que lo distinga de *Q. transmontana*, por lo que se consideró tomar ambas sinonimias como uno solo morfotipo. Además, en el caso de *Q. pallescens* y *Q. bipedalis* se optó por excluirlos debido a que no se identificaron ejemplares de herbario que se pudieran asignar a estas sinonimias, y para aquellos que presentaban cierta afinidad no fue posible localizar los lugares de colecta debido a la antigüedad del registro. Los ejemplares que cumplían con los caracteres diagnósticos de la especie fueron agrupados en morfotipos con base en las sinonimias. De tal manera, se caracterizaron cinco morfotipos como unidades base para el estudio, rescatando las sinonimias de *Q. centralis*, *Q. prinopsis*, y *Q. transmontana*, esta última se decidió dividirla en dos morfotipos con base en diferencias de la pubescencia en las ramillas, la forma del ápice de la lámina foliar y su distribución geográfica (Figura 1; Tabla 1).

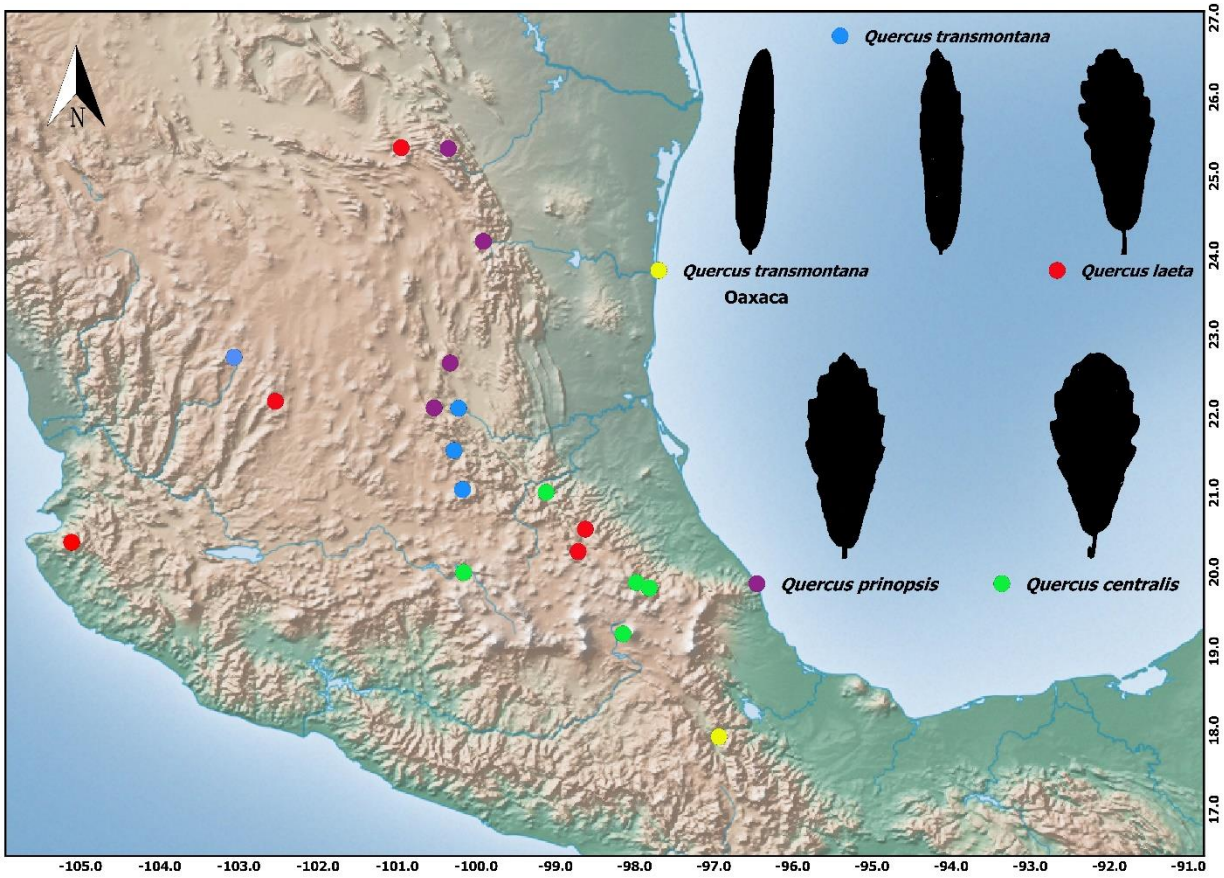


Figura 1. Distribución geográfica de los morfotipos propuestos para *Quercus laeta*.

Tabla 1. Diferencias morfológicas identificadas entre los morfotipos del complejo *Q. laeta*.

Carácter	*Q. transmontana*	*Q. transmontana*-coyula	*Q. laeta*	*Q. prinopsis*	*Q. centralis*
Peciolo	Glabro de color amarillo rojizo	Rojizo glabrescentes	Rojizo Pubescente	Densamente tomentoso de color amarillo dorado	Pubescente
Tamaño del peciolo (mm)	(2.7) 3.4-4.8 (6.3)	(2.3) 3.3 – 4.8 (5.3)	5.5 – 7.9 (11.5)	6 - 9.5 (10.93)	(4.9) 6.3 – 9.4 (12)
Envés de la hoja	Pubescente sin tricomas glandulares	Pubescente sin tricomas glandulares	Pubescente con tricomas glandulares ausentes o muy escasos.	Densamente tomentoso de color amarillo dorado	Pubescente, en ocasiones con tricomas glandulares
Forma de la hoja	Lanceolada a oblonga	Lanceolada	Oblanceolada, oblonga	Elíptica a obovada oblanceolada	Obovada a ovada
Largo de la lámina (cm)	(4.9) 6.3 - 8.7 (9.8)	(6) 6.3 – 8.4 (9.4)	(5.2) 7.3 – 9.2 (10.5)	(5.4) 7.3 – 8.2 (9.3)	(6.2) 7.4 - 10
Ancho de la lámina (cm)	1.4-2.1 (2.8)	(0.9) 1.4 – 1.7 (2.4)	(1.7) 2.4 – 3 (3.7)	2.3- 3.5	(2.5) 2.9 - 3.9 (5)
Ramillas	Glabras con lenticelas conspicuas	Pubescentes con lenticelas conspicuas	Pubescentes	Tomentosas de color amarillo dorado	Pubescentes, color café con lenticelas conspicuas
Forma del ápice	Redondeado	Agudo, en ocasiones obtuso	Agudo	Obtuso	Obtuso
Forma de la base	Oblicua a redondeada	Redondeada a veces oblicua	Redondeada	Redondeada a oblicua	Redondeada a oblicua
Forma de las yemas	Cónicas	Redondas con estípulas lineares	Redondas	Cónicas	Cónicas
Tamaño de las yemas (mm)	(1.4) 1.98- 2.83 (3.3)	1.4 – 1.83	1.8 – 2.6 (3.4)	2.4-3.7 (4.3)	(1.87) 2.2 – 2.7 (3.2)
Tipo de vegetación	Bosque de encino seco, Bosque de pino-encino	Ecotono bosque de encino con matorral xerófilo	Bosque de encino, Ecotono entre bosque de encino y matorral xerófilo	Bosque de encino seco	Bosque de pino-encino
Distribución geográfica	Durango, Guanajuato, San Luis Potosí, Zacatecas	Oaxaca	Aguascalientes, Chihuahua, Hidalgo, Jalisco, Nayarit y Nuevo León	Nuevo León, San Luis Potosí	Michoacán, Puebla y Tlaxcala; Ciudad de México

La aplicación de múltiples líneas de evidencia al estudio de poblaciones polimórficas es una estrategia útil para evaluar patrones de divergencia a través del continuo de especiación y, en el contexto de este trabajo, para probar la hipótesis de que los diferentes morfotipos identificados para *Quercus laeta* representan linajes evolutivamente independientes. En este sentido el complejo *Quercus laeta* es un modelo interesante para

comprender procesos de especiación incluso en presencia de flujo génico, como es probablemente común en el género *Quercus*. Este trabajo constituye un esfuerzo inicial para establecer límites y esclarecer las relaciones evolutivas en un grupo de encinos que históricamente ha sido difícil de abordar. Bajo este marco, el presente estudio se divide en tres capítulos:

El capítulo I (*Even more oak species in Mexico? Genetic structure and morphological differentiation support the presence of at least two specific entities within* Quercus laeta) tuvo como objetivo evaluar si algunos de estos morfotipos pueden considerarse entidades específicas distintas o son más bien parte de un continuo de variación.

A partir de las distintas entidades reconocidas en el capítulo anterior, en el capítulo II (*Phylogenomics and ecological niche contrasts lead to the identification of multiple and independent evolutionary lineages within the* Quercus laeta *complex (Fagaceae): new insights into the diversification of the Leucomexicana clade*) se emplea un enfoque que combina la reconstrucción de árboles de linajes, especies y la estimación de redes filogenéticas con el análisis de diferenciación ambiental para contrastar las hipótesis de especies genético-morfométrico, así mismo se proponen nuevas hipótesis filogenéticas entre los linajes detectados en el complejo *Q. laeta* y otras especies de encinos del clado *Leucomexicana*.

Una vez identificados diferentes linajes dentro del complejo, en el capítulo IV (*Micromorphological characterization as a taxonomical tool for delimitation of oak species: A case study of the* Quercus laeta *complex*) se profundiza en el estudio de rasgos micromorfológicos con el objetivo de i) Evaluar cualitativa y cuantitativamente la variación de tricomas entre los linajes genómicos detectados en *Quercus laeta* mediante microscopía electrónica de barrido, ii) probar la aplicabilidad de caracteres micromorfológicos (tipos y densidad de tricomas, variación de longitud de los brazos de tricomas, número de radios por tricomas y tipos de ceras) para el reconocimiento de especies en complejos de encinos taxonómicamente complicados y iii) determinar la combinación más útil de caracteres micromorfológicos para la identificación de los linajes presentes en *Quercus laeta*.

OBJETIVO GENERAL

Contribuir a la comprensión de la especiación en el género *Quercus* analizando bajo un marco evolutivo los patrones de diferenciación a distintos niveles entre poblaciones polimórficas del complejo *Q. laeta* a lo largo de su distribución geográfica.

OBJETIVOS PARTICULARES

- Analizar los patrones de diferenciación genética y morfométrica entre poblaciones polimórficas del complejo *Q. laeta* a lo largo de su distribución geográfica

- Definir las relaciones entre poblaciones del complejo *Q. laeta* para identificar hipotéticos linajes independientes y así evaluar la ubicación filogenética de estos linajes en relación con otras especies de encinos en el clado *Leucomexicana.*

- Determinar cualitativa y cuantitativamente la combinación más útil de caracteres micromorfológicos para el reconocimiento de los linajes identificados mediante datos genómicos

II. CAPÍTULO I

Even more oak species in Mexico? Genetic structure and morphological differentiation support the presence of atleast two specific entities within *Quercus laeta*

Publicado en Journal of Systematics and Evolution

JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12818

Research Article

Even more oak species in Mexico? Genetic structure and morphological differentiation support the presence of at least two specific entities within *Quercus laeta*

Saddan Morales-Saldaña[1], Susana Valencia-Ávalos[2], Ken Oyama[3], Efraín Tovar-Sánchez[4], Andrew L. Hipp[5,6], and Antonio González-Rodríguez[1]*

[1]Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Morelia 58190, México
[2]Herbario de la Facultad de Ciencias, Departamento de Biología Comparada, Universidad Nacional Autónoma de México, Circuito Exterior, s.n, Ciudad Universitaria, Coyoacán, México City 04510, México
[3]Escuela Nacional de Estudios Superiores Unidad Morelia, Universidad Nacional Autónoma de México (UNAM), Antigua Carretera a Pátzcuaro No. 8701, Col. Ex-Hacienda de San José de la Huerta, Morelia, Michoacán 58190, México
[4]Centro de Investigación en Biodiversidad y Conservación, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001, Col. Chamilpa, C. P., Cuernavaca 62209, México
[5]The Morton Arboretum, Lisle, IL 60532-1293, USA
[6]The Field Museum, Chicago, IL 60605, USA
*Author for correspondence. Email: agrodrig@cieco.unam.mx
Received 31 May 2021; Accepted 24 November 2021; Article first published online 29 November 2021

Abstract Differentiation among populations, sometimes despite ongoing gene exchange, is a key step in speciation. Therefore, comparison of intra- and interspecific differentiation patterns is of great significance to understanding speciation. The genus *Quercus* is an interesting system to test speciation models in the presence of gene flow, due to its weak interspecific reproductive barriers. The aim of the present study was to characterize the degree and pattern of morphological and genetic differentiation among different morphotypes in the white oak *Quercus laeta*, some corresponding to the previously described species *Quercus centralis*, *Q. laeta*, *Quercus prinopsis*, and *Quercus transmontana*, as well as geographically structured variation within *Q. transmontana* not described previously. Our goal was to evaluate if some of these can be considered distinct specific entities or are rather part of a continuum of variation. Nine microsatellite loci and two intergenic regions of chloroplast DNA were analyzed. Morphological differences were evaluated using geometric morphometrics. Chloroplast DNA showed low differentiation, suggesting introgression or sharing of ancestral haplotypes among the *Q. laeta* morphotypes. Nuclear microsatellites indicated differentiation into two distinct main genetic groups, which were congruent with morphological differentiation. In conclusion, nuclear markers and morphological variations suggest the existence of at least two different entities within *Q. laeta*.

Key words: gene flow, genetic differentiation, leaf morphometrics, microsatellite loci, morphotype, population assignment, species delimitation.

1 Introduction

Identifying the processes that promote differentiation among populations despite gene exchange is crucial to understanding the evolution of biodiversity because gene flow across the tree of life challenges our understanding of what species are and how they arise (Nosil et al., 2009; Abbott et al., 2013; Mallet et al., 2016; Winger, 2017; Pinheiro et al., 2018; Albarrán-Lara et al., 2019, McCauley et al., 2019). One such challenge involves pinpointing the genetic barriers that initiate speciation and understanding the temporal succession of barriers that ultimately lead to the completion of speciation. In this context, it can be problematic to distinguish species that have low levels of genetic divergence, either because speciation is recent or because they continue to exchange genes (Muir & Schlötterer, 2005; Lexer et al., 2006; Petit & Excoffier, 2009; Pinho & Hey, 2010).

Species complexes offer an opportunity to study micro- and macroevolutionary divergence processes and the early steps in speciation (Pinheiro et al., 2018; Villaverde et al., 2018). Nevertheless, within such complexes, it is particularly difficult to identify clear boundaries between putative species under two scenarios: (i) when species show small morphological differences due to non-adaptative radiation, morphological stasis or phenotypic convergence (Bickford et al., 2007; Nosil, 2012; Barley et al., 2013;

Valencia-A et al., 2016); and (ii) when species show considerable morphological variation but small genetic differentiation as a result of a fast and recent divergence, constant gene flow, and ecological niche changes (Schluter, 2001; Nosil, 2012; Albarrán-Lara et al., 2019).

Species complexes guide us to the nexus of taxonomy and speciation. One cannot readily study one without also drawing inferences from and regarding the other. Inference of genetic clusters, the detection of recent admixture events, and the evaluation of divergence among populations are hallmarks of studies of speciation (Mallet, 1995; Guichoux et al., 2011; Sousa & Hey, 2013; Albarrán-Lara et al., 2019; McCauley et al., 2019; Miller et al., 2020; Zumwalde et al., 2021). As such, many methods, ranging from Nei's population genetic distances (Nei, 1972), F-Statistics (Weir & Cockerham, 1984), analysis of molecular variance (Excoffier et al., 1992) and discriminant analysis of principal components (DAPC) (Jombart et al., 2010), to admixture and Bayesian clustering analyses (Pritchard et al., 2000; Alexander et al., 2009), have been developed to identify genetic breaks among populations and potentially help to disentangle species boundaries (Pritchard et al., 2000; Hubisz et al., 2009; Jombart et al., 2010; Ewédjè et al., 2020; Miller et al., 2020). Among these methods, there is a conceptual division between methods that assess *a priori* (predefined) populations (e.g., F_{ST}, analysis of molecular variance [AMOVA]) versus approaches that identify genetic clusters *de novo* (Miller et al., 2020). In the first case, *a priori* assignment could help to visualize differentiation between hypothesized groups (e.g., geographical regions, morphotypes), whereas in the *de novo* approach, the population structure is inferred from a given data set, to discover genetic groups (Miller et al., 2020).

In parallel to genetic methods, morphological characterization of populations and species has advanced through new methodologies such as geometric morphometrics (Adams et al., 2013; Habel et al., 2015; Widener & Frant, 2018; Desmond et al., 2021). Leaf morphology is particularly relevant for species delimitation in plant groups that show little variation in floral features (Elias, 1980; Valencia-A, 2004). For this purpose, the geometric morphometrics approach may provide statistically robust delimitations among groups of individuals and to identify patterns of hybrid morphology (Viscosi & Cardini, 2011; Viscosi & Fortini, 2011; Viscosi, 2015; Liu et al., 2018). For many oak species (genus *Quercus* L.), high morphological variability has been documented at various levels, for example, among leaves on a single tree, among trees within populations, and among populations within species, which often complicates interpretation of taxonomic patterns (Hardin, 1975; Valencia-A, 2004, 2021). Understanding these sources of variance is therefore important to evaluate how different adaptive and neutral processes influence leaf morphology both among and within species (Bruschi et al., 2003; González-Rodríguez & Oyama, 2005; Viscosi, 2015; Liu et al., 2018; Desmond et al., 2021).

Morphological variation within oak species has been of key interest to botanists even before Darwin, who considered the exceptional variation in, for example, *Quercus robur* to be a consequence of recent divergence (Darwin, 1859, ch. 2). In the century following Darwin, it became increasingly clear that introgressive hybridization is also an important player in oak morphological variation (Engelmann, 1876; MacDougal, 1907; Palmer, 1948). In the decades following the modern synthesis, the weak inherent interspecific reproductive barriers in oaks became a difficulty for our conception of oak species, reaching a head in the mid-1970s (Burger, 1975; Hardin, 1975; Van Valen, 1976). Since then, a much broader understanding of divergence with gene flow across the tree of life has helped resolve this problem (e.g., Mallet et al., 2016), allowing us to turn our attention to how introgressive hybridization works hand in hand with local adaptation, phenotypic plasticity, and adaptation to broad climatic gradients to shape the great foliar polymorphism in both European and American oaks (Tsukaia, 2005; Sork et al., 2016; Ramírez-Valiente et al., 2017; Sáenz-Romero et al., 2017; Martins et al., 2018; Cavender-Bares, 2019).

Quercus laeta Liebm. is endemic to Mexico and widely distributed in different habitat types. Across its geographic distribution, the populations of this taxon show morphological variation that has been challenging to sort, mainly in the degree of pubescence in the underside of the leaves, the shape of the leaf lamina, and the pubescence of the twigs. In an oak flora that is already notoriously challenging, *Q. laeta* shows exceptional morphological polymorphism among populations (Valencia-A, 2004). These populations constitute a complex from which seven descriptions have been proposed (Trelease, 1924; Liebmann, 1854; Valencia-A, 2004), although six of them have been synonymized under *Q. laeta*: *Quercus bipedalis* Trel., *Quercus centralis* Trel., *Quercus obscura* Trel., *Quercus pallescens* Trel., *Quercus prinopsis* Trel., and *Quercus transmontana* Trel. (McVaugh, 1974; Rangel et al., 2002). Even though the variation in *Q. laeta* populations in different mountain chains of Mexico is quite complex, it follows certain geographical patterns, which could be interpreted as the result of geographic diversification related to the ecological heterogeneity of these montane areas (Hipp et al., 2018, 2020).

From this perspective, the application of multiple lines of evidence for the study of polymorphic populations is a useful strategy to evaluate divergence patterns across the speciation continuum and, in the context of this work, to test the hypothesis about the existence of separate species within *Q. laeta* (Corl et al., 2010; Pinheiro et al., 2018). As explained, the *Q. laeta* complex is an interesting model to understand diversification even in the presence of gene flow, as is probably common in *Quercus*, if not typical (Whittemore & Schaal, 1991; McVay et al., 2017; Kremer & Hipp, 2020).

In this study, we contribute to the understanding of speciation in the genus *Quercus* and the origin of Mexican species diversity by analyzing the genetic and morphometric differentiation patterns among polymorphic populations of the *Q. laeta* complex throughout its geographical distribution. We addressed the following questions: (i) are there recognizable genetic boundaries among the previously identified morphotypes or do they correspond to a genetic continuum? (ii) Is there correspondence between genetic clusters and foliar morphometric differentiation? and (iii) based on the combined evidence, is it possible to propose distinct specific entities within the *Q. laeta* complex?

2 Material and Methods

2.1 Herbarium revision

Because the taxonomic assignment of specimens to *Quercus laeta* has been a matter of controversy, we first reviewed the herbarium collections from the Faculty of Sciences of the National Autonomous University of Mexico (FCME) and the National Herbarium of Mexico (MEXU) to discard specimens erroneously assigned to *Q. laeta*, on the basis of the diagnostic characters originally described for the species (Liebmann, 1854) and for the synonyms described by Trelease (1924) as well as the characteristics observed on type specimens available at JSTOR (http://plants.jstor.org). These characteristics mainly include the presence of papillae and absence of glands on the leaf undersides and sessile, slightly contorted trichomes on the abaxial surface. Once the erroneously determined specimens were discarded, a morphological examination and assignment of the specimens to different postulated species (hereafter morphotypes) were performed, using as a reference the names initially proposed by Trelease (1924). However, during this herbarium work, it was not possible to assign specimens and collection locations to the *bipedalis* and *pallescens* morphotypes, so these were excluded from the study. Likewise, we decided to divide the *transmontana* morphotype into two different morphotypes because the specimens from the Coyula locality in Oaxaca state showed a disjunct distribution and consistent morphological differences from the populations belonging to the *transmontana* morphotype from the center and north of Mexico. In particular, the "*transmontana coyula*" morphotype shows twig pubescence, glabrescent reddish petioles, and acute leaf apex, in contrast to the glabrous twigs, glabrous yellow-reddish petioles, and rounded leaf apex present in the *transmontana* morphotype. Thus, five morphotypes (*centralis*, *laeta*, *prinopsis*, *transmontana*, and *transmontana coyula*) were defined as base units for this study (Table S1).

2.2 Population samples

A total of 167 adult individuals of the *Q. laeta* complex were sampled from 19 sites (hereafter, "populations" refers to samples from these locations as geographically defined, unless otherwise specified). At each population, 5–11 randomly selected trees were sampled, separated from each other by a minimum of 30–50 m. Populations almost covered the complete geographical distribution of *Q. laeta* and the five morphotypes (Table S2; Fig. 1). At each site, the trees were identified based on the above-mentioned diagnostic morphological characteristics. Collected branches were pressed and dried to obtain three herbarium specimens. From each tree, we additionally sampled mature leaves with no apparent damage from different branches; these leaves were stored in plastic bags and placed on ice until final storage at −80 °C in the laboratory for genetic analysis. Likewise, between six and eight mature leaves were pressed per individual for the corresponding morphometric analyses. Vouchers of specimens were deposited in the FCME herbarium.

2.3 DNA extraction, chloroplast DNA sequencing, and nuclear SSR genotyping

DNA isolation was performed using a cetyltrimethyl ammonium bromide (CTAB) protocol with an additional phenol–chloroform cleaning step (Lefort & Douglas, 1999). The isolated DNA was diluted with deionized water to a final concentration of 20 ng/μl and stored at −20 °C. We amplified two chloroplast (cpDNA) intergenic spacers: *trnC-ycf6* and *ycf6-psbM* (Shaw et al., 2005). Polymerase chain reactions

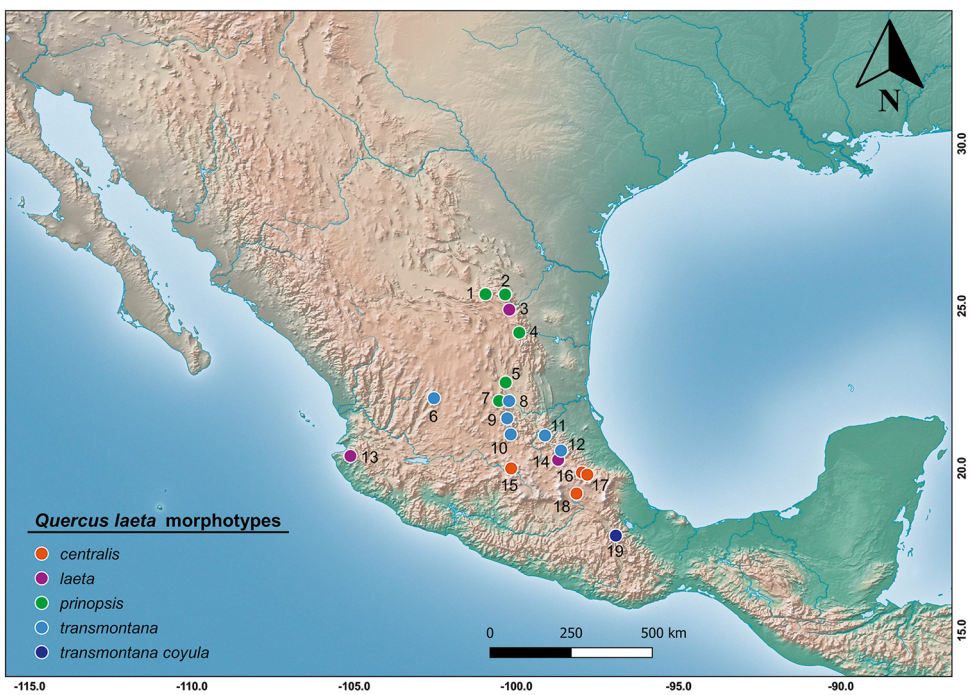


Fig. 1. Map of collected populations of the *Quercus laeta* complex in Mexico.

(PCR) were carried out using the Taq-Platinum master mix (Qiagen, CA, USA) as follows: 1 µl of each forward and reverse primer (0.2 µM), 10 µl of master mix, 5 µl of H_2O, and 3 µl of genomic DNA in a total volume of 20 µl. PCR amplifications began with 5 min denaturation at 94 °C, followed by 35 cycles of 1 min denaturation at 94 °C, 1.5 min annealing at 49 °C (*trnC-ycf6*) or 65 °C (*ycf6-psbM*), and 1.5 min extension at 72 °C. A final extension step of 10 min at 72 °C was included. Unpurified PCR products were sent to Macrogen Company (Rockville, MD, USA) for purification and sequencing.

To estimate the population genetic diversity and structure, nine nuclear microsatellite (nSSR) loci were amplified in three multiplex PCRs and two individual reactions. The first multiplex reaction included the QpZAG96 and QpZAG110 (Steinkellner et al., 1997) loci; the second included the QpZAG36 (Steinkellner et al., 1997) and QrZAG39 (Kampfer et al., 1998) loci; and the third included the quru-GA-IF02, quru-GA-OC11, and quru-GA-OMO5 (Aldrich et al., 2002) loci. Finally, the quru-GA-OC11 and quru-GA-2F05 (Aldrich et al., 2002) loci were amplified separately. Reactions were carried out with 3 µl of Taq-Platinum master mix (Qiagen CA, USA), 0.3 µl of each forward and reverse primer (0.2 µM), 1.4 µl of H_2O, and 1 µl of genomic DNA (10 ng/µl) in a total volume of 6 µl. The reaction began with 3 min denaturation at 94 °C, followed by 40 cycles of 30 s denaturation at 94 °C, 45 s annealing at 54 °C for quru-GA-OC11 and the first and third multiplex reactions, 48 °C for the second multiplex reaction, and 45 °C for quru-GA-OM05. A final extension step of 10 min at 72 °C was included. One microliter of each PCR product was combined with 9 µl of Hi-Di Formamide and 0.3 µl of GeneScan-600 LIZ (Applied Biosystems, CA, USA) and run on an ABI-PRISM 3100-Avant sequencer (Applied Biosystems, CA, USA) to obtain the size of the microsatellite fragments. Electropherograms were analyzed using GENEMARKER software v.1.91 (Softgenetics LLC, State College, PA, USA). The individual genotype assignments of the nine nSSR were verified at least three times to corroborate our genotyping.

2.4 Nuclear genetic diversity and structure

The presence and frequency of null alleles for each locus and population were determined using FreeNA (Chapuis & Estoup, 2007). Deviations from the Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each population were tested in GENEPOP v.3.4 (Raymond & Rousset, 1995) with 10 000 Markov Chain iterations. For each population, we calculated descriptive genetic diversity statistics as the mean number of alleles (N_a), the mean number of effective alleles (N_e), the mean expected heterozygosity (H_E), the mean observed heterozygosity (H_O), and the mean inbreeding coefficient (*F*) using GENALEX v.6.5 (Peakall & Smouse, 2012). Rarefied allelic richness was estimated in FSTAT (Goudet, 2001) to account for differences in sample size among populations. As some of the loci showed the presence of null alleles, we also calculated the inbreeding coefficient using the Bayesian procedure implemented in INEST v.2.0 (Chybicki & Burczyk, 2009).

To identify potential genetic breaks among populations that would be suggestive of the presence of distinct specific entities, we analyzed genetic differentiation and structure using methods that rely both on *a priori* defined populations and approaches that infer genetic clusters *de novo* from individual data. First, to visualize the relationships among populations, we constructed an unrooted NJ tree from nSSR data in POPULATIONS ver. 1.2.31 (Langella, 1999) with Nei's standard genetic distance (Nei, 1987). Support for genetic structure was estimated using bootstrapping with resampling across loci 1000 times. Second, to identify genetic clusters *de novo*, we used multivariate DAPC using the function *dapc* of the *adegenet* package (Jombart, 2008; Jombart et al., 2010; R Core Team, 2013). This method does not rely on a particular population genetics model and is thus free of assumptions about Hardy–Weinberg or linkage equilibrium. In addition, it generates a graphical representation of the multivariate distance among the inferred clusters (Jombart et al., 2010). In the analysis, we identified the optimal number of clusters using the find.clusters() function. We tested *K* values from 1 to 10, and the optimal number of clusters was selected based on the Bayesian information criterion (BIC) (Jombart et al., 2010). The multivariate distance among clusters was visualized using a minimum spanning tree. Cross validation was applied to avoid overestimation of the number of components to be used.

Additionally, we applied the Bayesian clustering algorithm implemented in the software STRUCTURE 2.3.4 (Pritchard et al., 2000), performing a hierarchical analysis (Pritchard & Wen, 2003; Vähä et al., 2007; Janes et al., 2017). First, analyses were performed on the complete data set running values of *K* from 1 to 6 with 10 independent runs for each *K*. The burn-in length was 100 000 steps, followed by 300 000 MCMC iterations. Following Pritchard et al. (2000) and Hubisz et al. (2009), we performed these initial runs under different models: (i) using morphotype grouping as a prior (LOCPRIOR option), (ii) using geographical populations as a prior, and (iii) not using the LOCPRIOR option. In the three cases, data analysis was performed under an admixture model and correlated allele frequencies. Second, we performed independent STRUCTURE runs on each genetic cluster identified in the previous step to detect further patterns of genetic structure. For these STRUCTURE runs, we assumed *K* from 1 to 6, also using the ancestry model with admixture and correlated allele frequencies. In all cases, the most likely *K*-value was selected by compiling runs with STRUCTURE HARVESTER v.0692 (Earl & vonHoldt, 2012) and using the *ad hoc* ΔK method (Evanno et al., 2005). The program CLUMPAK (Kopelman et al., 2015) was used to summarize the 10 runs for the best number of clusters found.

Finally, to quantify genetic variance components among the different inferred groups, we used AMOVA in Arlequin ver. 3.5.2.2 (Excoffier & Lischer, 2010). Therefore, we tested the following four different groupings: (i) populations grouped according to the morphotypes; (ii) populations grouped according to the results of unrooted neighbor-joining (NJ) using the software POPULATIONS ver. 1.2.31 (Langella, 1999); (iii) genetic clusters obtained by DAPC; and (iv) populations grouped according to the nuclear genetic groups obtained by STRUCTURE 2.3.4 (Pritchard et al., 2000). The significance of the genetic variation partitions was tested with 10 000 permutations. Additionally, we computed

pairwise fixation statistics (F_{ST}, R_{ST}) among morphotypes using the software SPAGeDi (Hardy & Vekemans, 2002). R_{ST} and F_{ST} have the same theoretical expectations when population genetic structuring results solely from genetic drift, while if stepwise mutations contribute to genetic differentiation (mutation rate non-negligible compared to the migration rate or the reciprocal of the number of generations since population isolation), we expect $R_{ST} > F_{ST}$ (Hardy et al., 2003). The latter situation implies that the mean allele sizes vary between populations for at least some loci, leading to a phylogeographic signal. The latter was tested using the allele size permutation test (Hardy et al., 2003) implemented in SPAGeDi.

2.5 cpDNA diversity and structure

Electropherograms were analyzed using BioEdit (Hall, 1999), and the obtained sequences were manually aligned using MEGA 6 (Tamura et al., 2007). Molecular diversity indices, such as the number of polymorphic sites (*S*), haplotype diversity (*Hd*), and nucleotide diversity (π) per population, were calculated using DnaSP v 6.0 software (Rozas et al., 2017). Arlequin 3.5.2.2 (Excoffier & Lischer, 2010) was used to perform four analyses of molecular variance according to the four groupings previously explained for nuclear genetic variation (see above). The significance of the genetic variation partitions was tested with 10 000 permutations.

The genealogical relationships among haplotypes were assessed using a median-joining algorithm implemented in PopART v 1.7 (Leigh & Bryant, 2015). Finally, the relative frequencies of haplotypes per population were mapped using QGis 3.8.3 (QGIS Development Team, 2019).

2.6 Morphological analysis of leaf shape

We quantified morphological differentiation in the *Q. laeta* complex using leaf shape analysis using geometric morphometrics methods. Analyses were performed in 868 mature leaves corresponding to 152 individuals. Each leaf was scanned using an HP Scanjet 300 with a ruler as a size reference. To obtain the overall leaf shape, in each image, a fan of 15 guidelines covering the entire leaf contour was created using the program MakeFan6 of the "Integrated Morphometrics Packaged" (IMP) series to record two landmarks (leaf base and leaf apex) and 26 semi-landmarks (along the leaf contour) with the program TpsDig (Rohlf, 2005). The Tps file includes an *x* and *y* coordinate for each landmark and semi-landmark per leaf, plus the two scale points. To remove the effects of rotation, translation, and scale, a Procrustes superimposition analysis was performed using the program MorphoJ version1.05d (Klingenberg, 2011), generating a covariance matrix. We used this matrix as input for a principal components analysis (PCA) and canonical discriminant analysis (CDA). For both the PCA and the CDA, all collected specimens were arranged according to the four different grouping criteria used for nuclear and chloroplast genetic variations. Finally, we evaluated the percentage of morphological assignment based on the shape of the leaf using *a priori* grouping that is classified into *a posteriori* grouping from a classification matrix.

3 Results

3.1 Nuclear genetic diversity

Because most microsatellite loci failed to amplify in La Malinche (population 18), we eliminated this population from the nuclear genetic diversity and structure analyses. The results from the FreeNA analysis indicated that all loci showed the presence of null alleles, varying in frequency in the populations from 0 to 0.43. However, there was no effect of null alleles on the genetic differentiation estimates, since the F_{ST} value not using ENA was 0.0628, while F_{ST} using ENA was 0.0625. In total, 103 alleles were identified in the 18 populations across nine microsatellites. Genetic diversity was high across all populations ($N_e = 3.223–5.132$; $H_E = 0.622–0.777$) (Table 1). N_e and H_E values were the highest in populations 14 and 3, respectively, and the lowest in populations 8 and 15, respectively. Rarefied allelic richness per locus ranged from 2.462 to 6.343, while at the population level, the highest allelic richness was found in population 14 ($A = 4.711$) and the lowest in population 10 ($A = 3.512$) (Table 1). Taking null alleles into account, the inbreeding coefficient (F_{IS}INEST) was 0.0071 in contrast to $F_{IS} = 0.095$, which was obtained without correction.

3.2 Nuclear genetic structure

The unrooted NJ tree based on Nei's standard genetic distance (Nei, 1987) showed that populations clustered into three groups (Figs. 2A, 2B). However, tree nodes were weakly supported (<30%). The first group (in green) was constituted by five populations, three of them classified as morphotype *centralis* (populations 15, 16, and 17) and two as morphotype *laeta* (populations 13 and 14). This group is distributed in the trans-mexican volcanic belt (TMVB). The second group (in yellow) was composed of seven populations that belong to morphotypes *transmontana* (central-northern Mexico, populations 6, 8, 9, 10, 11, and 12) and *transmontana coyula* (Oaxaca, south of Mexico; population 19). Finally, the third group (in red) is restricted to the north of Mexico and consists of six populations that belong to the morphotype *prinopsis* (populations 1, 2, 4, 5, and 7) and one population assigned to morphotype *laeta* (population 3).

In contrast, the DAPC analysis detected five clusters (Fig. 3) ($K = 5$; BIC = 221.9785; Fig. S1), which were in general incongruent with the morphotype assignment of the individuals. The first genetic cluster was composed of a high proportion of individuals that belong to the *transmontana* (0.70) and *transmontana coyula* morphotypes (0.22). The second genetic cluster was formed mainly by individuals belonging to the *prinopsis* (0.42) and *transmontana* morphotypes (0.24). The third cluster had a higher proportion of individuals included in the *transmontana* (0.70) and *centralis* morphotypes (0.15). The fourth cluster showed a considerable mixture of individuals belonging to morphotypes *transmontana* (0.34) and *prinopsis* (0.32), and the fifth cluster showed a higher proportion of individuals identified as *prinopsis* (0.52), *centralis* (0.18), and *laeta* (0.18) (Table S3). The multivariate distance represented by the minimum spanning tree suggests that clusters two and five are genetically closer to each other, while cluster one is genetically more different from the rest of the identified clusters (Fig. 3).

Table 1 Genetic diversity statistics for cpDNA sequences and nSSRs for 19 populations of the *Quercus laeta* complex

		cpDNA					nSSR					
Population number	*Morphotype*	*N*	*h*	*Hd*	π	*S*	*Na*	*Ne*	*A*	H_O	H_E	*F*
1	*prinopsis*	9	1	0	0	0	6.889	4.339	4.507	0.617	0.717	0.144
2	*prinopsis*	10	1	0	0	0	6.778	4.307	4.230	0.615	0.690	0.118
3	*laeta*	5	2	0.40	0.00022	1	5.000	4.197	4.456	0.578	0.777	0.216
4	*prinopsis*	10	2	0.40	0.00022	1	6.333	4.370	4.233	0.626	0.694	0.111
5	*prinopsis*	8	3	0.52	0.00032	2	5.333	3.812	3.957	0.673	0.661	−0.007
6	*transmontana*	10	4	0.64	0.00042	3	6.110	4.350	3.865	0.651	0.713	0.087
7	*prinopsis*	8	2	0.38	0.00022	1	5.333	3.332	3.893	0.621	0.625	0.040
8	*transmontana*	8	1	0	0	0	4.778	3.223	4.711	0.599	0.624	0.034
9	*transmontana*	9	2	0.35	0.00020	1	4.667	3.477	3.859	0.559	0.625	0.186
10	*transmontana*	10	1	0	0	0	6.222	4.458	4.247	0.678	0.694	0.025
11	*transmontana*	11	3	0.51	0.00042	3	6.222	4.083	3.718	0.693	0.682	−0.037
12	*transmontana*	11	1	0	0	0	7.111	4.421	4.233	0.573	0.739	0.229
13	*laeta*	10	1	0	0	0	5.667	3.775	3.608	0.608	0.632	0.012
14	*laeta*	10	2	0.50	0.00028	1	7.111	5.132	3.512	0.700	0.757	0.073
15	*centralis*	5	1	0	0	0	4.222	3.253	4.251	0.578	0.622	0.107
16	*centralis*	9	1	0	0	0	6.000	4.026	4.047	0.677	0.720	0.067
17	*centralis*	9	1	0	0	0	5.222	3.298	4.367	0.639	0.650	0.032
18	*centralis*	5	1	0	0	0	NA	NA	NA	NA	NA	NA
19	*transmontana coyula*	9	2	0.50	0.00085	3	5.667	4.108	4.083	0.500	0.693	0.278

For cpDNA: *N*, sample size; *h*, number of haplotypes; *Hd*, haplotypic diversity; π, nucleotide diversity; and *S*, number of polymorphic sites. For nSSRs: *Na*, mean number of alleles per locus; *Ne*, mean number of effective alleles per locus; *A*, allelic richness after rarefaction; H_O, mean observed heterozygosity; H_E, mean expected heterozygosity; and *F*, inbreeding coefficient.

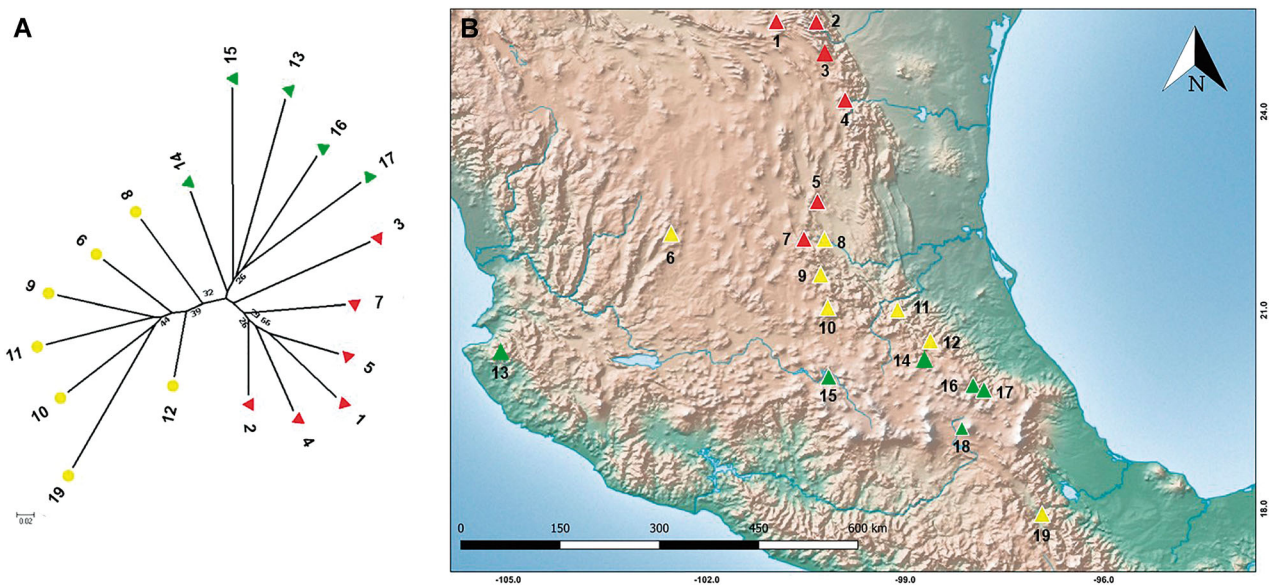


Fig. 2. Neighbor-joining tree for *Quercus laeta* complex. **A**, Unrooted neighbor-joining tree based on Nei's (1987) distance indicating the relationships among 18 populations of the *Q. laeta* complex. **B**, Geographic distribution of the three genetic groups found by POPULATIONS.

All STRUCTURE runs using the complete data set and under the different models converged on the same results, with $K = 2$ showing the highest support based on the ΔK statistic (Fig. S2). Therefore, here, we report the results obtained when geographical populations are used as a prior, while the results from using morphotype assignment as a prior and when no prior was used are presented in the supplementary information (Fig. S3). Most of the populations had an assignment index to their genetic group $Q > 0.80$, except for populations 3, 8, 12, 14, and 15, with *Q* values of

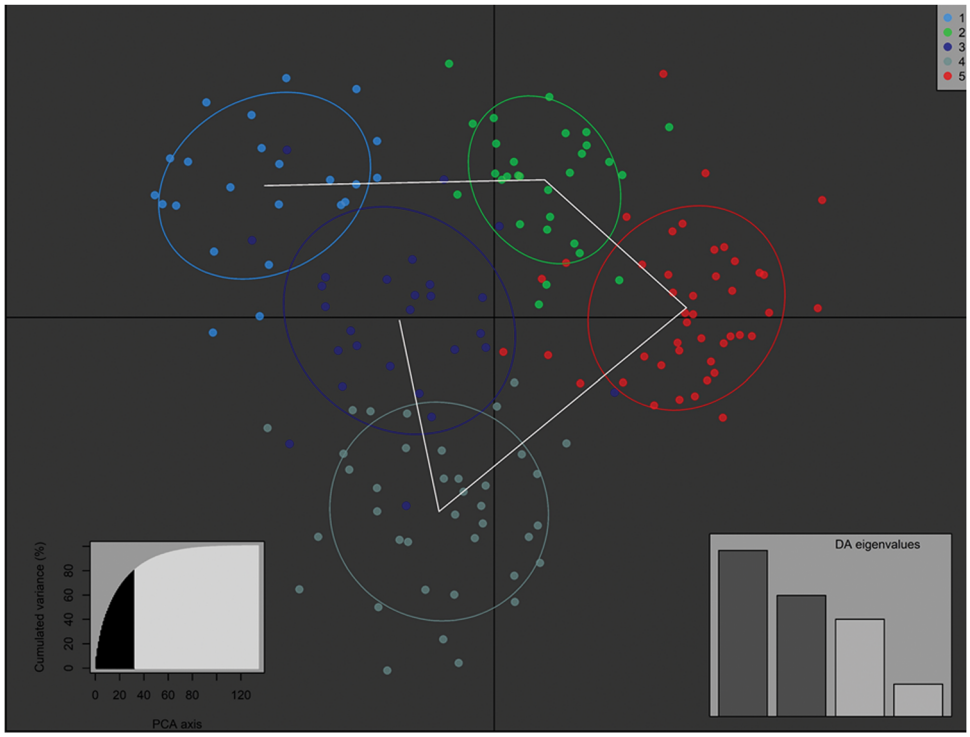


Fig. 3. Scatterplots of the discriminant analysis of principal components among individuals of five morphotypes within the *Quercus laeta* complex. The white line shows the relationships among genetic clusters based on a minimum spanning tree.

0.73, 0.66, 0.68, 0.69, and 0.55, respectively. The distribution map of the genotypic clusters for $K = 2$ across populations showed that the genetic group (GG) 1 (in green) is restricted to the central part of Mexico (populations 6, 8, 9, 10, 11, and 14) and population 19 from the south of Mexico (Fig. 4). The samples classified as morphotypes *transmontana* and *transmontana coyula* belong to this genetic group. The populations that comprised this genetic group are the same as the yellow group identified by the NJ tree analysis and the genetic cluster 1 identified by DAPC. In contrast, the GG2 (in red) is widely distributed mainly along the TMVB (populations 12, 13, 15, 16, and 17) and the north of Mexico (populations 1, 2, 3, 4, 5, and 7) (Fig. 4). The green and red genetic groups detected by NJ tree analysis and the genetic cluster 5 identified by DAPC are analogous to this STRUCTURE genetic group. The GG2 is formed by individuals assigned to morphotypes *centralis*, *laeta*, and *prinopsis*.

When evaluating the substructure within each GG, we identified different genetic subunits. For the GG1, we identified a substructure with $K = 2$, where the first genetic subunit (S1A) was represented in populations 8 and 9, while the second genetic subunit (S1B) had a greater proportion in populations 6, 10, 11, 14, and 19 (Fig. 5A). An interesting pattern was identified within the GG2, where the optimal value was $K = 4$ (Fig. 5B). The first genetic subunit (S2A) was represented in the north of Mexico within populations 1, 2, 4, 5, and 7 (*prinopsis* morphotype). The second genetic subunit (S2B) corresponded to population 13 (*laeta* morphotype), which is the most geographically isolated population located in western Mexico. The third genetic subunit (S2C) was composed mainly by individuals of populations 3, 12, and 15 (*centralis*, *laeta*, and *prinopsis* morphotypes). The fourth genetic subunit (S2D) was composed by individuals of populations 16 and 17 (*centralis* morphotype). The S2B, S2C, and S2D were located across the TMBV in the center of Mexico.

The different AMOVAs showed significant genetic differentiation at all levels tested. The magnitude of overall genetic differentiation was low ($F_{ST} = 0.070$; $P < 0.001$), as also differentiation among morphotypes ($F_{CT} = 0.032$; $P < 0.0001$), the two main genetic groups according to STRUCTURE ($F_{CT} = 0.031$; $P < 0.001$), genetic groups according to NJ tree analysis ($F_{CT} = 0.030$; $P < 0.0001$), and genetic clusters obtained by DAPC ($F_{CT} = 0.014$; $P < 0.04$) (Table 2).

Pairwise fixation indices were relatively high between morphotypes *prinopsis—transmontana* ($F_{ST} = 0.050$ and $R_{ST} = 0.076$), *prinopsis—transmontana coyula* ($F_{ST} = 0.110$ and $R_{ST} = 0.166$), *laeta—prinopsis* ($F_{ST} = 0.034$ and $R_{ST} = 0.099$), and *laeta—transmontana coyula* ($F_{ST} = 0.063$ and $R_{ST} = 0.067$) (Table 3). A significantly higher R_{ST} value than the corresponding F_{ST} value was only observed between *laeta—prinopsis* ($p = 0.01$) and *prinopsis—transmontana* morphotypes ($p = 0.02$). By contrast, low fixation indices were found between morphotypes *centralis* and *laeta* ($F_{ST} = 0.017$ and $R_{ST} = 0.035$) and *transmontana* and *transmontana coyula* ($F_{ST} = 0.046$ and $R_{ST} = 0.003$) (Table 3).

3.3 Chloroplast DNA population differentiation

The final alignment of the sequences of the intergenic regions *trnC-ycf6* and *ycf6-psbM* (GenBank Accession Numbers: MW999693–MW999830; MZ028207–MZ028344) was 1810 pb

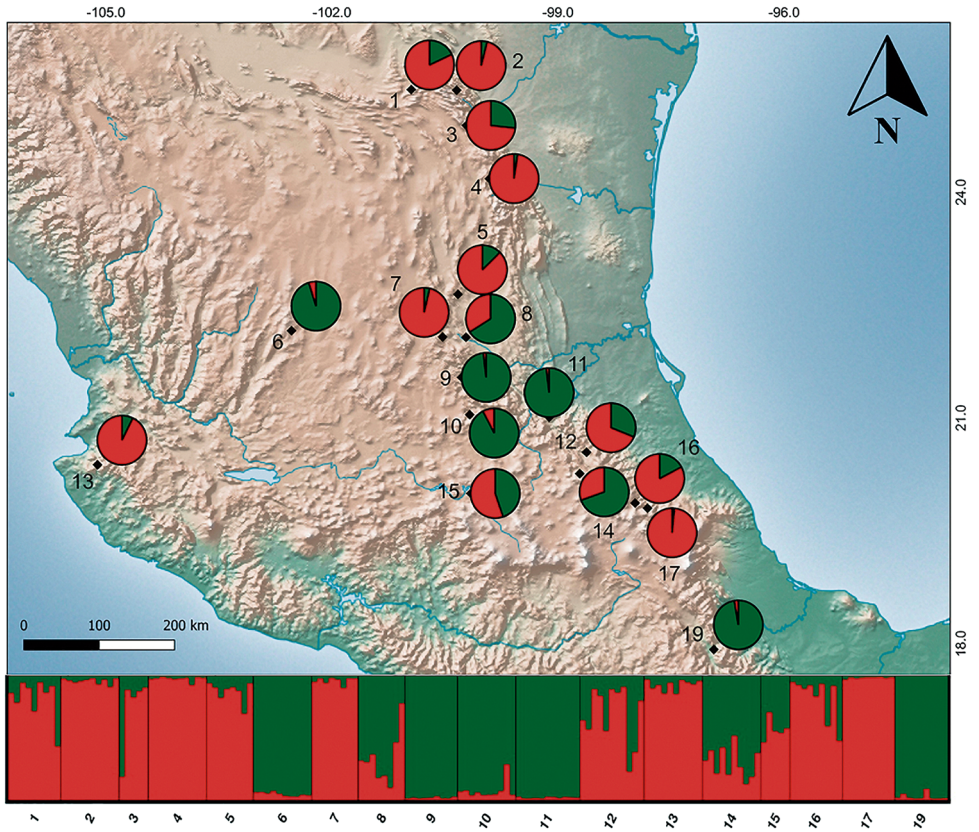


Fig. 4. Bayesian analysis of population structure for the *Quercus laeta* complex. Distribution map of ancestry proportions identified by structure for *Q. laeta* complex for $K = 2$.

in length. A total of 15 polymorphic sites were observed, defining 13 haplotypes. Haplotype and nucleotide diversities were in general low, ranging within populations from $Hd = 0$–0.644 and from $\pi = 0$ to 0.00085 (Table 1). The statistical parsimony haplotype network showed a star-like shape where haplotype A had a central position and was the most frequent and geographically widespread (present in 18 populations, and all morphotypes) (Fig. 6). Most of the other haplotypes had a low frequency and were restricted to single populations (private haplotypes). Nine populations were monomorphic, eight of which showed the central haplotype. Populations with the highest number of haplotypes

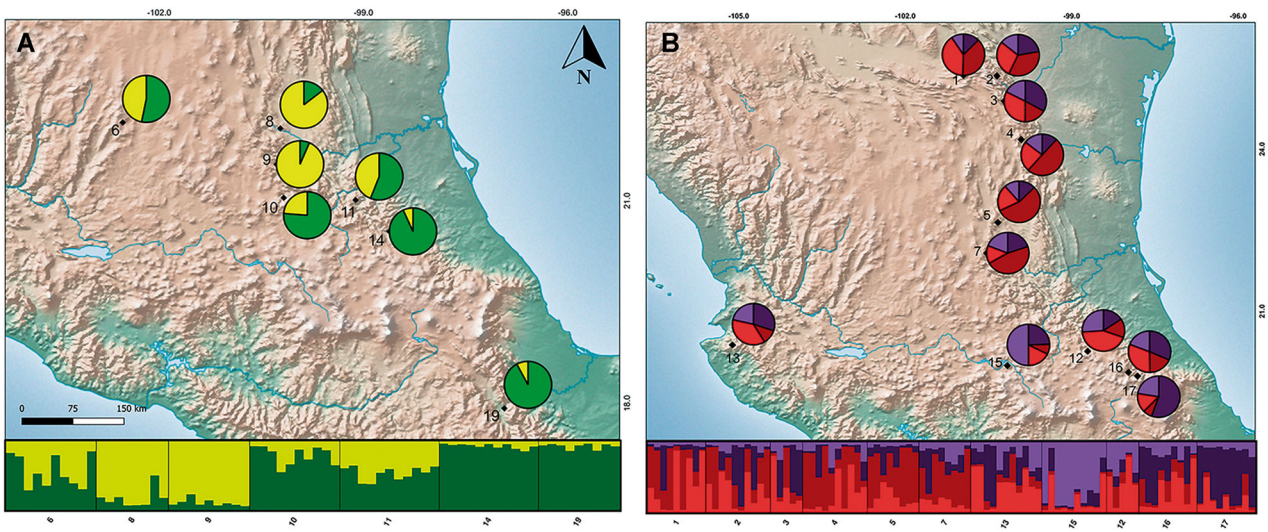


Fig. 5. Bayesian analysis of population substructure. **A**, Substructure within GG1 ($K = 2$). **B**, Substructure within GG2 ($K = 4$). Small rhombuses represent sampling localities.

Table 2 Analysis of molecular variance for nine nSSR loci at different levels within the *Quercus laeta* complex

Group	Source of variation	*d.f.*	Sum of squares	Variance components	Percentage of variation	
No groups defined	Among populations	17	131.317	0.248	7	$F_{ST} = 0.070$***
	Within populations	304	1004.950	3.306	93	
	Total	321	1136.267		100	
Morphotypes	Among groups	4	51.805	0.114	3.20	$F_{CT} = 0.032$***
	Among populations within groups	13	79.512	0.158	4.40	$F_{SC} = 0.046$***
	Within populations	304	1004.950	3.306	92.40	$F_{ST} = 0.076$***
	Total	321	1136.267		100	
Genetic groups (POPULATIONS)	Among groups	2	35.619	0.108	3	$F_{CT} = 0.031$***
	Among populations within groups	15	95.698	0.173	4.81	$F_{SC} = 0.050$***
	Within populations	304	1004.950	3.306	92.18	$F_{ST} = 0.078$***
	Total	321	1136.267		100	
Genetic groups (DAPC)	Among groups	4	30.766	0.039	1.47	$F_{CT} = 0.014$*
	Among populations within groups	13	66.668	0.149	5.53	$F_{SC} = 0.056$***
	Within populations	304	764.618	2.51	93.00	$F_{ST} = 0.070$***
	Total	321	862.053		100	
Genetic groups (STRUCTURE)	Among groups	1	24.571	0.113	3.12	$F_{CT} = 0.030$***
	Among populations within groups	16	106.747	0.189	5.24	$F_{SC} = 0.054$***
	Within populations	304	1004.950	3.306	91.63	$F_{ST} = 0.084$***
	Total	321	683.487		100	

*$p \leq 0.05$; ***$p \leq 0.001$.

were populations 5, 6, and 11 belonging to morphotypes *prinopsis* and *transmontana* (Table 1; Fig. 6).

The AMOVAs showed different levels of differentiation. The magnitude of differentiation among populations was high ($\Phi_{ST} = 0.31$; $P < 0.00001$) and moderate among morphotypes ($\Phi_{CT} = 0.20$, $P < 0.005$), while the differentiation among genetic groups according to NJ tree analysis ($\Phi_{CT} = -0.00624$), genetic clusters obtained by DAPC ($\Phi_{CT} = -0.00117$), and genetic groups according to STRUCTURE ($\Phi_{CT} = -0.00115$) was low and not significant (Table 4).

3.4 Morphological analysis of leaf shape

The principal components analysis of the foliar shape was unable to discriminate clear entities under any of the *a priori* groupings (Fig. S4). On the other hand, the analysis of canonical variables recovered four variates that explained 100% of the variation (62.32%, 24.19%, 10.19%, 3.3%, respectively). When we used the morphotypes as well the genetic clusters obtained by the NJ tree and DAPC analysis as grouping criteria, it was not possible to discriminate among entities because the boundaries between the different groups overlap (Figs. 7A–7C). However, when we used the two main genetic groups detected by STRUCTURE, the frequency distribution of the first discriminant function indicated a clear morphometric differentiation (Fig. 7D). The GG1 (in blue), including the morphotypes *transmontana* and *transmontana coyula*, have narrower (lanceolate) leaf forms, while populations in the GG2 (in red), including the

Table 3 Pairwise fixation indices among morphotypes of the *Quercus laeta* complex using F_{ST} (below diagonal) and R_{ST} (above diagonal) over nine nSSR loci

Q. laeta morphotypes					
centralis	*laeta*	*prinopsis*	*transmontana*	*transmontana coyula*	*Q. laeta* morphotypes
----------	0.035*	0.041*	0.016	0.049	*centralis*
0.017*	----------	**0.099***	0.023*	0.067*	*laeta*
0.025*	0.034*	----------	**0.076***	0.166*	*prinopsis*
0.047*	0.028*	0.050*	----------	0.003	*transmontana*
0.101*	0.063*	0.110*	0.046*	----------	*transmontana coyula*

*$p \leq 0.05$. Allele size permutation tests indicate whether R_{ST} is significantly larger than F_{ST} (highlighted in bold).

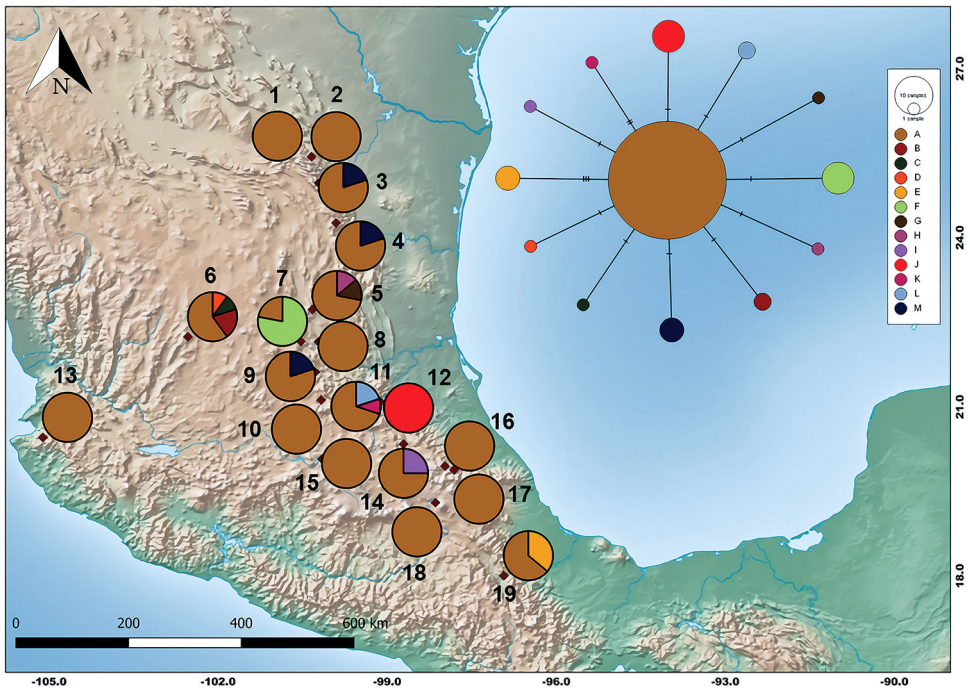


Fig. 6. Geographic distribution and statistical parsimony network of 13 haplotypes of the *Quercus laeta* complex. Small rhombuses represent sampling localities, and pie charts represent frequencies of the haplotypes found in each sampling locality.

Table 4 Analyses of molecular variance for cpDNA variation at different levels within the *Quercus laeta* complex

Group	Source of variation	*d.f.*	Sum of squares	Variance components	Percentage of variation	
No groups defined	Among populations	18	53.276	0.337	32	$\Phi_{ST} = 0.317$***
	Within populations	108	78.267	0.724	68	
	Total	126	131.543	1.062		
Morphotypes	Among groups	4	29.403	0.232	21	$\Phi_{CT} = 0.209$*
	Among populations within groups	14	23.874	0.154	14	$\Phi_{SC} = 0.175$***
	Within populations	108	78.267	0.724	65	$\Phi_{ST} = 0.348$***
	Total	126	131.543	1.111		
Genetic groups (POPULATIONS)	Among groups	2	5.658	−0.006	0	$\Phi_{CT} = -0.006$
	Among populations within groups	16	47.618	0.341	32	$\Phi_{SC} = 0.320$***
	Within populations	108	78.267	0.724	68	$\Phi_{ST} = 0.316$***
	Total	126	131.543	1.059		
Genetic groups (DAPC)	Among groups	4	13.861	−0.001	0	$\Phi_{CT} = -0.001$
	Among populations within groups	14	39.415	0.338	32	$\Phi_{SC} = 0.318$***
	Within populations	108	78.267	0.724	68	$\Phi_{ST} = 0.317$***
	Total	126	131.543	1.0617		
Genetic groups (STRUCTURE)	Among groups	1	3.284	−0.001	0	$\Phi_{CT} = -0.001$
	Among populations within groups	17	49.992	0.337	32	$\Phi_{SC} = 0.318$***
	Within populations	108	78.267	0.724	68	$\Phi_{ST} = 0.317$***
	Total	126	131.543	1.061		

*$p \leq 0.05$; ***$p \leq 0.0001$.

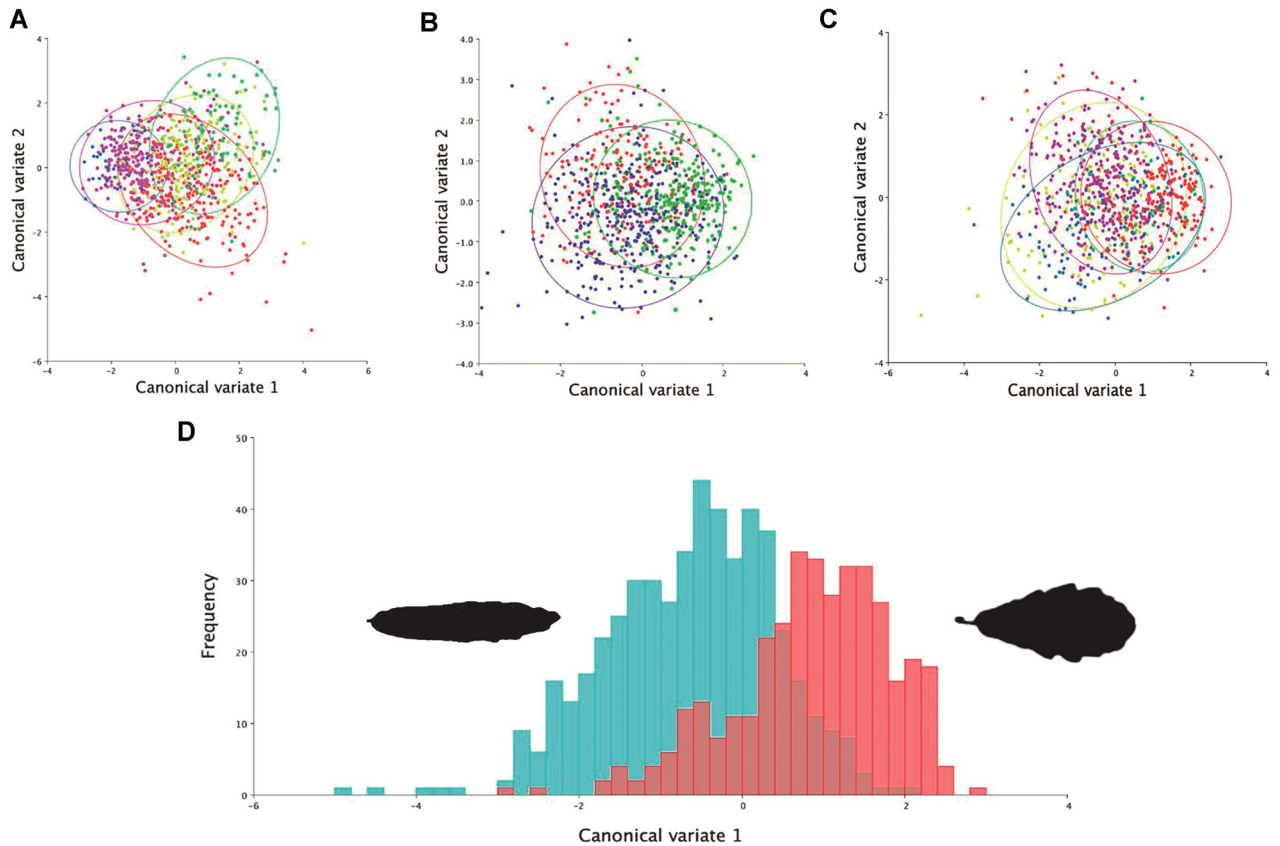


Fig. 7. Canonical variate analyses of two landmarks (leaf base and leaf apex) and 26 semi-landmarks (along the leaf contour) from 152 specimens of the *Quercus laeta* complex. **A**, Plot grouping individuals according to morphotypes; **B**, Plot grouping individuals according to the NJ tree; **C**, Plot with the DAPC groups; and **D**, Histogram of the first canonical discriminant function (CD 1) scores by individual for the leaf shape analysis of genetic group 1 (in blue) and genetic group 2 (in red).

morphotypes *laeta*, *prinopsis*, and *centralis*, are characterized by obovate and oblanceolate leaf forms (Fig. 7D). These results indicate that the genetic groups detected by STRUCTURE show morphometric differentiation within the *Q. laeta* complex. The assignment of the different individuals shows correspondence among the genetic groups and the leaf form of 74% for GG1 and 78% for GG2.

4 Discussion

In the present study, evidence from the nuclear genome (nSSRs), the plastome (cpDNA sequences), and morphology, quantified using geometric morphometrics, showed evidence for recognizing at least two different biological entities within the *Quercus laeta* complex that could be proposed as species (Table 5). The genus *Quercus* has long been a source of debate regarding species concepts (Engelmann, 1876; Ehrlich & Raven, 1969; Burger, 1975; Van Valen, 1976), mainly due to the high interspecific gene flow. As a consequence, low genetic differentiation is typical among oak species (Muir & Schlötterer, 2005; Lexer et al., 2006; Peñaloza-Ramírez et al., 2010; Owusu et al., 2015; Albarrán-Lara et al., 2019). Therefore, the fact that we found a relatively high genetic differentiation with few microsatellite loci may suggest that these biological entities are clearly distinct and with a considerable degree of genetic isolation (Pritchard et al., 2000; Craft et al., 2002).

4.1 CpDNA does not differentiate oak complexes

The results obtained for the cpDNA and nSSRs showed low congruence. For the cpDNA, the AMOVA indicated high and significant genetic differentiation among populations and moderate differentiation among morphotypes, but low and nonsignificant differentiation among the nuclear genetic groups obtained with POPULATIONS, DAPC, and STRUCTURE. Furthermore, a star-shaped haplotype network was observed, suggesting the existence of a single main cpDNA lineage in the *Q. laeta* complex with many low-frequency exclusive haplotypes found in the different morphotypes, with the *centralis* morphotype as an exception, since it is the only morphotype that did not show an exclusive haplotype. Haplotype sharing has been commonly reported among oak species and interpreted as evidence for interspecific gene flow (Petit et al., 1997, 2003; Dumolin-Lapègue et al., 1999; González-Rodríguez et al., 2004), even though incomplete lineage sorting may also play a role (Muir & Schlötterer, 2005; Lexer et al., 2006; Muir & Schlöetterer, 2006). In this context, we propose that the star-like shape of the network and haplotype sharing among populations within the *Q. laeta* complex could be the result of rapid population expansions

Table 5 Comparison of grouping patterns based on different data and methods

Data type	Method	Cluster number	Grouping patterns
cpDNA	AMOVA	1	*transmontana* + *transmontana coyula* + *centralis* + *laeta* + *prinopsis*
nSSR	DAPC	5	*transmontana* + *transmontana coyula*; *prinopsis* + *transmontana*; *transmontana* + *centralis*; *transmontana* + *prinopsis*; *prinopsis* + *centralis* + *laeta*
nSSR	NJ tree	3	*transmontana* + *transmontana coyula*; *centralis* + *laeta*; *prinopsis*
nSSR	Structure	2	*transmontana* + *transmontana coyula*; *laeta* + *centralis* + *prinopsis*
Morphology	Geometric morphometry	2	*transmontana* + *transmontana coyula*; *laeta* + *centralis* + *prinopsis*

AMOVA, analysis of molecular variance; DAPC, discriminant analysis of principal components.

coupled with instances of chloroplast capture (Tsitrone et al., 2003; Kremer & Hipp, 2020). Chloroplast capture is relatively frequent in species with a sympatric distribution and reproductive compatibility, as has been reported in other Fagales (King & Ferris, 2000; Okaura & Harada, 2002; Acosta & Premoli, 2010) and in European and American white oaks (Whittemore & Schaal, 1991; Petit et al., 1997, 2003). In conclusion, low haplotype differentiation within the *Q. laeta* complex supports the premise that incongruence between cpDNA variation and species limits in oaks generally makes it a poorly suited marker to track fine-scale phylogeny (Petit et al., 1993; Dumolin-Lapègue et al., 1999; Manos et al., 1999; Belahbib et al., 2001; Petit & Excoffier, 2009; Pham et al., 2017; McCauley et al., 2019; Kremer & Hipp, 2020).

4.2 Drawing species limits in the *Q. laeta* complex

The establishment of species boundaries through population assignment based on clustering is useful in the early stages of species delimitation, as it groups individuals that likely exchange genes or have similar allele frequencies due to shared ancestry (Pritchard et al., 2000; Rittmeyer & Austin, 2012; Carstens et al., 2013). In this context, our nuclear genetic data showed a consistent pattern of genetic differentiation across different methods based on both *a priori* and *de novo* designations. Even though the unrooted NJ analysis showed three groups, the DAPC *de novo* analysis identified five genetic clusters and the Bayesian analyses identified two clusters at the highest hierarchical level, and up to six genetic clusters, it is possible to identify a major constant genetic discontinuity within the *Q. laeta* complex. On the one hand, based on genetic cluster 1 identified by DAPC, GG1 (in green) detected by STRUCTURE, and the yellow group identified by the NJ analysis (Figs. 2–4), our results suggest that individuals from the *transmontana* and *transmontana coyula* morphotypes could be considered as the same entity (hereafter called *Q. transmontana*). On the other hand, based on the genetic clusters 2 and 5 identified by DAPC analysis, the GG2 (in red) detected by STRUCTURE, and the red and green groups detected by the NJ analysis (Figs. 2–4), the second major genetic cluster is formed by individuals belonging to morphotypes *centralis*, *laeta*, and *prinopsis* (hereafter called *Q. laeta*).

An interesting result of this study, besides the clear genetic discontinuity between *Q. transmontana* and *Q. laeta*, is that use of complementary analytical methods allowed us to detect potential recent admixture events and to identify nested clusters. For example, the DAPC analysis recovered genetic clusters with mixed membership (genetic clusters 3 and 4; Fig. 3; Table S3), which might suggests that although *Q. transmontana* and *Q. laeta* are two different biological entities, local introgression could have occurred between them in some instances.

In the case of the identification of nested clusters, the NJ and the substructure analysis suggested that *Q. laeta* itself is formed by two geographical groups segregated into up to four genetic clusters. The northern group (*prinopsis* morphotype), which was found in the Sierra Madre Oriental, is formed mainly by a single genetic group, while the geographic group of the TMBV (*centralis* + *laeta* morphotypes) is formed by three genetic groups, suggesting that these populations show significant genetic heterogeneity worthy of further study. For example, if we compare fixation indexes among morphotypes that conform to *Q. laeta*, we observe that (i) values between *prinopsis* and *laeta* (F_{ST} = 0.034, R_{ST} = 0.099) are within or above those typically reported between taxonomically valid Mexican oak species (e.g., F_{ST} = 0.016, R_{ST} = 0.042; Albarrán-Lara et al., 2019; F_{ST} = 0.022; Peñaloza-Ramírez et al., 2010) and (ii) significantly larger R_{ST} than F_{ST} values between *prinopsis* and *laeta* suggest that stepwise mutations have contributed to genetic differentiation between these entities, which could indicate low levels of gene flow between both groups. Intra-specific genetic clusters are generally interpreted as evidence of past population fragmentation (Hardy et al., 2003) and since historical fragmentation may be related to processes of divergence, we suggest that the *prinopsis* morphotype could represent a separate species from *Q. laeta* resulting from allopatric isolation, although a more thorough evaluation is necessary. In conjunction, our results suggest that although all morphotypes within the *Q. laeta* complex may not correspond to discrete or strongly differentiated genetic units, the complex is not a genetic continuum either.

Interestingly, the other important line of evidence, morphometric variation, did not indicate clear differentiation among morphotypes, or genetic groups detected by the NJ tree analysis. However, when we compared the above-described two main groups (*Q. transmontana* and *Q. laeta*), we found a clear differentiation between the narrow leaves of *Q. transmontana* and the obovate leaves of *Q. laeta*. Furthermore, when we contrasted these two genetic–morphometric groups to the species hypotheses (morphotypes) initially used based on qualitative morphological data, we observed that there is no

total correspondence. This suggests that the exclusive use of qualitative micro- and macromorphological characteristics that historically have been important to identify boundaries between Mexican oaks species might not be enough to establish boundaries in taxonomically complicated *Quercus* complexes, so we suggest using multiple lines of evidence.

5 Conclusion

The basic goal of this study was to assess species delimitation and relationships within the *Q. laeta* complex. Our results demonstrate that (i) not all the morphotypes belong to genetically and morphologically discrete units, but (ii) *Quercus laeta*, as currently recognized, comprises a minimum of two genetically, morphologically, and ecologically coherent populations, which likely warrant recognition as distinct species, *Quercus transmontana* and *Q. laeta*, and (iii) the morphotype *prinopsis*, distributed along dry oak forests in the north of the Sierra Madre Oriental, could represent a cryptic species from *Q. laeta*, although a more thorough evaluation is necessary. Our findings help explain the inconsistent results obtained in previous phylogenomic studies regarding these taxa (Hubert et al., 2014; Hipp et al., 2019, 2020), which did not consider these two species as different. Ongoing taxonomic re-evaluation of *Q. laeta* may show that the complex may even comprise more than two species.

Acknowledgements

We thank Oscar De Luna, Gonzalo Contreras, Ricardo Gaytán, and Maribel Arenas for their help with field collection. Jesús Llanderal provided technical help with microsatellite sequencing. Oscar De Luna provided help with morphometric analyses. Also, we thank Goretty Mendoza for her laboratory assistance and M. Socorro González Elizondo for providing feedback on taxonomic issues. The authors thank the LABORATORIO NACIONAL DE ANÁLISIS Y SÍNTESIS ECOLÓGICA (LANASE-UNAM) and Marco Antonio Villalda Quezada for technical support for data analyses. Comments by an anonymous reviewer and Olivier Lepais on previous versions of the manuscript are greatly appreciated. Saddan Morales-Saldaña thanks Posgrado en Ciencias Biológicas—UNAM and Consejo Nacional de Ciencia y Tecnología (CONACYT) for the PhD scholarship and financial support to develop graduate studies at UNAM (scholarship/CVU: 483720/624511). This study constitutes a partial fulfillment of the Graduate Program in Biological Sciences of the National Autonomous University of Mexico (UNAM). Partial economic support for this research was provided by project PAPIIT-DGAPA-UNAM IN 210020 to Antonio González-Rodríguez.

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle AC, Buggs R, Butlin KR, Dieckmann U, Eroukhmanoff F, Grill A, Cahan HA, Hermansen SJ, Hewitt G, Hudson G, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Möst M, Mullen S, Nichols R, Nolte WA, Parisod C, Pennig K, Rice MA, Ritchie GM, Seifert B, Smadja MC, Stelkens R, Symura MJ, Vainöla R, Wolf WBJ, Zinner D. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26: 229–246.

Acosta MC, Premoli AC. 2010. Evidence of chloroplast capture in south american *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). *Molecular Phylogenetics and Evolution* 54: 235–242.

Adams CD, Rohlf FJ, Slice DE. 2013. A field comes of age: Geometrics morphometrics in the 21st century. *Hystrix, The Italian Journal of Mammalogy* 24: 7–14.

Albarrán-Lara A, Petit R, Kremer A, Caron H, Peñaloza-Ramírez JM, Gugger PF, Dávila-Aranda PD, Oyama K. 2019. Low genetic differentiation between two morphologically and ecologically distinct giant-leaved Mexican oaks. *Plant Systematics and Evolution* 305: 89–101.

Aldrich PR, Sun CH, Romero-Severson J. 2002. Microsatellite markers for northern red oak (Fagaceae: *Quercus rubra*). *Molecular Ecology Notes* 2: 472–474.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.

Barley JA, White J, Diesmos CA, Brown RM. 2013. The challenge of species delimitation at the extremes: Diversification without morphological change in Philippine sun skinks. *Evolution* 67: 3556–3572.

Belahbib N, Pemonge MH, Ouassou A, Sbay H, Kremer A, Petit RJ. 2001. Frequent cytoplasmic exchanges between oak species that are not closely related: *Quercus suber* and *Q. ilex* in Morocco. *Molecular Ecology* 10: 2003–2012.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winter K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* 22: 148–155.

Bruschi P, Grossoni P, Bussotti F. 2003. Within- and among tree variation in leaf morphology of *Quercus petraea* (Matt.) Liebl. natural populations. *Trees* 17: 164–172.

Burger WC. 1975. The species concept in *Quercus*. *Taxon* 24: 45–50.

Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013. How to fail at species delimitation. *Molecular Ecology* 22: 4369–4383.

Cavender-Bares J. 2019. Diversification, adaptation, and community assembly of the American oaks (*Quercus*), a model clade for integrating ecology and evolution. *New Phytologist* 221: 669–692.

Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24: 621–631.

Chybicki J, Burczyk J. 2009. Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity* 100: 106–113.

Corl A, Davis AR, Kuchta SR, Sinervo B. 2010. Selective los of polymorphic mating types is associated with rapid phenotypic evolution during morphic speciation. *Proccedings of the National Academy of Sciences of the USA* 107: 4254–4259.

Craft KJ, Ashley MV, Koenig WD. 2002. Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. *American Journal of Botany* 89: 1792–1798.

Darwin C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.

Desmond SC, Garner M, Flannery S, Whittemore AT, Hipp AL. 2021. Leaf shape and size variation in bur oaks: An empirical study and simulation of sampling strategies. *American Journal of Botany* 108: 1540–1554.

Dumolin-Lapègue S, Kremer A, Petit RJ. 1999. Are chloroplast and mitochondrial DNA variation species independent in oaks? *Evolution* 53: 1406–1413.

Earl AD, von Holdt BM. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

Elias TS. 1980. *The complete trees of North America. Field guide and natural history*. New York: Van Nostrand Reinhold Company & Times Mirror Magazines Inc.

Ehrlich PR, Raven PH. 1969. Differentiation of populations: Gene flow seems to be less important in speciation than the neo-Darwinians thought. *Science* 165: 1228–1232.

Engelmann G. 1876. The oaks of the United States. *Transactions of the Academy of Sciences St. Louis* 3: 539–543.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the sofware STRUCTURE: A simulation study. *Molecular Ecology* 14: 2611–2620.

Ewédjè KBEE, Jansen, S, Kouame, KG, Staquet, A, Piñeiro, R, Abessole, ER, Laurier, EON, Daïnou, K, Bernand, BA, Doucet, JL, Hardy JO. 2020. Species delimitation in the African tree genus *Lophira* (Ochnaceae) reveals cryptic genetic variation. *Conservation Genetics* 21: 501–514.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5, a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.

González-Rodríguez A, Arias DM, Valencia S, Oyama K. 2004. Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks. *American Journal of Botany* 91: 401–409.

González-Rodríguez A, Oyama K. 2005. Leaf morphometric variation in *Quercus affinis* and *Q. laurina* (Fagaceae), two hybridizing Mexican red oaks. *Botanical Journal of the Linnean Society* 147: 427–435.

Goudet J. 2001. Fstat, a program to estimate and test gene diversities and fixation indices (v. 2.9.3). http://www.unil.ch/izea/softwares/fstat.html

Guichoux E, Lagache L, Wagner S, Léger P, Petit RJ. 2011. Two highly validated multiplexes (12-plex and 8-plex) for species delimitation and parentage analysis in oaks (*Quercus* spp.). *Molecular Ecology Resources* 11: 578–585.

Hardin JW. 1975. Hybridization and introgression in *Quercus alba*. *Journal of the Arnold Arboretum* 56: 336–363.

Hardy OJ, Charbonnel N, Fréville H, Heuertz M. 2003. Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation. *Genetics* 163: 1467–1482.

Hardy OJ, Vekemans X. 2002. SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618–620.

Habel CJ, Zanchos FE, Dapporto L, Rödder D, Radespiel U, Tellier A, Schmitt T. 2015. Population genetics revisited-towards a multidisciplinary research field. *Biological Journal of the Linnean Society* 115: 1–12.

Hall TA. 1999. BioEdit: A user-friendly biological sequence alignment editor program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.

Hipp AL, Manos PS, González-Rodríguez A, Hahn M, Kaproth M, McVay JD, Valencia-Avalos S, Cavender JB. 2018. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. *New Phytologist* 217: 439–452.

Hipp AL, Manos PS, Hahn M, Avishai M, Bodénes C, Cavender-Bares J, Crowl A, Deng M, Denk T, Fitz-Gibbon S, Gailing O, González-Elizondo MS, González-Rodríguez A, Grimm GW, Jiang XL, Kremer A, Lesur I, McVay JD, Plomion C, Rodríguez-Correa H, Schulze ED, Simeone MC, Sork VL, Valencia-Avalos S. 2020. Genomic landscape of the global oak phylogeny. *New Phytologist* 226: 1–15.

Hipp AL, Whittermore AT, Garner M, Hahn M, Fitzek E, Guichoux E, Cavender-Bares J, Gugger PF, Manos PS, Pearse IS, Cannon CH. 2019. Genomic identify of white oak species in an eastern North American syngameon. *Annals of the Missouri Botanical Garden* 104: 455–477.

Hubert F, Grimm GW, Jousselin E, Berry V, Franc A, Kremer A. 2014. Multiple nuclear genes stabilize the phylogenetic backbone of the genus *Quercus*. *Systematics and Biodiversity* 12: 405–423.

Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9: 1322–1332.

Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, Andrew RL. 2017. The K=2 conundrum. *Molecular Ecology* 26: 3594–3602.

Jombart T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.

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: 94–108.

Kremer A, Hipp AL. 2020. Oaks: And evolutionary success story. *New Phytologist* 226: 987–1011.

Kampfer S, Lexer C, Glöss J, Steinkellner H. 1998. Characterization of (GA)n microsatellite loci from *Quercus robur*. *Hereditas* 129: 183–186.

King RA, Ferris C. 2000. Chloroplast DNA and nuclear DNA variation in the sympatric alder species, *Alnus cordata* (Lois.) Duby and *A. glutinosa* (L.) Gaertn. *Biological Journal of the Linnean Society* 70: 147–160.

Klingenberg PC. 2011. MORPHOJ: An integrated software package for geometric morphometrics. *Molecular Ecology Resources* 11: 353–357.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. CLUMPAK: A programm for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15: 1179–1191.

Langella O. 1999. Populations, 1.2.32. http://bioinformatics.org/%7E;tryphon/populations/

Lefort F, Douglas GC. 1999. An efficient micro-method of DNA isolation from mature leaves of four hardwood tree species *Acer*, *Fraxinus*, *Prunus*, and *Quercus*. *Annals of Forest Science* 56: 259–263.

Leigh WJ, Bryant D. 2015. POPART: Full feature sofware for haplotype network construction. *Methods in Ecology and Evolution* 6: 1110–1116.

Lexer C, Kremer A, Petit RJ. 2006. Shared alleles in sympatric oaks: Recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology* 15: 2007–2012.

Liebmann FM. 1854. Oversigt Over det Kongelige Danske Videnskabernes Selskabs Forhandlinger og Dets Medlemmers Arbeider. *Copenhagen* 173.

Liu Y, Li Y, Song J, Zhang R, Yan Y, Wang Y, Du FK. 2018. Geometric morphometric analyses of leaf shapes in two sympatric Chinese oaks: *Quercus dentata* Thunberg and *Quercus aliena* Blume (Fagaceae). *Annals of Forest Science* 75: 90–102.

MacDougal DT. 1907. Hybridization of wild plants. *Botanical Gazette* 43: 45–58.

Mallet J. 1995. A species definition for the modern synthesis. *Trends in Ecology & Evolution* 10: 294–299.

Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? *BioEssays* 38(2): 140–149.

Manos SP, Doyle JJ, Nixon KC. 1999. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution* 12: 333–349.

Martins K, Gugger PF, Llanderal-Mendoza J, González-Rodríguez A, Fitz-Gibbon ST, Zhao JL, Rodriguez-Correa H, Oyama K, Sork VL. 2018. Landscape genomics provides evidence of climate associated genetic variation in Mexican populations of *Quercus rugosa*. *Evolutionary Applications* 11: 1842–1858.

McCauley RA, Cortés-Palomec AC, Oyama K. 2019. Species diversification in a lineage of Mexican red oak (*Quercus* section *Lobatae* subsection Racemiflorae) the interplay between distance, habitat, and hybridization. *Tree Genetics & Genomes* 15: 27.

McVaugh R. 1974. Fagaceae. In: McVaugh R. ed. *Flora Novo-galiciana*. Controbutions from the University of Michigan herbarium. 12: 50-30.

McVay JD, Hipp AL, Manos PS. 2017. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society B: Biological Sciences* 284: 20170300.

Miller MJ, Cullingham IC, Peery MR. 2020. The influence of a priori grouping on inference of genetic clusters: Simulation study and literature review of the DAPC method. *Heredity* 125: 269–280.

Muir G, Schlötterer C. 2005. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology* 14: 549–61.

Muir G, Schlöetterer C. 2006. COMMENT: Moving beyond single-locus studies to characterize hybridization between oaks (*Quercus* spp.). *Molecular Ecology* 15: 2301–2304.

Nei M. 1972. Genetic distance between populations. *The American Naturalist* 106: 283–292.

Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.

Nosil P. 2012. *Ecological speciation*. Oxford: Oxford University Press.

Nosil P, Harmon LJ, Seehausen O. 2009. Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution* 24: 145–156.

Okaura T, Harada K. 2002. Phylogeographical structure revealed by chloroplast DNA variation in Japanese beech (*Fagus crenata* Blume). *Heredity* 88: 322–329.

Owusu SA, Sullivan AR, Weber JA, Hipp AL, Gailing O. 2015. Taxonomic relationships and gene flow in four North American *Quercus* species (*Quercus* section *Lobatae*). *Systematic Botany* 40: 510–521.

Palmer EJ. 1948. Hybrid oaks of North America. *Journal of the Arnold Arboretum* 29: 1–48.

Peakall R, Smouse PE. 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic sofware for teaching and research-an update. *Bioinformatics* 28: 2537–2539.

Peñaloza-Ramírez JM, 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: 389–399.

Petit RJ, Bodénès C, Ducousso A, Roussel G, Kremer A. 2003. Hybridization as a mechanism of invasion in oaks. *New Phytologist* 161: 151–164.

Petit RJ, Excoffier L. 2009. Gene flow and species delimitation. *Trends in Ecology & Evolution* 24: 386–393.

Petit RJ, Kremer A, Wagner DB. 1993. Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theoretical and Applied Genetics* 87: 122–128.

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A. 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences USA* 94: 9996–10001.

Pinheiro F, Dantas-Queiroz MV, Palma-Silva C. 2018. Plant species complexes as models to understand speciation and evolution: A review of South American studies. *Critical Reviews in Plant Sciences* 37: 54–80.

Pham KK, Hipp AL, Manos PS, Cronn RC. 2017. A time and a place for everything: Phylogenetic history and geography as joint predictors of oak plastome phylogeny. *Genome* 60: 720–732.

Pinho C, Hey J. 2010. Divergence with gene flow: Models and data. *Annual Review of Ecology, Evolution, and Systematics* 41: 215–230.

Pritchard KJ, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Pritchard JK, Wen W. 2003. *Documentation for STRUCTURE Software: Version 2*. Chicago: University of Chicago Press. http://web.stanford.edu/group/pritchardlab/software/structure2_1.html

QGIS Development Team. 2019. QGIS Geographic information system. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Ramírez-Valiente JA, Center JP, Sparks A, Sparks KL, Etterson JR, Longwell T, Pilz G, Cavender-Bares J. 2017. Population-level differentiation in growth rates and leaf traits in seedlings of the neotropical live oak *Quercus oleoides* grown under natural and manipulated precipitation regimes. *Frontiers in Plant Science* 8: 585–249.

Raymond M, Rousset F. 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rangel SR, Carlos E, Zenteno R, de Lourdes M, Enríquez A. 2002. El género *Quercus* (Fagaceae) en el estado de México. *Annals of the Missouri Botanical Garden* 82: 551–593.

Rittmeyer EN, Austin CC. 2012. The effects of sampling on delimiting species from multi-locus sequence data. *Molecular Phylogenetics and Evolution* 65: 451–463.

Rohlf FJ. 2005. tpsDig, digitize landmarks and outlines, version 2.04. Department of Ecology and Evolution, State University of New York. http://www.sbmorphometrics.org/soft-dataacq.html

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* 34: 3299–3302.

Sáenz-Romero C, Lamy JB, Ducousso A, Musch B, Ehrenmann F, Delzon S, Cavers S, Chalupka W, Dagdas S, Hansen JK, Lee SJ, Liesebach M, Rau HM, Psomas A, Schneck V, Steiner W,

Zimmerman NE, Kremer A. 2017. Adaptive and plastic responses of *Quercus petraea* populations to climate across Europe. *Global Change Biology* 23: 2831–2847.

Schluter D. 2001. Ecology and the origin of species. *Trends in Ecology and Evolution* 16: 372–380.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Small RL. 2005. The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetics analysis. *American Journal of Botany* 92: 124–166.

Sork VL, Squire K, Gugger PF, Steele SE, Levy ED, Eckert AJ. 2016. Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *American Journal of Botany* 103: 33–46.

Sousa V, Hey J. 2013. Understanding the origin of species with genome-scale data: Modelling gene flow. *Nature Review Genetics* 14: 404–414.

Steinkellner H, Fluch S, Turetschek E, Lexer C, Streiff R, Kremer A, Glöss J. 1997. Identification and characterization of (GA/CT) n microsatellite loci from *Quercus petraea*. *Plant Molecular Biology* 33: 1093–1096.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596–1599.

Trelease W. 1924. A model for chloroplast capture. *Memoirs of the National Academy of Sciences* 20: 89–1782.

Tsitrone A, Kirkpatrick M, Levin DA. 2003. A model for chloroplast capture. *Evolution* 57: 1776–1782.

Tsukaia H. 2005. Leaf shape: Genetic controls and environmental factors. *The International Journal of Developmental Biology* 49: 547–555.

Vähä JP, Erkinaro J, Niemela E, Primmer CR. 2007. Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology* 16: 2638–2654.

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

Valencia-A S. 2021. Species delimitation in genus *Quercus* (Fagaceae). *Botanical Sciences* 99: 01–12.

Valencia-A S, Sabas JLR, Soto OJA. 2016. A new species of *Quercus*, section *Lobatae* (Fagaceae) from the Sierra Madre Oriental, Mexico. *Phytotaxa* 269: 120–126.

Van Valen L. 1976. Ecological species, multispecies and oaks. *Taxon* 25: 233–239.

Villaverde T, Pokorny L, Olsson S, Rincón-Barrado M, Johnson MG, Gardner EM, Wickett JN, Molero J, Riina R, Sanmartín I. 2018. Bridging the micro-and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* 220: 636–650.

Viscosi V. 2015. Geometric Morphometrics and leaf phenotypic plasticity: Assessing fluctuating asymmetry and allometry in European white oaks (*Quercus*). *Botanical Journal of the Linnean Society* 179: 335–348.

Viscosi V, Cardini A. 2011. Leaf morphology, taxonomy and geometric morphometrics: A simplified protocol for beginners. *PLoS ONE* 6: e25630.

Viscosi V, Fortini P. 2011. Leaf shape variation and differentiation in three sympatric white oak species revealed by elliptic Fourier analysis. *Nordic Journal of Botany* 29: 632–640.

Weir BS, Cockerham C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.

Whittemore AT, Schaal BA. 1991. Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences of the USA* 88: 2540–2544.

Widener L, Frant JB. 2018. Genetic differentiation and diversity of two sympatric subspecies of *Castilleja affinis*; a comparison between the endangered serpentine endemic (spp. *neglecta*) and its widespread congener (spp. *affinis*). *Conservation Genetics* 19: 365–381.

Winger BM. 2017. Consequences of divergence and introgression for speciation Andean cloud forest birds. *Evolution* 71: 1815–1831.

Zumwalde BA, McCauley RA, Fullinwider IJ, Duckett D, Spence E, Hoban S. 2021. Genetic, morphological, and environmental differentiation of an arid-adapted oak with a disjunct distribution. *Forests* 12: 465.

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12818/suppinfo:

Table S1. Morphological differences among morphotypes of *the Quercus laeta* complex.

Table S2. Number of individuals collected and genotyped by population.

Table S3. Proportion of individuals of each morphotype belonging to each of the five genetic clusters detected by discriminant analysis of principal components.

Fig. S1. Value of BIC versus number of clusters obtained by discriminant analysis of principal components analysis.

Fig. S2. Δ*K* statistic for different STRUCTURE runs. A, Using morphotype grouping as prior (LOCPRIOR option). B, Using geographical populations as a prior, and C, not using the LOCPRIOR option.

Fig. S3. Bayesian analysis of population structure for *Quercus laeta* complex. A, using morphotype grouping as prior (LOCPRIOR option) and B, not using the LOCPRIOR option.

Fig. S4. Principal components analysis of two landmarks (leaf base and leaf apex) and 26 semi-landmarks (along the leaf contour) from 152 specimens of the *Quercus laeta* complex. **A**, Plot grouping individuals according to morphotypes; **B**, plot grouping individuals according to the NJ tree; **C**, plot with the DAPC groups; **D**, plot grouping individuals according to the STRUCTURE analysis.

III. CAPÍTULO II

Phylogenomics and ecological niche contrasts lead to the identification of multiple and independent evolutionary lineages within the *Quercus laeta* complex (Fagaceae): new insights into the diversification of the *Leucomexicana* clade (*In prep*.)

Phylogenomics and Ecological Niche Contrasts Lead to the Identification of Multiple and Independent Evolutionary Lineages within the *Quercus laeta* complex (Fagaceae): New Insights into the Diversification of the *Leucomexicana* subsection

Saddan Morales-Saldaña[1], Andrew L. Hipp[2,3], Marlene Hahn[2], Susana Valencia-Ávalos[4], M. Socorro González-Elizondo[5], David S. Gernandt[6], Kasey K. Pham[7], Ken Oyama[8], and Antonio González-Rodríguez[1*]

[1]*Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Morelia, 58190 Michoacán, México*

[2]*The Morton Arboretum, Lisle, IL 60532-1293, USA*

[3] *The Field Museum, Chicago, IL 60605, USA*

[4]*Herbario de la Facultad de Ciencias, Departamento de Biología Comparada, Universidad Nacional Autónoma de México, Circuito Exterior, s.n., Ciudad Universitaria, Coyoacán, México City CP 04510, México*

[5]*CIIDIR Unidad Durango, Instituto Politécnico Nacional, Durango, México*

[6]*Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, 04510*

Ciudad de México, Mexico

[7] *Department of Biology, University of Florida, Gainesville, Florida 32611, USA*

[8]*Escuela Nacional de Estudios Superiores Unidad Morelia, Universidad Nacional Autónoma de México (UNAM), Antigua Carretera a Pátzcuaro No. 8701, Col. Ex-Hacienda de San José de la Huerta, Morelia, Michoacán 58190, México*

[*] Author for correspondence (agrodrig@cieco.unam.mx)

Abstract

Multispecies introgressive hybridization poses a challenge to taxonomic and phylogenetic understanding of taxa with high numbers of co-occurring, intercrossable species. The oaks, genus *Quercus* (Fagaceae), exemplify this situation. They are highly diverse in sympatry and cross freely, creating syngameons of interfertile species. Moreover, they are notoriously variable in morphology and have large effective population sizes and high heterozygosity. Their taxonomic and phylogenetic complications as a consequence derive not only from introgression, but also from plasticity and incomplete lineage sorting. Although a well-resolved, dated phylogeny is available for the monophyletic American oaks, evolutionary relationships within many of the more recently derived clades remain to be defined, particularly for the young and exceptionally diverse Mexican white oak clade. Here, we used ecological data and sequences of 155 low-copy nuclear genes to identify distinct lineages within the *Quercus laeta* complex of central and northern Mexico. We analyzed the concatenated molecular data and used coalescent analysis of gene trees to assess the phylogenetic placement of these lineages relative to other oak species in the Mexican white oak clade. We used phylogenetic network methods that account for both incomplete lineage sorting and introgression to evaluate the timing and genomic significance of recent or historical introgression among lineages. These analyses support three major conclusions. First, the *Quercus laeta* complex is polyphyletic, explaining the contradictory resolutions for putative taxon in previous phylogenomic studies; second, the *Quercus laeta* complex comprises six well-supported lineages, each restricted geographically and third, historical introgression and niche divergence in the context of the whole suite of sympatric Mexican white oak species have shaped the diversification of the *Quercus laeta* complex. The *Quercus laeta* complex is, in other words, a morphological

and ecologically interconnected group of species rather than a clade, and one that connects reproductively to species from across the Mexican white oak syngameon. Our work demonstrates the importance of both ecological divergence and reticulate gene flow on woody plant diversity across the temperate / tropical Mexican forests.

Key words: Coalescence, Introgression, Species Complex, Phylogenetic Network, *Quercus* subsection *Leucomexicana*, Target Enrichment

Typically, phylogenetic relationships are graphically represented by bifurcating trees, which suggest a split of population-level lineages that diverged and remained independent (Linder et al. 2003; Rancilhac et al. 2021). However, hybridization is prevalent and widespread and has played an important role in diversification across diverse lineages of the tree of life (Abbott et al. 2013; Mallet et al. 2016; Solís-Lemus and Ané 2016; Gernandt et al. 2018; Crowl et al. 2020; Kleinkopf et al. 2019; Taylor and Larson 2019; Rancilhac et al. 2021). Phylogenetic inference in the presence of hybridization, introgression, and incomplete lineage sorting requires modelling evolutionary relationships as networks, i.e., phylogenetic trees with interconnected branches, where interconnections represent gene flow in the context of an otherwise divergent process (Maddison and Knowles 2006; Wen et al. 2008; Solís-Lemus and Ané 2016; McVay et al. 2017a; Crowl et al. 2020; Rancilhac et al. 2021). Inferring these networks from individual gene trees and Single Nucleotide Polymorphisms (SNP) variation is a major challenge for phylogenetics.

The genus *Quercus* (the oaks) is renowned for the facility of species to exchange genes within taxonomic sections, almost irrespective of the time elapsed since their divergence from a common ancestor within that section (Palmer 1948; Hardin 1975), exemplifying divergence in the face of ongoing gene flow (Belahbib et al. 2001; Eaton et al. 2015; Sullivan et al. 2016; Li et al. 2021). Partly because of this, early attempts to reconstruct the oak phylogeny based on chloroplast DNA and limited numbers of nuclear loci faced limited success (Manos et al.1999; Mayol and Roselló 2001; Bellarosa et al. 2005; Simeone et al. 2013; Hubert et al. 2014). More recently, analyses based on restriction-site associated DNA sequencing (RAD-seq) data have been useful in reconstructing the relationships and patterns of diversification of the main oak clades (Hipp et al. 2014; Cavender-Bares et al. 2015; Hipp et al. 2018; Hipp et al. 2020). Although these

studies have provided a solid, well-resolved and dated phylogeny for the monophyletic American oak clade, they also demonstrated that the evolutionary relationships within some groups remain to be defined, particularly for Mexican species (Valencia-A 2004; Hipp et al. 2018; Hipp et al. 2020). These studies have recognized the need to use a large number of loci across the entire genome to accurately assess the evolutionary relationships of closely related species (Hipp et al. 2014; Hipp et al. 2018; Crowl et al. 2020; Hipp et al. 2020) and have pointed to introgression events that have yet to be explored across the tree using dense sampling of individuals analyzed using phylogenomic methods.

RAD-seq phylogenetic studies are constrained in part by the relatively short loci that they recover (Ree and Hipp 2015). They have enabled targeted studies of oak introgression hypotheses (e.g., Eaton et al. 2015; McVay et al. 2017a, b) but not yet taken advantage of the full power of gene-tree based methods for reconstructing phylogenetic networks. By contrast with RAD-seq, targeted enrichment methods can provide both a large number of low copy nuclear loci capable of resolving conflicting phylogenetic relationships of non-model organisms, making the inference of reticulate phylogenetic histories against the background of incomplete lineage sorting more straightforward (Weitemier et al. 2014; Buddenhagen et al. 2016; Gernandt et al. 2018; Folk et al. 2017; Vatanparast et al. 2018; Villaverde et al. 2018; Dodsworth et al. 2019; Loiseau et al. 2019; Hale et al. 2020; Breinholt et al. 2021; Ma et al. 2021). These methods use small RNA or DNA oligonucleotides (called “baits” or “probes”) to enrich next-generation sequencing libraries for selected target loci, such that the DNA for target loci is preferentially sequenced while much of the non-target DNA is discarded (Weitemier et al. 2014; Breinholt et al. 2021). Target enrichment has been shown to be cost-effective, efficient in high throughput sequencing contexts, and useful with historical specimens such as herbarium samples

(Gardner et al. 2016; Villaverde et al. 2018; Kleinkopf et al. 2019; Hale et al. 2020; Breinholt et al. 2021).

Understanding phylogenetic networks separate from ecology, however, provides little insight into the processes shaping biodiversity, as each species is the realization of a niche (Van Valen 1976). Ecological niche models (ENMs) have proven to be complementary to phylogenomic approaches in addressing questions such as whether recently diverged taxa that show limited genetic differences are nonetheless ecologically distinct, supporting the idea that they are separately evolving lineages (Rissler and Apodaca 2007; Su et al. 2015; Nunes and Pearson 2017; Lin et al. 2021), or to examine the roles that ecological differences have played in divergence and speciation (Glor and Warren 2010; Wooten and Gibbs 2012; Blair et al. 2013; Gutiérrez-Ortega et al. 2020; Calixto-Rojas et al. 2021; Lin et al. 2021). Although most niche models are constructed at the species level, niche conservatism operates at the macroevolutionary level, while local adaptation operates at the population level, within species (Smith et al. 2019). In cases where niche evolution is rapid or local adaptation is suspected, creating separate niche models for lineages within species may best capture environmental relationships (Smith et al. 2019; Gutiérrez-Ortega et al. 2020). If these lineages occupy different regions of environmental space, niche divergence may be shaping the future of biodiversity (Goudarzi et al. 2019).

In this study, we employed an approach bridging micro- and macroevolutionary scales to resolve evolutionary relationships in the *Quercus laeta* complex, a taxonomically problematic group within *Quercus* subsect. *Leucomexicana*, a rapidlydiversifying clade endemic to Mexico, Central America, and the southwestern U.S. (Hipp et al. 2018; Manos and Hipp 2021). The *Q. laeta* complex is endemic to Mexico and widely distributed in a broad variety of environmental conditions. It is considered one of the most polymorphic

Mexican oak groups (Valencia-A, 2004), and as many has six taxa—*Quercus bipedalis* Trel., *Quercus centralis* Trel., *Quercus obscura* Trel., *Quercus pallescens* Trel., *Quercus prinopsis* Trel. and *Quercus transmontana* Trel.— have been variously teased out of or synonymized under *Q. laeta* (McVaugh 1974; Romero et al. 2002). Previous taxonomic work in the complex has been based mainly on micromorphological characters such as papillae and glands on the leaf undersides, and the conformation of trichomes on the upper surfaces of the leaves (Liebmann 1854; Trelease 1924; Morales-Saldaña et al. 2022). More recently, phylogenetic and population-level studies have suggested the existence of at least two different entities within *Q. laeta* that may constitute separate species (Hipp et al. 2020, Morales-Saldaña et al. 2022). Understanding the evolutionary history of this complex is clearly essential to understanding its taxonomy and its role in the ecology of Mexican woodlands and savannas.

Here, we assessed evolutionary relationships among populations within the *Q. laeta* complex by combining ecological niche comparisons with phylogenetic inference based on target enrichment sequencing data. Our specific objectives were to: 1) infer relationships among populations in the *Q. laeta* complex; 2) assess the phylogenetic placement of population lineages relative to other oak species in the *Leucomexicana* subsect.; 3) infer recent or historical introgression in this group and among the *Q. laeta* complex and sympatric white oak species; and 4) estimate the importance of environmental differentiation in the diversification of the retrieved oak lineages.

MATERIALS AND METHODS

Taxon Sampling

A data set of 56 samples grouped into two subsets was used (online Appendix 1, Fig. 1). The first (subset 1) corresponded to 36 samples collected for this study from 12

populations of the *Q. laeta* complex throughout its range in Mexico. The second (subset 2) consisted of 20 samples comprising 13 species from the *Leucomexicana* clade that have been identified as closely related to the *Q. laeta* complex in previous phylogenomic studies (Hipp et al., 2018; Hipp et al., 2020). The latter 20 samples were obtained from field work performed for the present study and from previous collections of the second and senior authors. Vouchers specimens were deposited at the Herbarium of the Facultad de Ciencias of the Universidad Nacional Autónoma de México (FCME). Geographical information and voucher details for the 56 samples are shown in online Appendix 1.

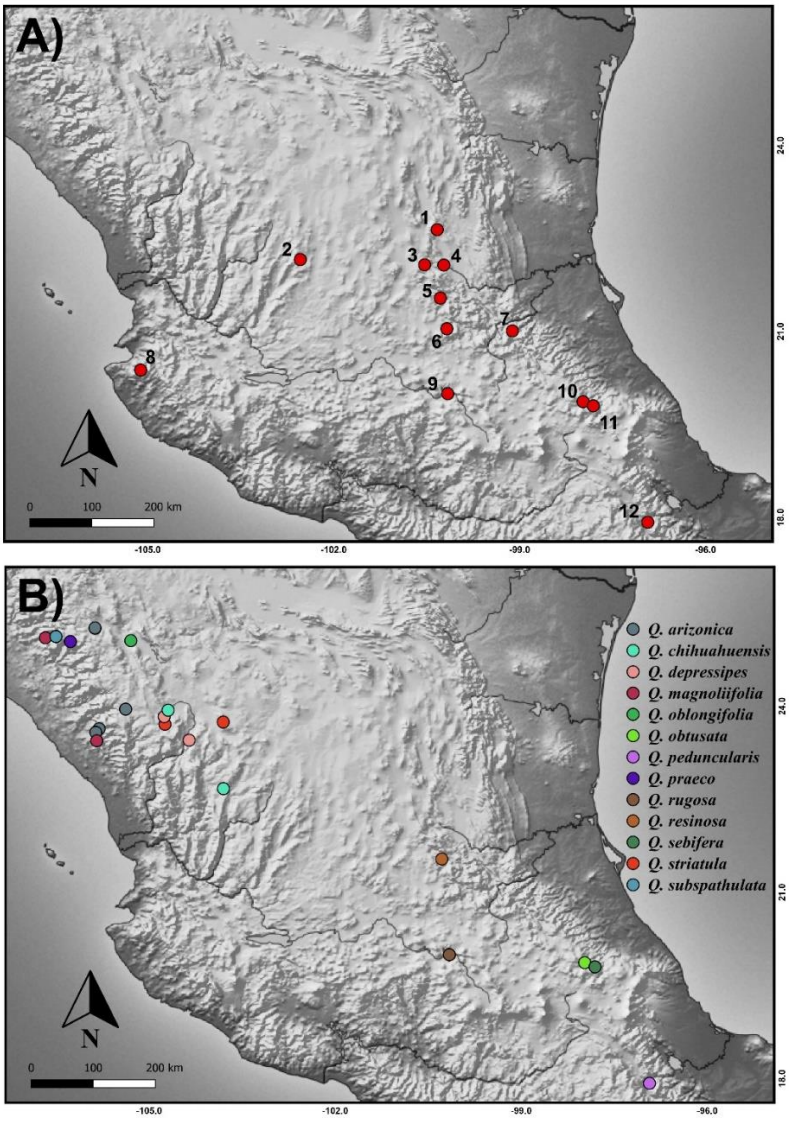


Figure 1. Geographical distribution of the different samples collected. a) Populations for the *Quercus laeta* complex (subset 1), 1: Guadalcazar, 2: La Congoja, 3: San Nicolás Tolentino, 4: San José Gallinas, 5: Vergel de Bernalejo, 6: Cieneguilla, 7: Silao, 8: Sierra El Cuale, 9: Contepec, 10: Waterfalls of Tuliman, 11: Tetela de Ocampo, 12: Coyula*;* b) Localities collected for species from the subsection *Leucomexicana* (subset 2).

DNA Extraction, Library Preparation and Sequencing

Genomic DNA was extracted from fresh leaves using the cetyltrimethyl ammonium bromide (CTAB) protocol with an additional phenol-chloroform cleaning step (Lefort and Douglas 1999). We checked the DNA concentration of each extraction with a Qubit fluorometer 4.0 (Thermo Fisher Scientific) using a high-sensitivity kit and with a Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA). Samples with ≥ 200 ng of DNA and an A260/A280 between 1.8 and 2.2 were carried on to the next step. Isolated genomic DNA was sonicated to obtain fragments of size > 550 pb using a Covaris E220 Focused-ultrasonicator (Wohurn, MA, USA) with Covaris microtubes.

Library preparation was done with the KAPA Hyper Prep Kit (KAPA biosystems Wilmington, MA, USA) following the manufacturer's protocol. Library concentrations were quantified using a Qubit fluorometer 4.0 (Thermo Fisher Scientific) with a high-sensitivity kit. Solution-based hybridization and enrichment was carried out through the MYBaits target enrichment system (Arbor Biosciences, Ann Arbor, Michigan) following the standard MYBaits version 5.02 protocol (https://arborbiosci.com/mybaits-manual/). Libraries were pooled according to phylogenetic proximity, following previous phylogenomic studies (Hipp et al. 2018). We used custom baits designed for oaks by Pham et al. (in prep.; published for the first time here) that target 465 putatively single copy genes. The target-enriched libraries were sequenced on an Illumina MiSeq at The Field

Museum of Natural History (Chicago, IL, United States) using MiSeq reagent kit v3 (600 cycles, 2 × 300 bp paired-end reads).

Gene Assembly, Alignment, and Filtering

We used FastQC version 0.11.8 (Andrews, 2010) to assess the quality of Illumina raw. Multiplexed reads were separated using their barcode sequences and combined into paired fastq files (R1 and R2) for each sample individually. We then used Trimmomatic v. 0.36 (Bolger et al., 2014) to remove adapter sequences and low-quality reads using a 4 bp sliding window, a quality threshold of Q30, and a minimum sequence length of 30. Consensus target sequences were assembled using HybPiper ver. 1.2 (Johnson et al. 2016), which defaults to ignoring heterozygotic positions generating a single consensus sequence per individual, with potentially heterozygous bases called as the nucleotide with the highest read frequency. HybPiper used BWA v. 0.7.1 (Li and Durbin 2009) to align reads to the reference nuclear gene sequences and SAMtools v. 0.1.19 (Li et al. 2009) was used to sort the reads into separate directories for each gene. Subsequently, reads mapped for each locus were de novo assembled into contigs with the best k-mer automatically detected by SPAdes version 3.10.1 (Bankevich et al. 2012). HybPiper was use to output the exon regions for each gene separately.

The resulting gene files were imported to Geneious v. 8.1.9 and aligned individually with MAFFT version 7.3 (Katoh et al. 2002; Katoh and Standley 2013). Following Gernandt et al. (2018) and Villaverde et al. (2018), after a visual inspection, we proposed the following *ad hoc* criteria to identify and exclude alignments with poor quality: i) alignment length < 500 pb, ii) <20% of sites identical, iii) pairwise identity <90% and iv) genes detected as paralogs. The percentage of identical sites and pairwise identity statistics were calculated in the Geneious alignment view (Kearse et al. 2012); paralogs were

detected using HybPiper version 1.2 (Johnson et al., 2016).

Inference of Phylogenetic Relationships and Population Structure within the Quercus laeta *complex*

First, we inferred phylogenetic relationships at the population level within the *Q. laeta* complex and estimated boundaries among hypothetical lineages. For this, we evaluated phylogenetic relationships by maximum likelihood (ML) including only samples belonging to the *Q. laeta* complex (subset 1; see online Appendix 1). For ML analysis, all 155 loci were concatenated into a super-matrix and a phylogenetic tree was inferred in RAxML v.8.2.9 (Stamatakis 2014), using the GTRCAT implementation of the general time reversible model that affords substantial savings in computational time for large phylogenetic data sets, with branch support assessed using 1000 fast bootstraps. Additionally, to assess the genetic distinctiveness and cohesion of groups within the *Q. laeta* complex, a non-metric multidimensional scaling (NMDS) clustering approach was applied on a pairwise genetic distance matrix for *Quercus laeta* samples. The distance matrix for NMDS was calculated under the GTRGAMMA model in RAxML v.8.2.9 (Stamatakis 2014), then converted from long- to wide-format for use in the monoMDS function of the R package vegan v. 2.5–2 (Oksanen et al. 2018). NMDS ordination was conducted using default parameters with a maximum number of random starts in search of stable solution of 1000 (maxit = 1000) from $K = 1$ through $K = 5$.

Inference of Phylogenetic Relationships among the Quercus laeta *Lineages and other Species from the Leucomexicana subsection*

Once different lineages were identified within the *Q. laeta* complex, the next step was to evaluate if these lineages are more closely related to each other than to other species of the subsect. *Leucomexicana*. For this purpose, we evaluated phylogenetic relationships by

both concatenation and multispecies coalescent (MSC) approaches to compare differences in topologies reconstructed under differents assumptions. For the concatenation analysis, we conducted maximum likelihood (ML) analysis in RAxML v 8.2.9 (Stamatakis 2014), using the GTRCAT approximation with 1000 fast bootstraps to evaluate branch support.

Because concatenation approaches can produce erroneous tree topologies in the presence of incomplete lineage sorting (ILS, Liu et al. 2015, Edwards et al. 2016), potentially producing high support for the incorrect tree (Bapteste et al. 2008, Warnow 2015), we used two coalescent-based species tree methods that accommodate ILS, ASTRAL-III v5.7.8 (Zhang et al. 2018) and SVDQuartets (Chifman and Kubatko 2014) as implemented in PAUP* v4.0a167 (Swofford 2002). Both ASTRAL-III and SVDQuartets allow analysis of multiple individuals per species, while estimating the relationship amongst species, not individuals. We assigned individuals based on i) the identity of their corresponding taxonomic species and ii) the results of the concatenated phylogenetic analyses of the section 2.4. For the ASTRAL-III analysis, a single gene tree was reconstructed for each locus by maximum likelihood in RAxML using the GTRCAT model with 1000 fast bootstrapping. The 155 gene trees reconstructed were used as input for the species tree reconstruction in ASTRAL-III version 5.7.3 (Zhang et al. 2018). Finally, for each branch in the ASTRAL species tree, we recovered the local posterior probability (LPP) support (Sayyari and Mirarab 2016). For the SVDQuartets analysis, PAUP version 4.0a169 (Swofford 2002) was applied to the concatenated matrix with 100 bootstrap replicates for branch support and all possible quartets evaluated. Both the ML and MSC trees were displayed graphically in Figtree v. 1.4.0 (Rambaut 2012).

Assessing Incongruence among Gene Trees and Species Trees

Bootstrap support may be high even when a low number of genes support a clade

(Minh et al. 2020; Pease et al. 2018). Therefore, we examined gene and site concordance factor (gCF and sCF, respectively) to quantify topological conflict around each branch of the ML concatenated tree in IQ-TREE-2 version 2.1.2 (Minh et al. 2020). For every branch of the concatenated tree, the gCF and sCF represents the percentage of decisive gene trees and alignment sites, respectively, containing that branch (Minh et al. 2020). In addition, the ASTRAL quartet score and the normalized quartet score were computed to summarize the proportion of induced quartet trees (from individual single-locus gene trees) in the ASTRAL species tree. Also, we recovered the quartet support (QS; Sayyari and Mirarab 2016) which measures the conflict of gene around each branch of the ASTRAL species tree (Zhang et al. 2018)

Inferring Ancient Introgression

We used the maximum pseudolikelihood method Species Networks applying Quartets (SNaQ) (Solís-Lemus and Ané 2016) to estimate a phylogenetic network among the Mexican white oak species sampled, allowing for both hybridization and incomplete lineage sorting (ILS). To reduce the computational burden of estimating a large network (> 30 taxa) and account for sensitivity of SNaQ to taxon sampling, we analyzed four datasets that reflect a wide range of possible reticulation events. We selected for our analysis samples that showed conflicting phylogenetic signals in the concatenated analysis and in the species trees. Networks were inferred with the SNaQ method implemented in the Julia v1.7.2 (Bezanson et al. 2017) package PhyloNetworks v0.11.0 (Solís-Lemus and Ané 2016; Solís-Lemus et al. 2017), based on concordance factors from 155-gene-tree sets for each of the four taxon samples. We used RAXML v.8 (Stamatakis 2014) to obtain gene trees with bootstrap support for each gene (with the GTR+GAMMA model and 100 bootstrap replicates) and used these trees to estimate quartet concordance factors (CFs) with

PhyloNetworks (function readTrees2CF). These CFs are then used to reconstruct phylogenetic networks under incomplete lineage sorting (ILS). Using the species tree obtained with ASTRAL-III as a starting tree, we then estimated the best phylogenetic network with varying number of hybridization events (h) allowed ($0 \leq h \leq 3$). To ensure convergence, we performed 10 independent runs under each value of h. Pseudolikelihood scores for each value of h were plotted in R v.3.4.4 (R Development Core Team 2013), and the best network model was selected by examining at what value of h the pseudolikelihood score plateaus, as recommended by Solís-Lemus et al. (2017).

Niche Differentiation among Lineages within the Quercus laeta *complex*

Localities for the ecological niche models (ENM) for each of the identified lineages in the *Q. laeta* complex were based on the 14 populations included in this study (see online Appendix 1), combined with records obtained from the careful revision of the specimens conducted by the authors present in the collections of the Herbarium of Facultad de Ciencias, UNAM (FCME) and the National Herbarium of Mexico (MEXU). We did not consider records from online databases, because visual inspection of specimens was necessary to differentiate among the lineages here identified. Because the spatial correlation of the localities may cause model overfitting (Dormann et al. 2007; Dormann et al. 2013), specimen records were thinned using the spThin package in R (Aiello-Lammens et al. 2015), which randomly selects points of presence with at least 5 km distance among locations using 100 replicas. In the end, we obtained 56 presence data points partitioned into the *centralis TMVBc* (7), *centralis TMVBe* (7), *laeta SMS* (6), *laeta SMOc* (10), *prinopsis* (17) and *transmontana* (9) lineages. While this sampling is low, several studies have revealed that as few as five localities can produce biologically meaningful models (Pearson et al. 2007; Shcheglovitova and Anderson 2013; Galante et al. 2018).

The accessibility area "M" was defined based on the shapefile of ecoregions from the World Wildlife Fund (Olson et al. 2001), selecting those ecoregions where at least one point of occurrence was found. We used the 19 bioclimatic variables obtained from the WorldClim database at 30 arc-second spatial resolution (~1 km^2; Fick and Hijmans 2017; available at https://worldclim.org/data/worldclim21.html). For selection of the bioclimatic variables, we explored three data sets based on different criteria: (Set 1) removal of one variable from each pair of variables for which Pearson product-moment correlations were high ($r \geq 0.8$); (Set 2) identification of variables that contribute most strongly to models using the jackknife in MaxEnt, followed by removal of one variable from each highly correlated pair of variables ($r > 0.80$); (Set 3) inclusion of only variables with variance inflation factor (VIF) values < 10 (following Brauner and Shacham 1998; Guisan et al. 2002).

Model calibration consisted of evaluating candidate models created with 13 distinct regularization multipliers (0.1 to 1.3 at intervals of 0.1), different combinations of three feature classes (Linear; Linear and Quadratic, Linear, Quadratic and Hinge), and the three sets of environmental variables (set1, set2, set3) with distinct number of variables each. Best parameter settings were selected considering statistical significance (partial ROC; Peterson et al. 2008), predictive power (omission rates E = 5%; Anderson et al., 2003), and complexity level (AICc; Warren et al. 2010), in that order (Cobos et al. 2019). Final models were generated with 10 bootstrap replicates, using Maxent version 3.4.1 (Phillips et al. 2006) and the results of model calibration obtained from application of the kuenm R package (Cobos et al. 2019).

To test the degree of climate niche differentiation among lineages, we used the niche similarity test implemented in ENM Tools (Warren et al. 2008, 2010). This test employs

randomization to compute a null distribution and estimate whether the ENMs from different lineages are like each other than expected by chance, based on environmental differences in the environment in which they occur. The test was run using the "background.test" function in R package "ENMTools" (Warren et al. 2010) with 100 replicates, sampling a total of 30,000 random points and was analyzed with a two-tailed test (Warren et al. 2008). Niche overlap among lineages was measured via Schoener's *D* and a modified Hellinger's *I* metric (Warren et al. 2008).

RESULTS

Targeted Enrichment

A total of 465 genes from 465 targets were assembled successfully for at least one sample. From the 465 genes assembled, four genes (0.86%) showed alignment length < 500 pb, 16 genes (3.44%) were identified as paralogs, 93 genes (20%) showed pairwise identity <90% and, 197 genes (42.36%) showed <20% of sites identical so that three-hundred and ten loci were removed. The remaining 155 genes were carried forward for subsequent analysis. The largest aligned locus was 4689 bp and the shortest was 606 bp. The aligned length of the concatenated 56-taxon and 155-loci super-matrix was 302,321 bp with 8078 parsimony informative sites and 23.02% missing data

Inference of Phylogenetic Relationships and Population Structure within the Quercus laeta *Complex*

The ML analysis recovered up to six well-supported lineages. Lineage designations were based on the synonyms of *Q. laeta* and on the geographical regions where these lineages were sampled. Accordingly, these six lineages were called *centralis TMVBc* (Trans-Mexican Volcanic Belt central) (BS = 100), *centralis TMVBe* (Trans-Mexican Volcanic Belt eastern) (BS = 99), *laeta SMS* (Sierra Madre del Sur) (BS = 91), *laeta SMOc*

(Sierra Madre Occidental) (BS = 99), *prinopsis* (BS = 96) and *transmontana* (BS = 96) (Fig. 2). In the tree, bootstrap support was high (BS > 90) for both the backbone and the relationships among the six major lineages (Fig. 2).

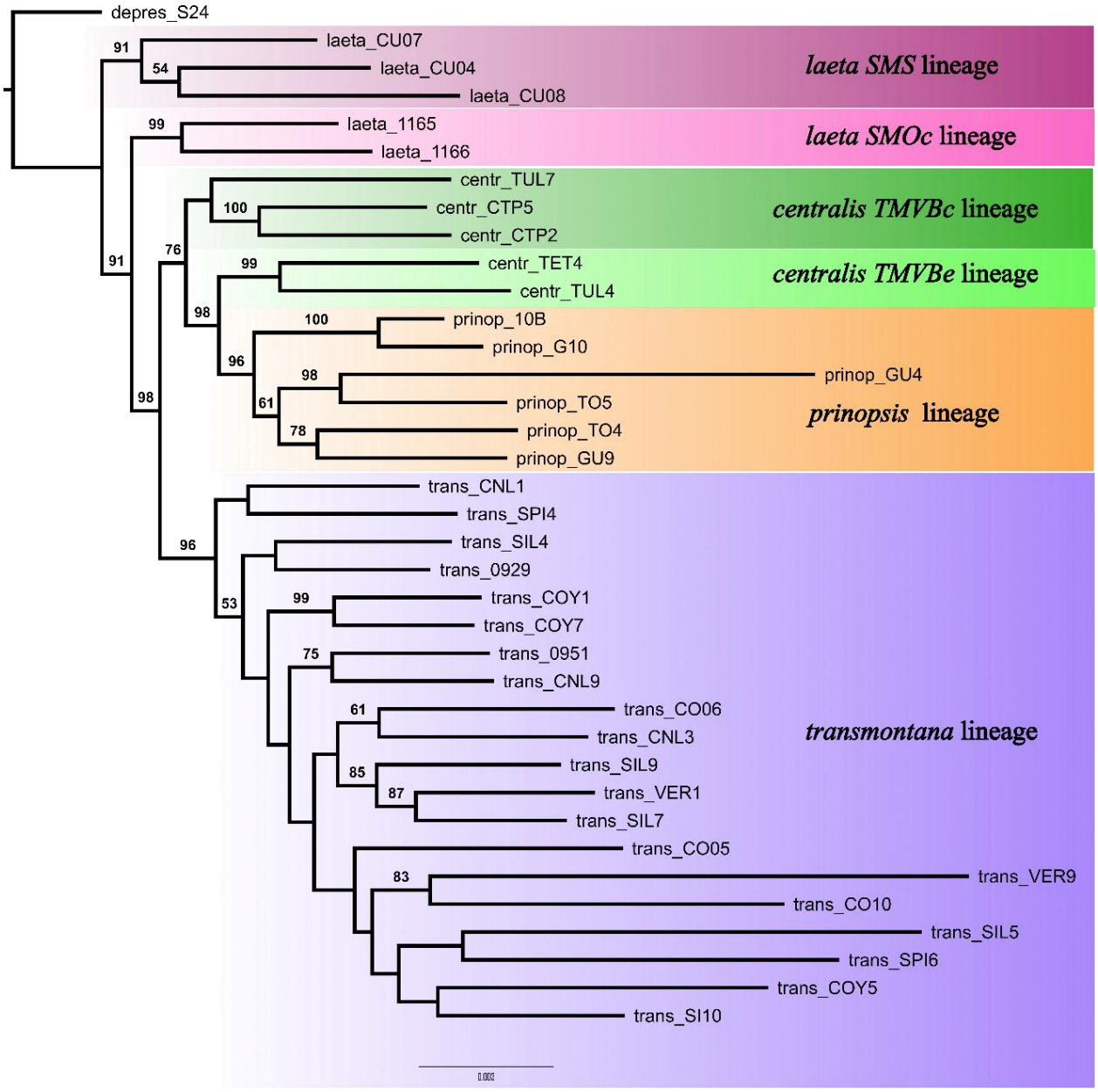


Figure 2. Tree based on maximum likelihood (155 loci: 302,321 pb) with the GTRCAT model and 1000 bootstraps. Bootstrap support values >50% are indicated above the branches. Colors and labels to the right of the tree indicate the different lineages proposed.

The NMDS genetic ordination (Fig. 3) showed that the stress values decrease rapidly from *K* = 1 to 2 to 3 (stress = 0.254, 0.162 and 0.118 respectively), showing decreases of less than 0.028 per additional dimension after that, indicating little added information for more than three dimensions. These three NMDS axes (Fig. 3 as two 2-dimensional plots) showed a clear separation of the *transmontana* lineage from *centralis*, *laeta* and *prinopsis* lineages, though with weaker separation among the other groups (Fig 3).

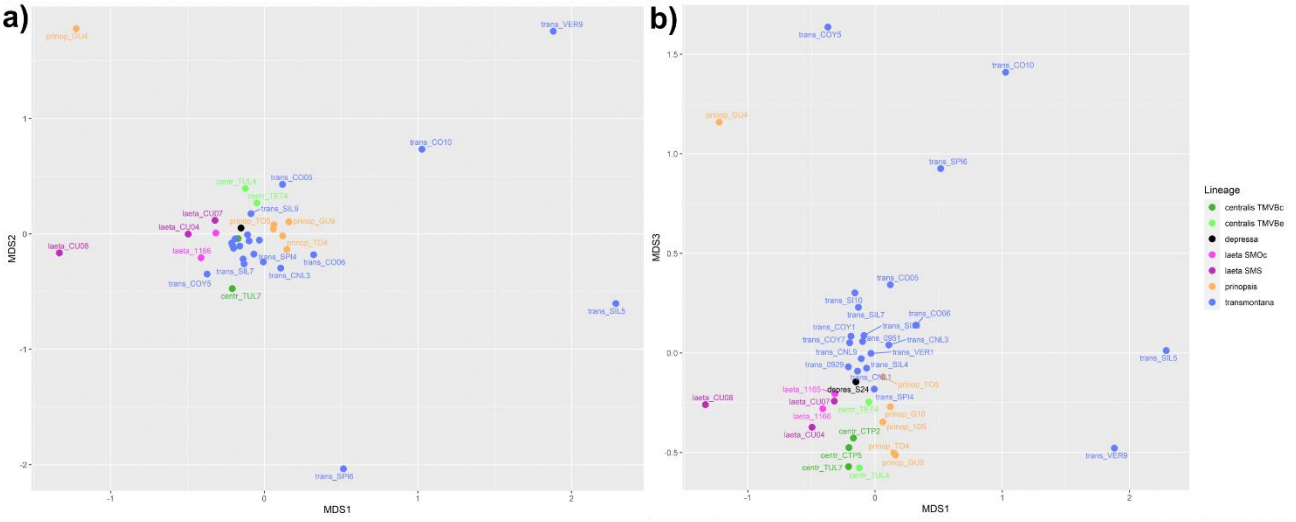


Figure. 3. Ordination of *Quercus laeta* individuals sequenced, based on genetic data. Ordination is based on a three-dimensional non-metric multidimensional scaling (NMDS) analysis of the pairwise genetic distance estimated using the GTRCAT nucleotide substitution model in RAxML.

Phylogenetic Relationships among Lineages of the Q. laeta *Complex and other Leucomexicana Species.*

The trees recovered by both ML and MSC suggest that the *Q. laeta* complex is polyphyletic (Fig. 4; Fig. 5 a-b). Furthermore, all methods support *Q. sebifera* as sister to *Q. rugosa* and a clade containing *Q. magnoliifolia*, *Q. resinosa*, and *Q. subspathulata* with high bootstrap support.

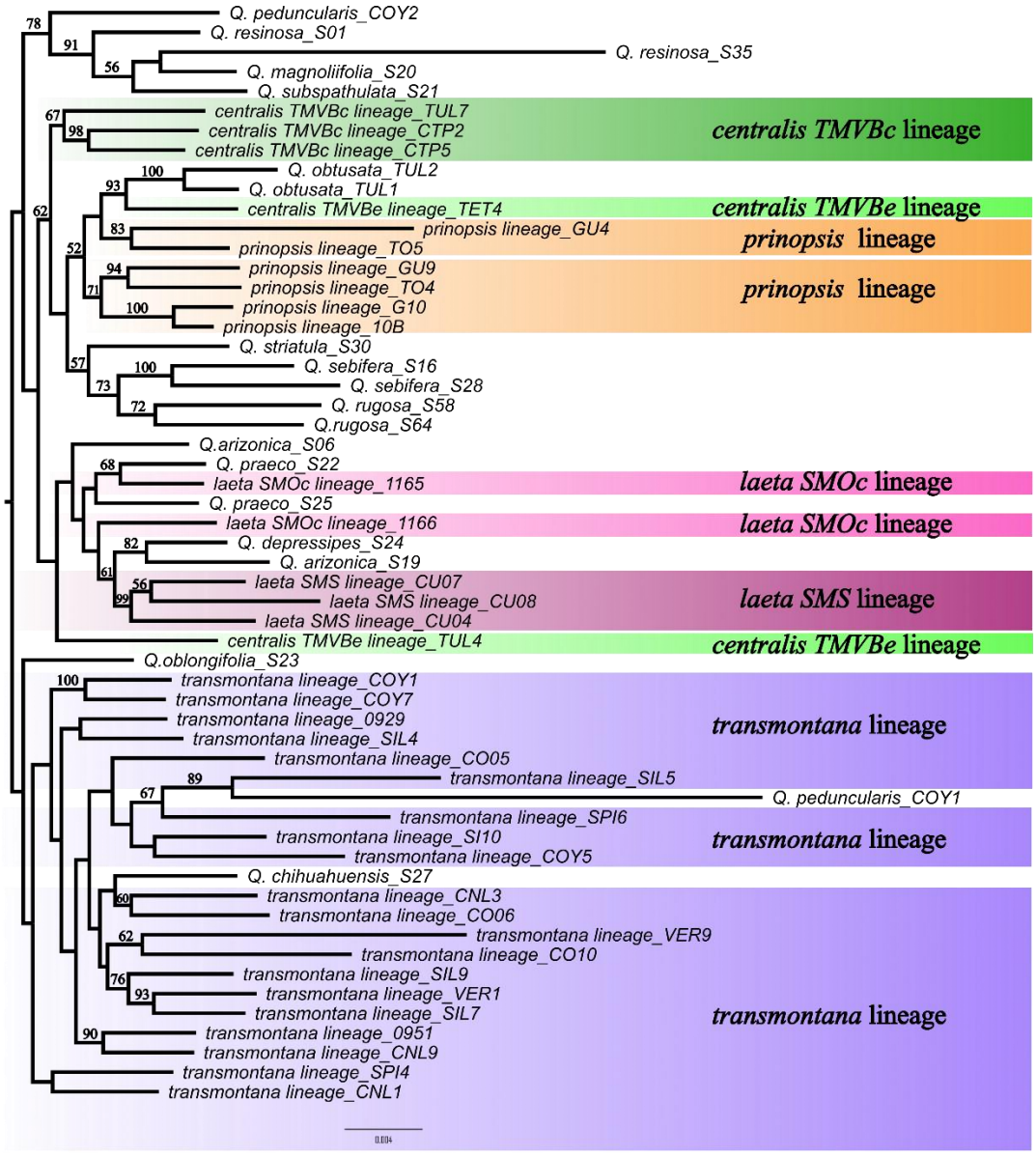


Figure 4. Concatenated ML tree inferred in RAxML. with concatenated matrix (155 loci; 302,321 pb). Bootstrap support values >50% are indicated above the branches.

ML analysis of the concatenated matrix under the GTRCAT substitution model

yielded a topology where *Q. chihuahuensis* and one sample of *Q. peduncularis* (pedun_Coy1) were placed in a clade with *transmontana* individuals, and this clade was sister to *Q. oblongifolia*. In turn, the other sample of *Q. peduncularis* (pedun_Coy2) was sister to a clade containing *Q. resinosa*, *Q. magnoliifolia*, and *Q. subspathulata* (BS = 78) (Fig. 4). The samples of *laeta* SMS were supported as monophyletic (BS = 99) and more closely related to the clade of *Q. depressipes – Q. arizonica* (BS = 61) than to *laeta* SMOc individuals, with one of these (laeta_1165) being placed with *Q. praeco* (BS = 68). In addition, *prinopsis* samples formed a clade (BS = 71) except for two samples that were sister to a clade containing *Q. obtusata* and one sample of *centralis* TMVBe (centr_TET4) (BS = 49). Finally, *centralis* TMVBc individuals were recovered as monophyletic (BS = 98) and placed sister to one sample of *centralis* TMVBe (centr_TUL7) (BS = 67) (Fig. 4).

Gene and site concordance factors calculated for the ML concatenated tree were generally low, indicating poor concordance among gene trees despite relatively high bootstrap values in shallow nodes of the tree. Gene concordance factors (gCF) ranged from 0% to 18.39% across the data set (online Appendix 2a), indicating that few gene trees contain the branches present in the ML tree. In contrast, site concordance factors (sCF) ranged from 29.57 % to 69.57 % with sCF values close to the neutral value (~ 33% = no informative decisive sites; Minh et al. 2020) mainly in the deeper nodes, pointing to a lack of a clear signal at this level (online Appendix 2a). A comparison between bootstrap, gCF and sCF values showed that low bootstrap values coincide with the lowest gCF and sCF values, but also revealed that high bootstrap values occurred with low gCF and sCF values, illustrating that bootstrap values do not entirely capture the variation in the underlying data (online Appendix 2b).

In the SVDQuartets tree (Fig 5a), individuals of *transmontana* formed a clade sister to *Q. chihuahuensis* (BS = 40), while *Q. peduncularis* was sister to a *centralis* TMVBc clade (BS = 73). Furthermore, a *laeta* SMOc clade was strongly supported as sister to *Q. praeco* (BS = 99) and in turn, this clade was sister to *laeta* SMS (BS = 57). As found in the ML analysis of the concatenated matrix, *centralis* TMVBe was sister to *Q. obtusata*, although with a low support (BS = 20), and *prinopsis* was recovered as sister to the *Q. sebifera – Q. rugosa* clade (BS = 58).

The ASTRAL species tree recovered relationships partially consistent with the results from the SVDQuartets tree (Fig. 5b). The *transmontana* clade was sister to *Q. chihuahuensis* (LPP = 0.41), and this clade in turn was sister to *Q. arizonica* (LPP= 0.29). The *laeta* SMOc clade was recovered as sister to *Q. praeco* (LPP = 0.67) and in turn this clade was sister to *laeta* SMS (LPP = 0.54). In contrast to the SVDQuartets tree, *prinopsis* was placed sister to *Q. obtusata* (LPP = 0.40) and this clade was sister to *centralis TMVBe* (LPP = 0.43). Like in the concatenated analysis, *Q. peduncularis* was more closely related to a clade containing *Q. magnoliifolia*, *Q. resinosa,* and *Q. subspathulata* than with the *centralis TMVBc* lineage.

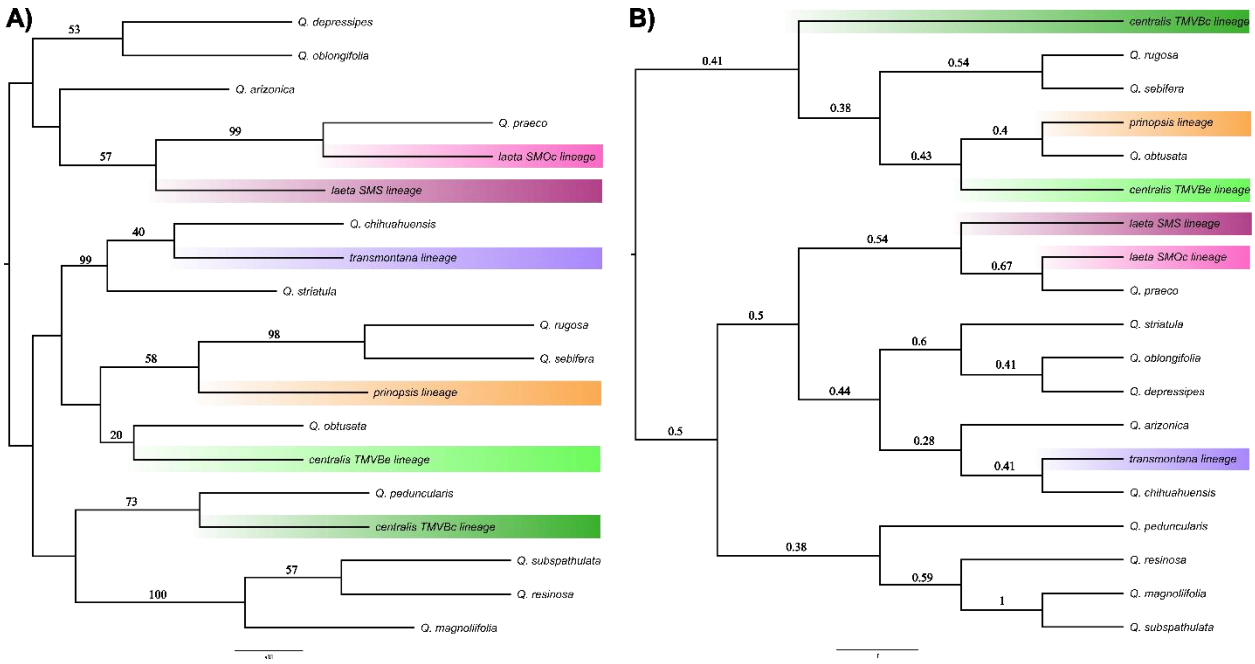


Figure 5. Phylogenetic hypotheses of relationships among the different lineages of the *Quercus laeta* complex and other species of the *Leucomexicana* clade. a) Species tree obtained by SVDQUARTETS. Bootstrap support values are indicated above the branches b) Species tree obtained by ASTRAL III. Local posterior probability values expressed as percentages are shown above branches.

We calculated quartet scores from ASTRAL, which are a measure of gene tree discordance that can be used as a proxy for ILS (Singh et al. 2022). The quartet score and the normalized quartet score were 7955760 and 0.38 respectively, suggesting that the level of ILS in the species tree was high. Further, the quartet support (QS) computed for all nodes in the species tree showed high levels of gene tree discordance for many nodes, where branch-specific quartet scores ranged from 32 to 100% (online Appendix 2c).

Evidence of Reticulation in the Q. laeta *Complex*

Phylogenetic networks estimated to ascertain whether introgression occurred over the evolutionary history of *Quercus laeta* lineages pinpointed introgression events in each of

the four datasets (online Appendix 3). Based on the slope heuristic, SNaQ analysis favored networks with a single reticulation edge (h1) for all datasets. In all datasets the phylogenetic networks recovered the *Q. laeta* complex as polyphyletic (Fig. 6). Trees recovered by specifying zero reticulation events were consistent with those recovered using MSC approaches. The four datasets recovered different reticulate relationships through several *Leucomexicana* lineages. Two analyses show *Q. transmontana* arising from a lineage that is either of hybrid origin with *Q. laeta* as one parent (Fig 6a) or derived from the ancestor of the lineage that gave rise to *Q. laeta*, with introgression from a distant relative (Fig 6b). The other two networks (Fig 6c, d) point to introgression histories for *Q. striatula* or *Q. rugosa*. Numerous species—*Q. chihuahuensis, Q. oblongifolia, Q. peduncularis, Q. rugosa, laeta SMOC lineage* and *transmontana* lineages—were implicated in multiple networks (Fig 6a-d).

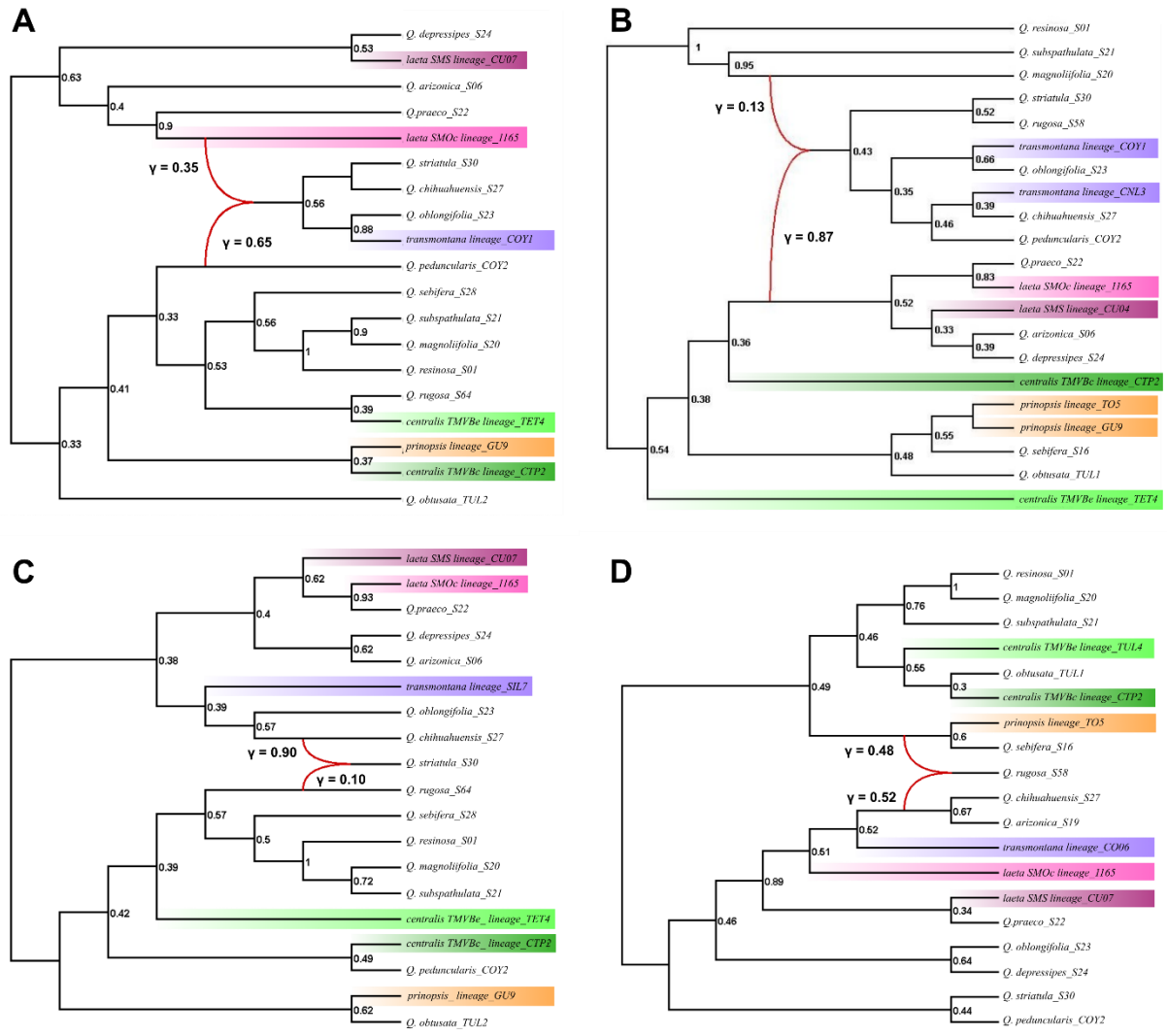


Figure 6. Phylogenetic networks as inferred by SNaQ from four different dataset. γ = inheritance probabilities.

Environmental Differentiation among Lineages

A total of 234 candidate models were generated and tested for each lineage (1,404 in total), of which 37 (2.63% of all models tested) passed the three selection criteria (online Appendix 4a). The evaluation of multiple models identified different sets of environmental variables and parameters to produce the best model for each lineage (online Appendix 4b). The ENMs for the different lineages showed a high predictive power based on AUC values

(AUC *centralis TMVBc.* =0.93; AUC *centralis TMVBe.* =0.91; AUC *laeta SMOc.* = 0.97; AUC *laeta SMS.* = 0.95; AUC *prinopsis* =0.97; AUC *transmontana* =0.96). Overall, the ENMs suggest that the different lineages are allopatric, although certain lineages show parapatric patterns (Table 1). The ENM showed that the highest environmental suitability both for *centralis TMVBc* and *centralis TMVBe* occurred in the temperate forests of the center and eastern region of the Trans-Mexican Volcanic Belt, while *laeta SMS* and *laeta SMOc.* were restricted to the temperate forests of western Mexico (Jalisco, Nayarit, and Durango). In contrast, *transmontana* and *prinopsis* have more xeric affinities. The highest environmental suitability for the *transmontana* lineage was on the mountain areas of the central and southern region of the Sierra Madre Oriental (SMOr), as well as in isolated mountain ranges of the Mexican Plateau, while *prinopsis* was restricted to the center and northern region of the SMOr. Also, the ENMs showed potential contact zones among *prinopsis* and *transmontana* on the central region of the SMOr, *centralis TMVBc* and *centralis TMVBe* on Trans-Mexican Volcanic Belt and *laeta SMS* and *laeta SMOc* on western Mexico (Fig. 7).

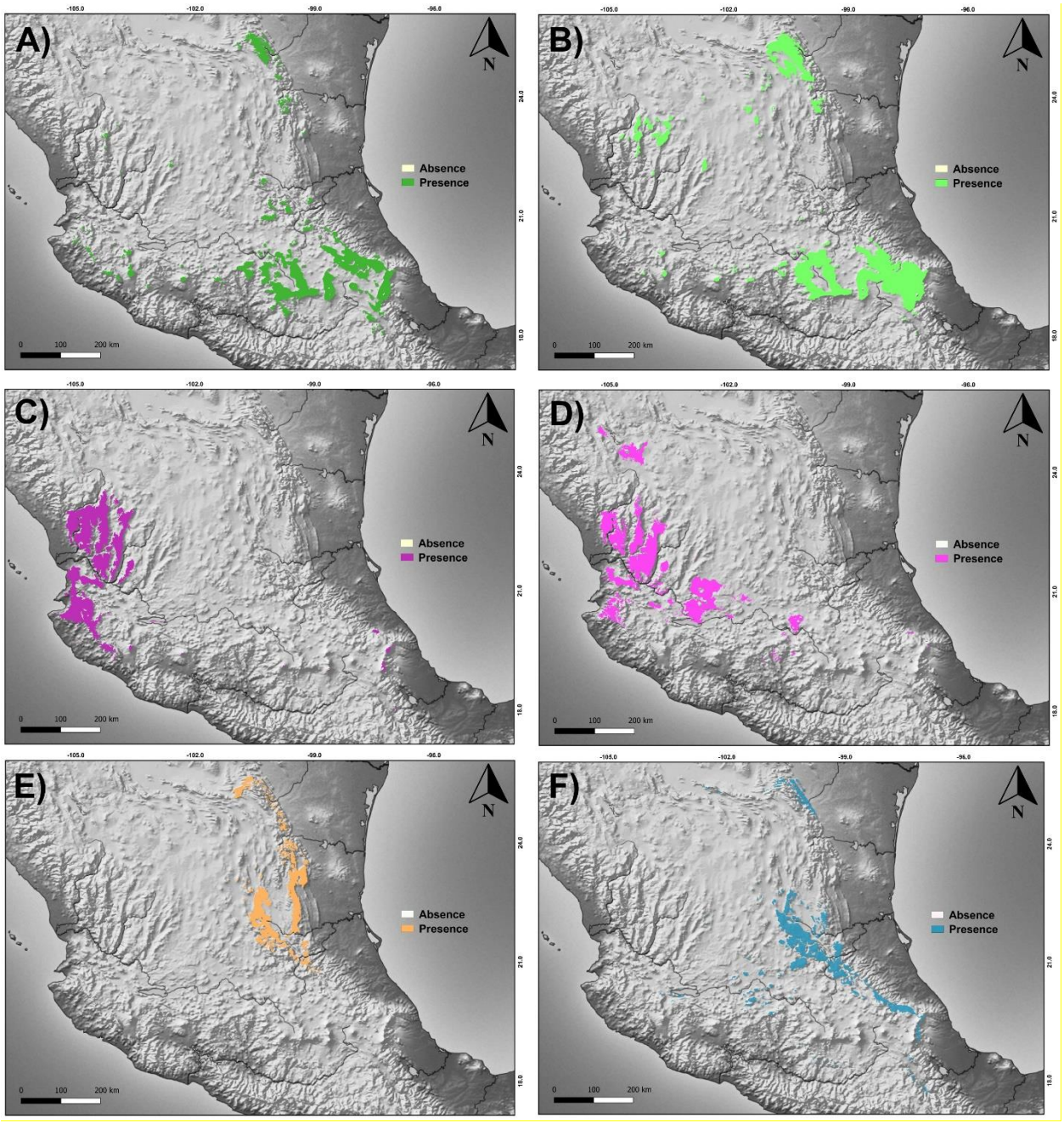


Figure 7. Reconstruction of the ecological niche models (ENM) for the six major lineages detected in the *Quercus laeta* complex. a) Binary distribution for *centralis TMVBc* lineage; b) Binary distribution for *centralis TMVBe* lineage; c) Binary distribution for *laeta* SMS lineage; d) Binary distribution for *laeta* SMOc lineage; e) Binary distribution for *prinopsis* lineage; and f) Binary distribution for *transmontana* lineage.

Table 1. Results of background tests for pairwise comparisons of *Quercus laeta* lineages based on 100 replicates.

Lineages pairwise comparison	**Range overlap**	**Schoener D**	**p-value Background test**	**Inference**
centralis TMVBc – centralis TMVBe	Parapatric	0.60	0.029	Convergent
centralis TMVBc – laeta SMS	Allopatric	0.61	0.019	Convergent
centralis TMVBc – laeta SMOc	Allopatric	0.38	0.009	Divergent
centralis TMVBc – prinopsis	Allopatric	0.27	0.009	Divergent
centralis TMVBc – transmontana	Allopatric	0.55	0.009	Convergent
centralis TMVBe – laeta SMS	Allopatric	0.40	0.009	Divergent
centralis TMVBe – laeta SMOc	Allopatric	0.26	0.009	Divergent
centralis TMVBe – prinopsis	Allopatric	0.24	0.009	Divergent
centralis TMVBe – transmontana	Allopatric	0.49	0.009	Divergent
laeta SMS – laeta SMOc	Parapatric	0.65	0.019	Convergent
laeta SMS – prinopsis	Allopatric	0.14	0.009	Divergent
laeta SMS – transmontana	Allopatric	0.36	0.009	Divergent
laeta SMOc – prinopsis	Allopatric	0.09	0.009	Divergent
laeta SMOc – transmontana	Allopatric	0.22	0.009	Divergent
prinopsis – transmontana	Parapatric	0.48	0.009	Divergent

According to the jackknife test, different combinations of variables explain environmental suitability for the different lineages. For *centralis TMVBc*, temperature annual range (bio7) and mean temperature of driest quarter (bio9) were the most important factors influencing climatic suitability, while *centralis TMVBe*, annual mean temperature (bio1) and precipitation seasonality (bio15). In *laeta SMOc*, temperature of coldest month (bio6), precipitation of driest month (bio14), precipitation seasonality (bio15), precipitation of wettest quarter (bio16), precipitation of driest quarter (bio17) and precipitation of coldest quarter (bio19), were the most important variables influencing the model. For *laeta SMS*, annual mean temperature (bio1) and precipitation seasonality (bio15) and precipitation of coldest quarter (bio19) were the most important factors influencing climatic suitability. Likewise, for *prinopsis*, mean diurnal range (bio2), temperature seasonality (bio4), temperature annual range (bio7) and precipitation of driest month (bio14) and were the

most significant factors in the ENM building. Lastly, for *transmontana*, mean diurnal range (bio2), isothermality (bio3), temperature seasonality (bio4), precipitation seasonality (bio15) and precipitation of warmest quarter (bio18) i were the most important factors influencing the model. The results the background test showed Schoener's D values ranged 0.09 to 0.65 depending on the pair of lineages compared where *laeta SMOc – prinopsis* comparison had lowest similarity value of all pairwise comparisons, while *laeta SMS – laeta SMOc* comparison had greatest similarity value (Table 1). All observed values of *D* were significantly lower than the null distributions, so that the null hypothesis could be rejected in all cases (Table 1; Fig. 8).

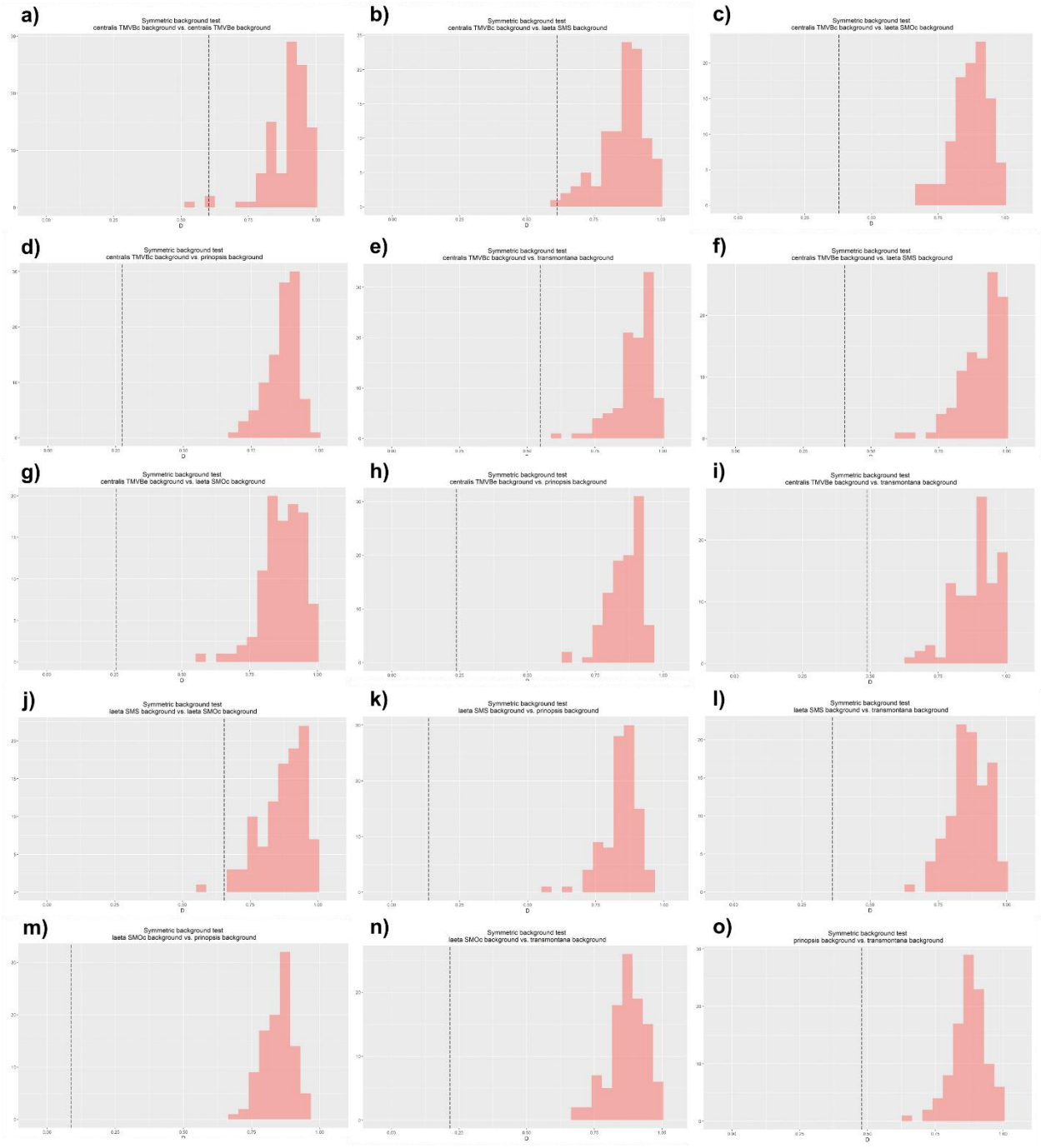


Fig 8. Climatic niche differentiation between the major lineages identified in the *Quercus laeta* complex. Dotted line shown observed values of Schoener's D statistics versus values generated from 100 replicates (pink histograms). a) pairwise tests between *centralis* TMVBc and *centralis* TMVBe lineages; b) pairwise tests between *centralis* TMVBc and

laeta SMS lineages; c) pairwise tests between *centralis* TMVBc and *laeta* SMOc lineages; d) pairwise tests between *centralis* TMVBc and *prinopsis* lineages; e) pairwise tests between *centralis* TMVBc and *transmontana* lineages; f) pairwise tests between *centralis* TMVBe and *laeta* SMS lineages; g) pairwise tests between *centralis* TMVBe and *laeta* SMOc lineages; h) pairwise tests between *centralis* TMVBe and *prinopsis* lineages; i) pairwise tests between *centralis* TMVBe and *transmontana* lineages, j) pairwise tests between *laeta* SMS and *laeta* SMOc lineages; k) pairwise tests between *laeta* SMS and *prinopsis* lineages; l) pairwise tests between *laeta* SMS and *transmontana* lineages; m) pairwise tests between *laeta* SMOc and *prinopsis* lineages; n) pairwise tests between *laeta* SMOc and *transmontana* lineages; o) pairwise tests between *prinopsis* and *transmontana* lineages.

DISCUSSION

Species are the basis of biodiversity research, underlying our inferences in evolution, conservation, and biogeography. Identifying distinct evolutionary lineages that correspond with species is consequently a key challenge in systematics (De Queiroz 2007; Fujita et al. 2012; Luo et al. 2018). Traditional species in *Quercus* have been delimited mainly based on morphology (Valencia-A. et al. 2016; González-Villarreal 2018). However, morphological traits are subject to plasticity and convergence, which makes them difficult to use at times in species delimitation (Deng et al. 2017; Valencia-A. 2021). DNA-based microsatellite markers have clarified relationships within many oak complexes, particularly in Mexico (e.g., McCauley et al. 2019; McCauley and Oyama 2020), shedding light on the gene flow that undermines species boundaries. Occasionally, however, allele frequency data from microsatellites provide insufficient data to establish species boundaries, and use of gene trees from hundreds of loci provide the power needed to estimate evolutionary relationships

and identify hybridization histories (Morales-Saldaña et al. 2022).

Our study uses fine scale sampling at the population level, combining multilocus data based on target enrichment sequencing with the evaluation of ecological niche differentiation to identify evolutionary lineages at multiple levels—populations, species, and deeper clades—in a plant group characterized by introgressive hybridization. Three points are worth highlighting based on the results. First, our data suggest the existence of up to six lineages within the *Q. laeta* complex, each geographically and ecologically restricted and potentially recognizable as a distinct species. Second, our study demonstrates that the *Q. laeta* complex is polyphyletic. Finally, the low support for relationships within the Mexican white oaks is due at least in part to introgressive hybridization. While our analyses did not partition tree support between processes of incomplete lineage sorting and historical or recent introgression, they demonstrate that introgression is playing a role. The difficulty of recognizing species and deeper lineages in the Mexican white oaks is most likely tied up with the processes, as introgressive hybridization, that have contributed to their diversification.

Relationships among lineages in the Q. laeta *complex and other Leucomexicana species: Evolutionary and Ecological Implications*

Recent phylogenomic analyses have greatly contributed to our understanding of the relationships among American oaks (Hipp et al. 2018; 2020). However, these analyses have largely disregarded the potential effects of lineage sorting and introgression (though cf. McVay et al. 2017a, b; Crowl et al. 2020), which may be disproportionately affecting phylogenetic resolution in the relatively young and highly diverse clades of Mexican oaks (Hipp et al. 2020). Within the subsect. *Leucomexicana*, *Quercus laeta* is a clear example of the degree of conflict among gene trees within a young oak clade.

Previous studies have recovered inconsistent relationships among species of the *Q. laeta* complex and close relatives of the subsect. *Leucomexicana*. One of the first phylogenetic studies to include *Q. laeta* (Manos et al. 1999) reported *Q. rugosa* as sister to *Q. laeta*, a result also reported by Hubert et al. (2014) although with a lower support (BS < 50). Both studies had limited taxonomic sampling from Mexico. Later studies with more samples reported *Q. arizonica* as sister to *Q. laeta* (Hipp et al. 2018) and suggested that the *Q. laeta* complex was paraphyletic, with a close relationship to *Q. chihuahuensis* (Hipp et al. 2020).

Our results also support the polyphyly of the *Q. laeta* complex, suggesting that the *Q. laeta* lineages recognized here are more deeply divergent than previously suspected, supporting the hypothesis that these lineages correspond to up to six different species. In addition, although the objective of our study was not to present an exhaustive phylogeny of the subsect. *Leucomexicana* and even though we found widespread gene trees discordance in certain clades, our results also provide advances in the understanding of the relationships among American oaks. The *Q. subspathulata – Q. resinosa – Q. magnoliifolia* and *Q. rugosa – Q. sebifera* clades, as well as the close relationship between the *laeta* SMOc lineage and *Q. praeco*, are stable across the different analyses.

The addition of numerous samples of the *Q. laeta* complex, however, introduces phylogenetic complexity. One of our most interesting results was the detection of four lineages that showed a clear geographic pattern with independent phylogenetic affinities. The *laeta SMS* and *laeta SMOc* lineages, located in west Mexico, are suggested in the MSC trees to have a previously unsuspected close relationship with *Q. praeco*. However, the ML concatenated analysis also suggested phylogenetic affinities with *Q. arizonica – Q. depressipes* (BS = 61), although concordance factors demonstrate discordance at the gene

and nucleotide levels. The other geographically structured lineages correspond to *centralis TMVBc* and *centralis TMVBe*, for which recovered phylogenetic relationships were ambiguous and with poorly supported. This low support in conjunction with low gCF and sCF values suggests that there is not enough information at this level to resolve phylogenetic affinities for these lineages. Nevertheless, all analyses placed *centralis TMVBc* and *centralis TMVBe* in separate clades, where they showed closer phylogenetic affinities to other species of the subsect. *Leucomexicana* subsection than to each other, supporting that they are polyphyletic, a result also shown in the molecular ordination (Fig 3, left panel).

The types of discrepancies observed among analyses conducted on different taxon samples in our study are common among recently and rapidly diverged lineages due to processes like ILS, introgression, and stochastic error (Murphy et al. 2020; Shee et al. 2020; Thomas et al. 2021), so that disentangling phylogenetic relationship in these groups remains challenging. In this context, we considered three hypotheses to reconcile the slightly different relationships inferred by the ML tree and MSC methods and unsupported nodes.

Hypothesis 1: Non-decisive data. Despite our somewhat large amount of data (155 low-copy nuclear genes), our datasets could be non-informative at these nodes. This could be particularly true for those nodes where we found low support and gCF values as well as neutral sCF values (~ 33%) for the ML concatenated tree suggesting that stochastic error could be a source of conflict in the phylogenetic signal; so, our dataset is not informative in some areas of the tree, mainly at a deep level.

Hypothesis 2: ILS. Incomplete lineage sorting is expected to be particularly problematic under scenarios of relatively short branches between speciation events and

large effective population sizes, as has been reported for American oaks (Hipp et al. 2018; Sork et al. 2022). In this context, we found substantial discordance between gene trees and the ASTRAL species tree, with only 38% agreement of gene tree quartets with the species tree; further, the low quartet support values found for certain branches of the ASTRAL tree (<50 Fig. S3) sustains the hypothesis of ILS as one cause of incongruence within *Q. laeta* lineages.

Hypothesis 3: Hybridization and introgressive gene flow. Considering that ASTRAL can be statistically inconsistent when gene trees evolve on a phylogenetic network (Long and Kubatko 2018), and that *Quercus* is renowned for interspecific gene flow (Belahbib et al. 2001; Eaton et al. 2015; Sullivan et al. 2016; Crowl et al. 2020; Li et al. 2021), hybridization may account for low phylogenetic support and discrepancies among analyses. Phylogenetic networks model both ILS and hybridization as sources of gene tree discordance. In this context, our results strongly suggest that reticulation has taken place between *Leucomexicana* lineages in the past, causing phylogenetic discordance within the clade, leading to conflict in the phylogenetic placement of some lineages.

Our results based on different phylogenetic networks are sensitive to taxa sampling (cf. Karimi et al. 2020), yet in combination they suggest events of ancient introgression that involved the ancestors of three present-day *Q. laeta* lineages (*laeta, prinopsis* and *transmontana*), but none among *centralis* lineages. In this context, the discordant phylogenetic relationships proposed among *Q. chihuahuensis, Q. oblongifolia, Q. peduncularis,* and the *transmontana* lineage could be explained by ancient introgression events, which together with rapid diversification of the Mexican white oaks (Hipp et al. 2018, 2020), probably has given these lineages too little time to accumulate informative genomic polymorphisms.

In addition to the introgression processes that have contributed to the diversification of oak species (Hipp et al. 2020; Kremer and Hipp 2020), it has been proposed that the high rates of lineage diversification in the Mexican oaks are associated with high rates of evolution along moisture gradients (Hipp et al. 2018). In support of this claim, the climatic niche differentiation results suggest that niche divergence is the most common pattern observed among lineages detected within *Q. laeta* complex. This inference of niche differentiation can also be seen in the ENM projection, where 12 of 15 comparisons suggests allopatric patterns, occupying geographically and ecologically distinct habitats with the exception of the *centralis TMVBc – centralis TMVBe and laeta SMS – laeta SMOc* lineages who shown parapatric patterns with a significant degree of niche similarity. In the case of *prinopsis – transmontana* lineages a parapatric pattern was also observed with a potential contact zone in the central region of the Sierra Madre Oriental (SMOr) although evidence of genetic admixture between these two lineages was not observed. Our results thus suggest that mountain barriers may have played key roles in speciation and diversification through the effects of topographic complexity on ecological stratification, environmental heterogeneity, and limitation of gene flow (Rodríguez-Correa et al. 2015; Hipp et al. 2018; Barret et al. 2019).

Delimitation of Major Lineages: The Need for a Taxonomic Reappraisal

Our work supports recent inferences from improved sampling at the population level and multilocus molecular analyses that oak species diversity in Mexico is underestimated, despite recent taxonomic descriptions and reappraisal (Valencia et al. 2016; González-Villarreal 2018; McCauley et al. 2019; McCauley and Oyama 2020; Valencia and Coombes 2020; Morales-Saldaña et al. 2022). *Quercus laeta* has previously been recognized as an oak species with exceptional population polymorphism (Valencia-A 2004), but the fact

that its potentially distinct morphotypes overlap in qualitative micromorphological characters regarded as diagnostic characters (e.g., papillae and glands on the leaf abaxial, sessile and slightly contorted trichomes on the abaxial surface) led taxonomists to consider *Q. laeta* to be a single, highly polymorphic species (McVaugh 1974; Romero et al. 2002). Our phylogenetic (ML) and non-phylogenetic (NMDS ordination) results demonstrate that the *Q. laeta* complex constitutes as many as six genetically and ecologically divergent populations that are as distinct genetically as other closely related species. This work expands on recent phylogenomic and population genetic studies demonstrating the existence of at least two different entities, *Q. transmontana* and *Q. laeta* (Morales-Saldaña et al. 2022) and serves as a crucial step toward taxonomic revision of a taxonomically difficult clade of Mexican white oaks.

Quercus transmontana, delimited using genetic and morphometric data by Morales-Saldaña et al. (2022), was recovered in this study as a monophyletic group with high BS values in the ML analyses (Fig 2) and clearly differentiated in the NMDS ordination (Fig 3). Furthermore, the species exhibits strong ecological niche differentiation from the rest of the lineages (Fig 8; Table 1). Together, these lines of evidence suggest that *Q. transmontana* is as distinct a species by phylogenetic, morphological, genetic, and ecological criteria as any other closely related Mexican white oak. On the other hand, Q. *laeta*, was also recognized by genetic and morphometric data (Morales-Saldaña et al. 2022), but in the present study it was split into up to five lineages. Based on results from both ML and NMDS methods and strong ecological niche differentiation relative to the rest of the lineages (Fig 8; Table 1), we suggest that the *prinopsis* lineage may also merit recognition as an independent species (*Q. prinopsis*).

By contrast, the status of the *centralis* TMVBc, *centralis* TMVBe, *laeta* SMS, and *laeta* SMOc lineages is ambiguous. Although all analyses recovered each of these lineages as non-monophyletic (Fig 2,4,5), but their paraphyly in at least the ML analyses might be a consequence of both introgression with other species of the *Leucomexicana* subsection (as demonstrated in Eaton et al., 2015 for the oaks of section *Virentes*) and ILS, which together may render otherwise monophyletic lineages paraphyletic (e.g., Rieseberg and Bouillet 1994; Funk and Onlamd 2003; Hörandl and Stuessy 2010; Carnicero et al. 2019; Kato et al. 2019; Bard et al. 2021). Considering that our results suggest a significant degree of niche similarity and that there is no evidence of morphological divergence among *laeta* SMS, and *laeta* SMOc lineages, we consider that both should be maintained under the single name *Q. laeta* for the time being. Similarly, as it is unclear whether *centralis* TMVBc and *centralis* TMVBe are reciprocally monophyletic (Fig. 4, 5 and 6) and there are no demonstrable ecological or morphological differences between the putative *centralis* lineages, we recommend maintaining *centralis TMVBc* and *centralis TMVBe* as a single species for now. Although the putative species name designations presented in this paper are based on the names proposed by Trelease (1924), older names exist that might have nomenclatural priority. Detailed morphological diagnosis and nomenclatural research will be necessary for the lineages proposed here before they can be formally redescribed.

CONCLUSION

The present study highlights the importance of utilizing approaches that bridge the micro- and macroevolutionary scales to resolve evolutionary relationships and species boundaries in young, potentially reticulate species complexes. Using target enrichment and ecological niche modeling, we recognized up to six evolutionarily independent lineages, greatly increasing the recognized species diversity of the *Q. laeta* complex. Our work

supports the existence of at least four species in the complex—*Q. centralis, Q. laeta, Q. prinopsis* and *Q. transmontana*—and lays a foundation for the taxonomic revision required to validate these names. Moreover, although the objective of the present study was not to estimate an exhaustive phylogeny for the subsect. *Leucomexicana,* the inferred phylogenies corroborated that the different lineages comprising the *Q. laeta* complex are more closely related to other species than to each another. Also, the existence of significant ecological niche differentiation among the *Q. laeta* complex lineages suggests that ecological differentiation has been a part of evolutionary divergence among these closely related white oaks, either driving speciation or resulting from it.

ACKNOWLEDGMENTS

We thank Oscar De Luna, Gonzalo Contreras, Ricardo Gaytán, and Maribel Arenas for their help with field collection. We thank the at the Field Museum of Natural History for use of sequencing equipment and for equipment training and support. Saddan Morales-Saldaña thanks Posgrado en Ciencias Biológicas—UNAM and Consejo Nacional de Ciencia y Tecnología (CONACYT) for the PhD scholarship and financial support to develop graduate studies at UNAM (scholarship/CVU: 483720/624511). This study constitutes a partial fulfillment of the Graduate Program in Biological Sciences of the National Autonomous University of Mexico (UNAM).

REFERENCES

Abbott R., Albach D., Ansell S., Arntzen J.W., Baird S.J., Bierne N., Boughman J., Brelsford A., Buerkle C.A., Buggs R., Butlin R.K., Dieckmann U., Eroukhmanoff F., Grill A., Cahan S.H., Hermansen J.S., Hewitt G., Hudson A.G., Jiggins C., Jones J., Keller B., Marczewski T., Mallet J., Martinez-Rodriguez P., Möst M., Mullen S., Nichols R., Nolte A.W., Parisod C., Pfennig K., Rice A.M., Ritchie M.G., Seifert B., Smadja C.M., Stelkens R., Szymura J.M., Väinolä R., Wolf J.B.W., Zinner D. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Aiello-Lammens M.E., Boria R.A., Radosavljevic A., Vilela B., Anderson R.P. 2015. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography. 38:541–545.

Anderson R.P., Lew D., Peterson A.T. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecol. Modell. 162:211–232.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Leslin V.M., Nikolenko I.S., Pham S., Prjibelski D.A., Phyhkin V.A., Sirotkin V.A., Vyahhi N., Tesler G., Alekseyev A.M., Pevzner A.P. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477.

Bard N.W., Miller C.S., Bruederle L.P. 2021. High genomic diversity maintained by populations of *Carex scirpoidea* subsp. *convoluta*, a paraphyletic Great Lakes ecotype. Conserv. Genet. 22:169–185.

Bapteste E., Susko E., Leigh J., Ruiz-Trillo I., Bucknam J., Doolittle W.F. 2008. Alternative methods for concatenation of core genes indicate a lack of resolution in deep nodes of the prokaryotic phylogeny. Mol. Biol. Evol. 25:83–91.

Barrett C.F., Sinn B.T., King L.T., Medina J.C., Bacon C.D., Lahmeyer S.C., Hodel D.R. 2019. Phylogenomics, biogeography and evolution in the American genus *Brahea* (Arecaceae). Bot. J. Linn. 190:242–259.

Belahbib N., Pemonge M.H., Ouassou A., Sbay H., Kremer A., Petit R.J. 2001. Frequent cytoplasmic exchanges between oak species that are not closely related: *Quercus suber* and *Q. ilex* in Morocco. Mol. Ecol. 10:2003–2012.

Bellarosa R., Simeone M.C., Papini A., Schirone B. 2005. Utility of ITS sequence data for phylogenetic reconstruction of Italian *Quercus* spp. Mol. Phylogenet. Evol. 34:355–370.

Bezanson J., Edelman A., Karpinski S., Shah V.B. 2017. Julia: A fresh approach to numerical computing. SIAM Rev Soc Ind Appl Math. 59:65–98.

Blair M.E., Sterling E.J., Dusch M., Raxworthy C.J., Pearson R.G. 2013. Ecological divergence and speciation between lemur (*Eulemur*) sister species in Madagascar. J. Evol. Biol. 26:1790–1801.

Brauner N., Shacham M. 1998. Role of range and precision of the independent variable in regression of data. AICHE J. 44:603–611.

Breinholt J.W., Carey S.B., Tiley G.P., Davis E.C., Endara L., McDaniel S.F., Neves G.L., Sessa B.E., Konrat M.V., Chantanaorrapint S., Fawcett S., Ickert-Bond M.S., Labiak H.P., Larraín J., Lehnert M., Lewis R.L., Nagalingum S.N., Patel N., Rensing S.A., Testo W., Vasco A., Villarreal C.J., Williams W.E., Burleigh J.G. 2021. A target enrichment probe set for resolving the flagellate land plant tree of life. Appl. Plant Sci. 9:e11406.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Buddenhagen C., Lemmon A.R., Lemmon E.M., Bruhl J., Cappa J., Clement W.L., Donoghue J.M., Edwards J.E., Hipp L.A., Kortyna M., Mitchell N., Moore A., Prychid J.C., Segovia-Salcedo C.M., Simmons P.M., Soltis S.P., Wanke S., Mast A. 2016. Anchored phylogenomics of angiosperms I: assessing the robustness of phylogenetic estimates. BioRxiv. 086298.

Calixto-Rojas M., Lira-Noriega A., Rubio-Godoy M., Pérez-Ponce de León G., Pinacho-Pinacho C.D. 2021. Phylogenetic relationships and ecological niche conservatism in killifish (Profundulidae) in Mesoamerica. J. Fish Biol. 99:396–410.

Carnicero P., Schönswetter P., Garcia-Jacas N., Galbany-Casals M. 2019. Is there a need for accepting paraphyletic taxa? A case study in the Sardinian endemic *Cymbalaria muelleri* (Plantaginaceae). Bot. J. Linn. Soc. 191:325–338.

Cavender-Bares J., González-Rodríguez A., Eaton D.A., Hipp A.A., Beulke A., Manos P.S. 2015. Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): a genomic and population genetics approach. Mol. Ecol. 24:3668–3687.

Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics. 30:3317–3324.

Cobos M.E., Peterson A.T., Barve N., Osorio-Olvera L. 2019. kuenm: an R package for detailed development of ecological niche models using Maxent. PeerJ, 7:e6281.

Crowl A.A., Manos P.S., McVay J.D., Lemmon A.R., Lemmon E.M., Hipp A.L. 2020. Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*). New Phytol. 226:1158–1170.

Deng M., Jiang X.L., Song Y.G., Coombes A., Yang X.R., Xiong Y.S., Li Q.S. 2017. Leaf epidermal features of *Quercus* Group Ilex (Fagaceae) and their application to species identification. Rev. Palaeobot. Palynol. 237:10–36.

De Queiroz K. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–886.

Dodsworth S., Pokorny L., Johnson M.G., Kim J.T., Maurin O., Wickett N.J., Forest F., Baker W.J. 2019. Hyb-Seq for flowering plant systematics. Trends Plant Sci. 24:887–891.

Dormann C.F., McPherson J., Araújo M., Bivand R., Bolliger J., Carl G., Davies G.R., Hirzel A., Jetz W., Kissling D., Ohlemüller R., Peres-Neto P.R., Reineking B., Schröder B., Schurr M.F., Wilson R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography. 30:609–628.

Dormann C.F., Elith J., Bacher S., Buchmann C., Carl G., Carré G., García-Marquéz J.R., Gruber B., Lafourcade B., Leitão J., Münkemüller T., McClean C., Osborne E.P., Reineking B., Schröder B., Skidmore K.A., Zurell D. Lautenbach S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography. 36:27–46.

Eaton D.A., Hipp A.L., González-Rodríguez A., Cavender-Bares J. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution. 69:2587–2601.

Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack E.J., Glenn C.T., Zhong B., Wu S., Lemmon M.E., Lemmon R.A., Leaché D.A., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462.

Fick S.E., Hijmans R.J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37:4302–4315.

Folk R.A., Mandel J.R., Freudenstein J.V. 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of

angiosperms. Syst. Biol. 66:320–337.

Fujita M.K., Leaché A.D., Burbrink F.T., McGuire J.A., Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27:480–488.

Funk D.J., Omland K.E. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Evol. 34:397–423.

Galante P.J., Alade B., Muscarella R., Jansa S.A., Goodman S.M., Anderson R.P. 2018. The challenge of modeling niches and distributions for data-poor species: a comprehensive approach to model complexity. Ecography. 41:726–736

Gardner E.M., Johnson M.G., Ragone D., Wickett N.J., Zerega N.J. 2016. Low-coverage, whole-genome sequencing of *Artocarpus camansi* (Moraceae) for phylogenetic marker development and gene discovery. Appl. Plant Sci. 4: 1600017.

Gernandt D.S., Aguirre-Dugua X., Vázquez-Lobo A., Willyard A., Moreno-Letelier A., Pérez de la Rosa J.A., Piñero D., Liston, A. 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection *Australes*. Am. J. Bot. 105:711–725.

Glor R.E., Warren D. 2010. Testing ecological explanations for biogeographic boundaries. Evol.; int. j. org. evol. 65:673–683.

González-Villarreal L.M. 2018. Dos nuevas especies de encinos (*Quercus*: Fagaceae), adicionales para la Flora de Jalisco y Áreas Colindantes, en el Occidente de México. Ibugana 9:47–71.

Goudarzi F., Hemami M.R., Rancilhac L., Malekian M., Fakheran S., Elmer K.R., Steinfartz S. 2019. Geographic separation and genetic differentiation of populations are not coupled with niche differentiation in threatened Kaiser's spotted newt (*Neurergus kaiseri*). Sci. Rep. 9:1–12.

Guisan A., Edwards T.C., Hastie T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model. 157:89–100.

Gutiérrez-Ortega J.S., Salinas-Rodríguez M.M., Ito T., Pérez-Farrera M.A., Vovides A.P., Martínez J.F., Molina-Freaner F., Hernández-López A., Kawaguchi L., Nagano J.A., Kajita T., Watano Y., Tsuchimatsu T., Takahashi Y., Murakami M. 2020. Niche conservatism promotes speciation in cycads: the case of *Dioon merolae* (Zamiaceae) in Mexico. New Phytol. 227:1872–1884.

Hörandl E., Stuessy T.F. 2010. Paraphyletic groups as natural units of biological classification. Taxon. 59:1641–1653.

Hale H., Gardner E.M., Viruel J., Pokorny L., Johnson M.G. 2020. Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. Appl. Plant Sci. 8:e11337.

Hardin J.W. 1975. Hybridization and introgression in *Quercus alba*. J. Arnold Arbor. 56:336–363

Hipp A.L., Eaton D.A., Cavender-Bares J., Fitzek E., Nipper R., Manos P.S. 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. PloS one. 9:e93975.

Hipp A.L., Manos P.S., González-Rodríguez A., Hahn M., Kaproth M., McVay J.D., Valencia A.S., Cavender-Bares J. 2018. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. New Phytol. 217:439–452.

Hipp A.L., Manos P.S., Hahn M., Avishai M., Bodénès C., Cavender-Bares J., Crowl A. A., Deng M., Denk T., Fitz-Gibbon S., Gailing O., González-Elizondo S.M, González-Rodríguez A., Grimm W.G., Jiang L.X., Kremer A., Lesur I., McVay D.J., Plomion C.,

Rodríguez-Correa H., Schulze D.E., Simeone C.M., Sork L.V., Valencia-Avalos S. 2020. Genomic landscape of the global oak phylogeny. New Phytol. 226:1198–1212.

Hubert F., Grimm G.W., Jousselin E., Berry V., Franc A., Kremer A. 2014. Multiple nuclear genes stabilize the phylogenetic backbone of the genus *Quercus*. Syst. Biodivers. 12:405–423.

Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C., Wickett N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl. Plant Sci. 4:1600016.

Karimi N., Grover C.E., Gallagher J.P., Wendel J.F., Ané C., Baum D.A. 2020. Reticulate evolution helps explain apparent homoplasy in floral biology and pollination in baobabs (Adansonia; Bombacoideae; Malvaceae). Syst. Biol. 69:462–478.

Kato M., Werukamkul P., Won H., Koi S. 2019. Paraphyletic species of Podostemaceae: *Cladopus fallax* and *Polypleurum wallichii*. Phytotaxa. 401:33–48.

Katoh K., Misawa K., Kuma K., Miyata T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30: 3059–3066.

Katoh K., Standley D. M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649.

Kleinkopf J.A., Roberts W.R., Wagner W.L., Roalson E.H. 2019. Diversification of

Hawaiian Cyrtandra (Gesneriaceae) under the influence of incomplete lineage sorting and hybridization. J Syst Evol. 57:561–578.

Kremer A., Hipp A.L. 2020. Oaks: an evolutionary success story. New Phytol. 226:987–1011.

Lefort F., Douglas G.C. 1999. An efficient micro-method of DNA isolation from mature leaves of four hardwood tree species *Acer*, *Fraxinus*, *Prunus* and *Quercus*. Ann. For. Sci., 56:259–263.

Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25:1754–1760.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25:2078–2079.

Li X., Wei G., El-Kassaby Y.A., Fang Y. 2021. Hybridization and introgression in sympatric and allopatric populations of four oak species. BMC Plant Biol. 21:1–14.

Liebmann F.M. 1854. Oversigt Over det Kongelige Danske Videnskabernes Selskabs Forhandlinger og Dets Medlemmers Arbeider. Copenhagen 173.

Lin H.Y., Gu K.J., Li W.H., Zhao Y.P. 2021. Integrating coalescent-based species delimitation with ecological niche modeling delimited two species within the *Stewartia sinensis* complex (Theaceae*)*. J. Syst. Evol., 60: 1037-1048.

Linder C.R., Moret B.M.E., Nakhleh L. 2003. Network (reticulate) evolution: Biology, models, and algorithms. The Pacific Symposium.

Liang L., Zhenxiang X., Charles C.D. 2015. Coalescent Methods Are Robust to the Simultaneous Effects of Long Branches and Incomplete Lineage Sorting. Mol. Biol. Evol. 32: 791–805.

Loiseau O., Olivares I., Paris M., de La Harpe M., Weigand A., Koubínová D., Rolland J., Bacon D.C., Balslev H., Borchsenius F., Cano A., Couvreur P.L.T., Denatte C., Fardin F., Gayot M., Mejía F., Mora-Machado T., Perret M., Roncal J., Sanín J.M. Stauffer F., Lexer C., Kessler M., Salamin, N. 2019. Targeted capture of hundreds of nuclear genes unravels phylogenetic relationships of the diverse Neotropical palm tribe Geonomateae. Front. Plant Sci. 10:864.

Long C., Kubatko L. 2018. The effect of gene flow on coalescent-based species-tree inference. Syst. Biol. 67:770–785.

Luo A., Ling C., Ho S.Y., Zhu C.D. 2018. Comparison of methods for molecular species delimitation across a range of speciation scenarios. Syst. Biol. 67:830–846.

Ma Z.Y., Nie Z.L., Ren C., Liu X.Q., Zimmer E.A., Wen J. 2021. Phylogenomic relationships and character evolution of the grape family (Vitaceae). Mol. Phylogenet. Evol. 154:106948.

Maddison W.P., Knowles L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21–30.

Mallet J., Besansky N., Hahn M.W. 2016. How reticulated are species?. BioEssays. 38:140–149.

Manos P.S., Doyle J.J., Nixon K.C. 1999. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). Mol. Phylogenet. Evol. 12:333–349.

Manos P.S., Hipp A.L. 2021. An Updated Infrageneric Classification of the North American Oaks (*Quercus* Subgenus *Quercus*): Review of the Contribution of Phylogenomic Data to Biogeography and Species Diversity. Forests. 12:786.

Mayol M., Rosselló J.A. 2001. Why nuclear ribosomal DNA spacers (ITS) tell different

stories in *Quercus*. Mol. Phylogenet. Evol. 19:167–176.

McCauley R.A., Cortés-Palomec A.C., Oyama K. 2019. Species diversification in a lineage of Mexican red oak (*Quercus* section *Lobatae* subsection *Racemiflorae*)—the interplay between distance, habitat, and hybridization. Tree Genet. Genomes 15:1–23.

McCauley R.A., Oyama K. 2020. A re-evaluation of taxonomy in *Quercus* section *Lobatae* subsection *Racemiflorae* (Fagaceae), resurrection of the name *Q. pennivenia* and description of a new taxon, *Q. huicholensis*. Phytotaxa. 471:247–257.

McVay J.D., Hipp A.L., Manos P.S. 2017a. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. Proc. Royal Soc. B. 284:20170300.

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

McVaugh R. 1974. Fagaceae. In: McVaugh R. ed. Flora Novo-galiciana. Controbutions from the University of Michigan herbarium. 12:50-30.

Minh B.Q., Hahn M.W., Lanfear R. 2020. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37:2727–2733.

Morales-Saldaña S., Valencia-Ávalos S., Oyama K., Sánchez E.T., Hipp A.L., González-Rodríguez A. 2022. Even more oak species in Mexico? Genetic structure and morphological differentiation support the presence of at least two specific entities within *Quercus laeta*. J. Syst. Evol. 60:1124–1139.

Murphy B., Forest F., Barraclough T., Rosindell J., Bellot S., Cowan R., Golos M., Jebb M., Cheek M. 2020. A phylogenomic analysis of *Nepenthes* (Nepenthaceae). Mol. Phylogenet. Evol. 144:106668.

Nunes L.A., Pearson R.G. 2017. A null biogeographical test for assessing ecological niche

evolution. J. Biogeogr. 44:1331–1343.

Oksanen J., Blanchet F.G., Friendly M., Kindt R., Legendre P., McGlinn D., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Szoecs E., Wagner H. 2018. vegan: Community ecology package. R package version 2.5–2. https://CRAN.Rproject.org/package=vegan

Olson D.M., Dinerstein E., Wikramanayake E.D., Burgess N.D., Powell G.V.N., Underwood E.C., D'Amico J.A., Itoua I., Strand H.E., Morrison J.C., Loucks C.J., Allnutt T.F., Ricketts T.H., Kura Y., Lamoreux J.F., Wettengel W.W., Hedao P., Kassem K.R. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. Bioscience 51:933–938.

Palmer E.J. 1948. Hybrid oaks of North America. J. Arnold Arbor. 29:1–48.

Pearson R.G., Raxworthy C.J., Nakamura M., Peterson A.T. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. 34:102–117.

Pease J.B., Brown J.W., Walker J.F., Hinchliff C.E., Smith S.A. 2018. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. Am. J. Bot. 105:385–403.

Peterson A.T., Papeş M., Soberón J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol Modell. 213:63–72

Phillips S.J., Anderson R.P., Schapire R.E. 2006. Maximum entropy modeling of species geographic distributions. Ecol Modell. 190:231–259.

Rambaut A. 2012. FigTree. Tree figure drawing tool version 1.4.0. http://tree.bio.ed.ac.uk/software/figtree/.690

R Core Team. 2013. R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rancilhac L., Irisarri I., Angelini C., Arntzen J.W., Babik W., Bossuyt F., Künsel S., Lüddecke T., Pasmans F., Sanchez E., Weisrock D. Veith M., Wielstra B., Steinfartz S., Hofreiter M., Philippe H., Vences M. 2021. Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders. Mol. Phylogenet. Evol. 155:106967.

Ree R.H., Hipp A.L. 2015. Inferring phylogenetic history from restriction site associated DNA (RADseq). In: Hoerandl E, Appelhaus M, eds. Next Generation Sequencing in Plant Systematics. Koenigstein: Koeltz Scientific Books, 181–204.

Rieseberg L.H., Brouillet L. 1994. Are many plant species paraphyletic?. Taxon. 43:21–32.

Rissler L.J., Apodaca J.J. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). Syst. Biol. 56:924–942.

Romero S.R., Carlos E., Zenteno R., de Lourdes M., Enríquez A. 2002. El género *Quercus* (Fagaceae) en el estado de México. Ann. Missouri Bot. Gard. 89:551–593.

Rodríguez-Correa H., Oyama K., MacGregor-Fors I., González-Rodríguez A. 2015. How are oaks distributed in the Neotropics? A perspective from species turnover, areas of endemism, and climatic niches. Int. J. Plant Sci. 176:222–231.

Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol., 33:1654–1668.

Shcheglovitova M., Anderson R.P. 2013. Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. Ecol Modell. 269:9–17

Shee Z.Q., Frodin D.G., Cámara-Leret R., Pokorny L. 2020. Reconstructing the complex

evolutionary history of the Papuasian Schefflera radiation through herbariomics. Front. Plant Sci. 11:258

Simeone M.C., Piredda R., Papini A., Vessella F., Schirone B. 2013. Application of plastid and nuclear markers to DNA barcoding of Euro-Mediterranean oaks (*Quercus*, Fagaceae): problems, prospects and phylogenetic implications. Bot. J. Linn. Soc. 172:478–499.

Singh P., Irisarri I., Torres-Dowdall J., Thallinger G.G., Svardal H., Lemmon, E. M., Koblmüller S., Meyer A., Sturmbauer, C. 2022. Phylogenomics of trophically diverse cichlids disentangles processes driving adaptive radiation and repeated trophic transitions. Ecol. Evol. 12:e9077.

Smith A.B., Godsoe W., Rodríguez-Sánchez F., Wang H.H., Warren D. 2019. Niche estimation above and below the species level. Trends Ecol. Evol. 34:260–273.

Solís-Lemus C., Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS Genet. 12:e1005896.

Solís-Lemus C., Bastide P., Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. Mol. Biol. Evol. 34:3292–3298.

Sork V.L., Cokus S.J., Fitz-Gibbon S.T., Zimin A.V., Puiu D., Garcia, J. A., Gugger F.P., Henriquez L.C., Zhen Y., Lohmueller E.K., Pellegrini M., Salzberg, S. L. (2022). High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. Nat. Commun. 13:1–15.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Su X., Wu G., Li L., Liu J. 2015. Species delimitation in plants using the Qinghai–Tibet Plateau endemic *Orinus* (Poaceae: Tridentinae) as an example. Ann. Bot. 116:35–48.

Sullivan A.R., Owusu S.A., Weber J.A., Hipp A.L., Gailing O. 2016. Hybridization and divergence in multi-species oak (*Quercus*) communities. Bot. J. Linn. Soc. 181:99–114.

Swofford D. L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods), version 4.0b10 win32. Sunderland, MA: Sinauer Associates

Taylor S.A., Larson E.L. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nat. Ecol. Evol. 3:170–177.

Thomas A.E., Igea J., Meudt H.M., Albach D.C., Lee W.G., Tanentzap A.J. 2021. Using target sequence capture to improve the phylogenetic resolution of a rapid radiation in New Zealand Veronica. Am. J. Bot. 108:1289–1306.

Trelease W. 1924. The American oaks. Mem. Natl. Acad. Sci. 20:1–255

Valencia-A.S. 2004. Diversidad del género *Quercus* (Fagaceae) en México. Bot. Sci.75:33–53.

Valencia-A.S. 2021. Species delimitation in the genus *Quercus* (Fagaceae). Bot. Sci. 99:1–12.

Valencia S., Coombes A.J. 2020. Nomenclatural revision and lectotypification of five species of Mexican oaks: *Quercus* (Fagaceae). Phytotaxa. 428:81–92.

Valencia S., Rosales J.L.S., Arellano O.J.S. 2016. A new species of *Quercus*, section *Lobatae* (Fagaceae) from the Sierra Madre Oriental, Mexico. Phytotaxa. 269:120–126.

Van Valen L. 1976. Ecological species, multispecies, and oaks. Taxon 25:233–239.

Vatanparast M., Powell A., Doyle J.J., Egan A.N. 2018. Targeting legume loci: A comparison of three methods for target enrichment bait design in Leguminosae phylogenomics. Appl. Plant Sci. 6: e1036.

Villaverde T., Pokorny L., Olsson S., Rincón-Barrado M., Johnson M.G., Gardner E.M., Wickett J.N., Molero J., Riina R., Sanmartín, I. 2018. Bridging the micro-and

macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. New Phytol. 220:636–650.

Warnow T. 2015. Concatenation analyses in the presence of incomplete lineage sorting. PLoS Curr.7.

Warren D.L., Glor R.E., Turelli M. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evol.; int. j. org. evol. 62:2868–2883.

Warren D. L., Glor R.E., Turelli M. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. Ecography 33:607–611.

Weitemier K., Straub S.C., Cronn R.C., Fishbein M., Schmickl R., McDonnell A., Liston, A. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. Appl. Plant Sci. 2:1400042.

Wen D., Yu Y., Zhu J., Nakhleh L. 2008. Inferring phylogenetic networks using PhyloNet. Syst. Biol., 67:735–740.

Wooten J.A., Gibbs H.L. 2012. Niche divergence and lineage diversification among closely related Sistrurus rattlesnakes. J. Evol. Biol. 25:317–328.

Zhang C, Rabiee M., Sayyari E., Mirarab S.2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinform. 19:153.

Figure captions

Figure 1. Geographical distribution of the different samples collected. a) Populations for the *Quercus laeta* complex (subset 1), 1:Guadalcazar, 2:La Congoja, 3:San Nicolás Tolentino, 4:San José Gallinas, 5:Vergel de Bernalejo, 6:Cieneguilla, 7:Silao, 8:Sierra El Cuale, 9:Contepec, 10:Waterfalls of Tuliman, 11:Tetela de Ocampo, 12:Coyula*;* b) Localities collected for species from the subsect. *Leucomexicana* (subset 2).

Figure 2. Tree based on maximum likelihood (155 loci: 302,321 pb) with the GTRCAT model and 1000 bootstraps. Bootstrap support values >50% are indicated above the branches. Colors and labels to the right of the tree indicate the different lineages proposed.

Figure. 3. Ordination of *Quercus laeta* individuals sequenced, based on genetic data. Ordination is based on a three-dimensional non-metric multidimensional scaling (NMDS) analysis of the pairwise genetic distance estimated using the GTRCAT nucleotide substitution model in RAxML.

Figure 4. Concatenated ML tree inferred in RAxML. with concatenated matrix (155 loci; 302,321 pb). Bootstrap support values >50% are indicated above the branches

Figure 5. Phylogenetic hypotheses of relationships among the different lineages of the *Q. laeta* complex and other species of the *Leucomexicana* clade. a) Species tree obtained by SVDQUARTETS. Bootstrap support values are indicated above the branches b) Species tree obtained by ASTRAL III . Local posterior probability values expressed as percentages are shown above branches

Figure 6. Phylogenetic networks as inferred by SNaQ from four different dataset. γ = inheritance probabilities

Figure 7. Reconstruction of the ecological niche models (ENM) for the six major lineages detected in the *Quercus laeta* complex. a) Binary distribution for *centralis TMVBc* lineage;

b) Binary distribution for *centralis TMVBe* lineage; c) Binary distribution for *laeta* SMS lineage; d) Binary distribution for *laeta* SMOc lineage; e) Binary distribution for *prinopsis* lineage; and f) Binary distribution for *transmontana* lineage.

Fig 8. Climatic niche differentiation between the major lineages identified in the *Quercus laeta* complex. Dotted line shown observed values of Schoener's D statistics versus values generated from 100 replicates (pink histograms). a) pairwise tests between *centralis* TMVBc and *centralis* TMVBe lineages; b) pairwise tests between *centralis* TMVBc and *laeta* SMS lineages; c) pairwise tests between *centralis* TMVBc and *laeta* SMOc lineages; d) pairwise tests between *centralis* TMVBc and *prinopsis* lineages; e) pairwise tests between *centralis* TMVBc and *transmontana* lineages; f) pairwise tests between *centralis* TMVBe and *laeta* SMS lineages; g) pairwise tests between *centralis* TMVBe and *laeta* SMOc lineages; h) pairwise tests between *centralis* TMVBe and *prinopsis* lineages; i) pairwise tests between *centralis* TMVBe and *transmontana* lineages, j) pairwise tests between *laeta* SMS and *laeta* SMOc lineages; k) pairwise tests between *laeta* SMS and *prinopsis* lineages; l) pairwise tests between *laeta* SMS and *transmontana* lineages; m) pairwise tests between *laeta* SMOc and *prinopsis* lineages; n) pairwise tests between *laeta* SMOc and *transmontana* lineages; o) pairwise tests between *prinopsis* and *transmontana* lineages.

Online Appendices

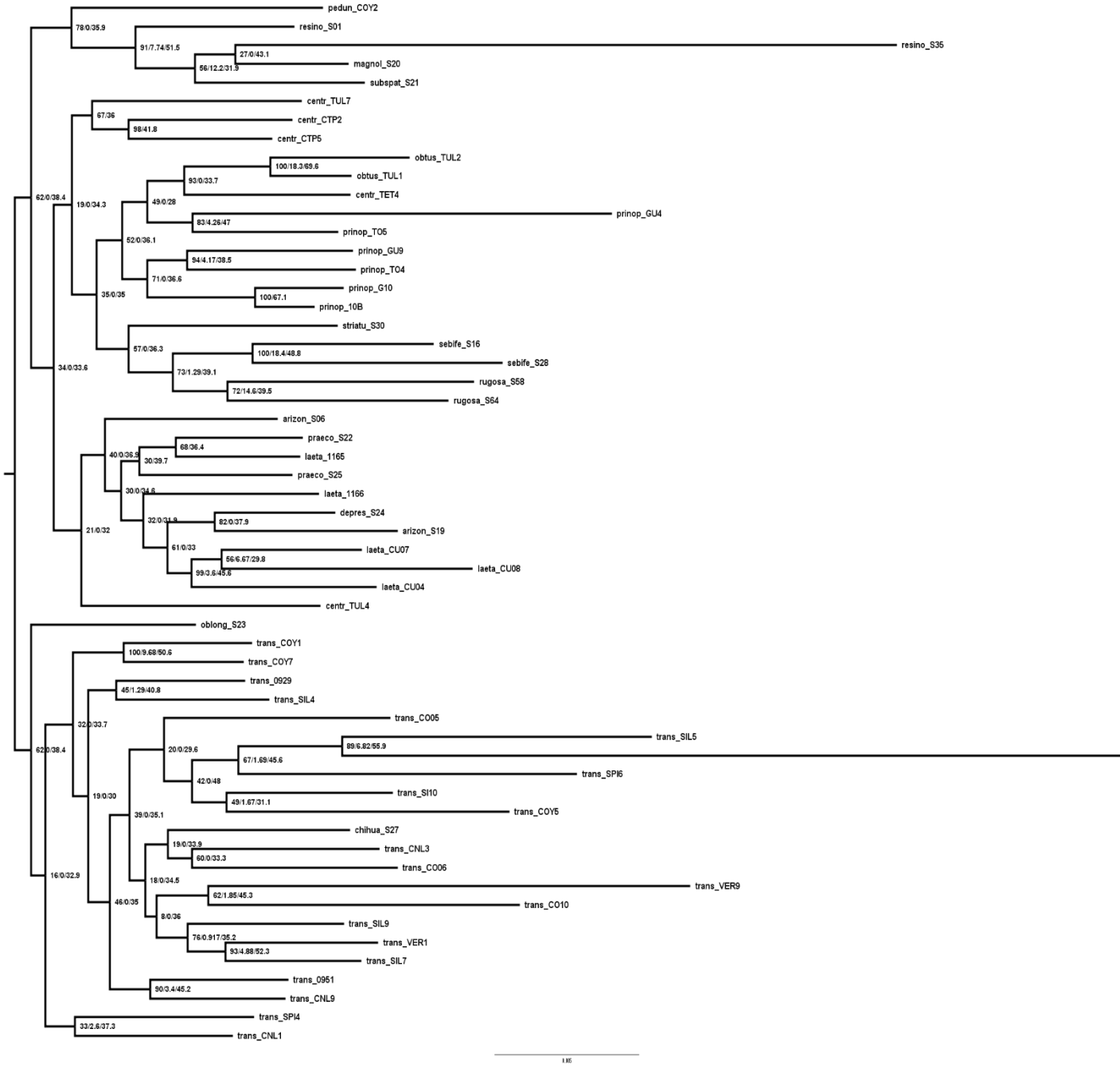
Phylogenomics and ecological niche contrasts lead to the identification of multiple and independent evolutionary lineages within the *Quercus laeta* complex (Fagaceae): new insights into the diversification of the *Leucomexicana* clade

Appendix 1. List of samples used in this study for *Q. laeta* complex and for species from the subsection *Leucomexicana* that were closely related to the *Q. laeta* complex with GPS coordinates of localities. Individual label includes abbreviation for species, site abbreviation, and individual number.

Species	Sample ID	Locality	Coordinates	
Q. laeta	centr_CTP2	Mexico: Michoacán: Contepec	19°59'45.60"	100°9'28.80"
Q. laeta	centr_CTP5	Mexico: Michoacán: Contepec	19°59'45.60"	100°9'28.80"
Q. laeta	centr_TUL4	Mexico: Puebla: waterfalls of Tuliman	19°52'8.40"	97°58'30.00"
Q. laeta	centr_TUL7	Mexico: Puebla: waterfalls of Tuliman	19°52'8.40"	97°58'30.00"
Q. laeta	centr_TET4	Mexico: Puebla: Tetela de Ocampo	19°48'0.00"	97°48'36.00"
Q. laeta	laeta_CU04	Mexico: Jalisco: Sierra El Cuale	20°22'40.80"	105°6'39.60"
Q. laeta	laeta_CU07	Mexico: Jalisco: Sierra El Cuale	20°22'40.80"	105°6'39.60"
Q. laeta	laeta_CU08	Mexico: Jalisco: Sierra El Cuale	20°22'40.80"	105°6'39.60"
Q. laeta	laeta_1165	Mexico: Durango: Canelas	25°07'33.6"	106°30'03.6"
Q. laeta	laeta_1166	Mexico: Durango: Canelas	25°07'33.6"	106°30'03.6"
Q. laeta	prinop_10B	Mexico: San Luis Potosi: Guadalcazar	22°38'31.20"	100°19'37.20"
Q. laeta	prinop_G10	Mexico: San Luis Potosi: Guadalcazar	22°38'31.20"	100°19'37.20"
Q. laeta	prinop_GU4	Mexico: San Luis Potosi: Guadalcazar	22°38'31.20"	100°19'37.20"
Q. laeta	pinop_GU9	Mexico: San Luis Potosi: Guadalcazar	22°38'31.20"	100°19'37.20"
Q. laeta	prinop_TO4	Mexico: San Luis Potosi: San Nicolás Tolentino	22° 4'30.00"	100°31'48.00"
Q. laeta	prinop_TO5	Mexico: San Luis Potosi: San Nicolás Tolentino	22° 4'30.00"	100°31'48.00"
Q. laeta	trans_CNL1	Mexico: Guanajuato: Cieneguilla	21° 2'34.80"	100°10'19.20"
Q. laeta	trans_CNL3	Mexico: Guanajuato: Cieneguilla	21° 2'34.80"	100°10'19.20"
Q. laeta	trans_CNL9	Mexico: Guanajuato: Cieneguilla	21° 2'34.80"	100°10'19.20"
Q. laeta	trans_CO05	Mexico: Aguascalientes: La Congoja	22° 9'32.40"	102°32'9.60"
Q. laeta	trans_CO06	Mexico: Aguascalientes: La Congoja	22° 9'32.40"	102°32'9.60"
Q. laeta	trans_CO10	Mexico: Aguascalientes: La Congoja	22° 9'32.40"	102°32'9.60"
Q. laeta	trans_COY1	Mexico: Oaxaca: Coyula	17°55'19.20"	96°55'51.60"
Q. laeta	trans_COY5	Mexico: Oaxaca: Coyula	17°55'19.20"	96°55'51.60"
Q. laeta	trans_COY7	Mexico: Oaxaca: Coyula	17°55'19.20"	96°55'51.60"
Q. laeta	trans_SIL4	Mexico: Guanajuato: Silao	21° 0'32.40"	99°6'57.60"
Q. laeta	trans_SIL5	Mexico: Guanajuato: Silao	21° 0'32.40"	99°6'57.60"
Q. laeta	trans_SIL7	Mexico: Guanajuato: Silao	21° 0'32.40"	99°6'57.60"
Q. laeta	trans_SIL9	Mexico: Guanajuato: Silao	21° 0'32.40"	99°6'57.60"
Q. laeta	trans_SI10	Mexico: Guanajuato: Silao	21° 0'32.40"	99°6'57.60"

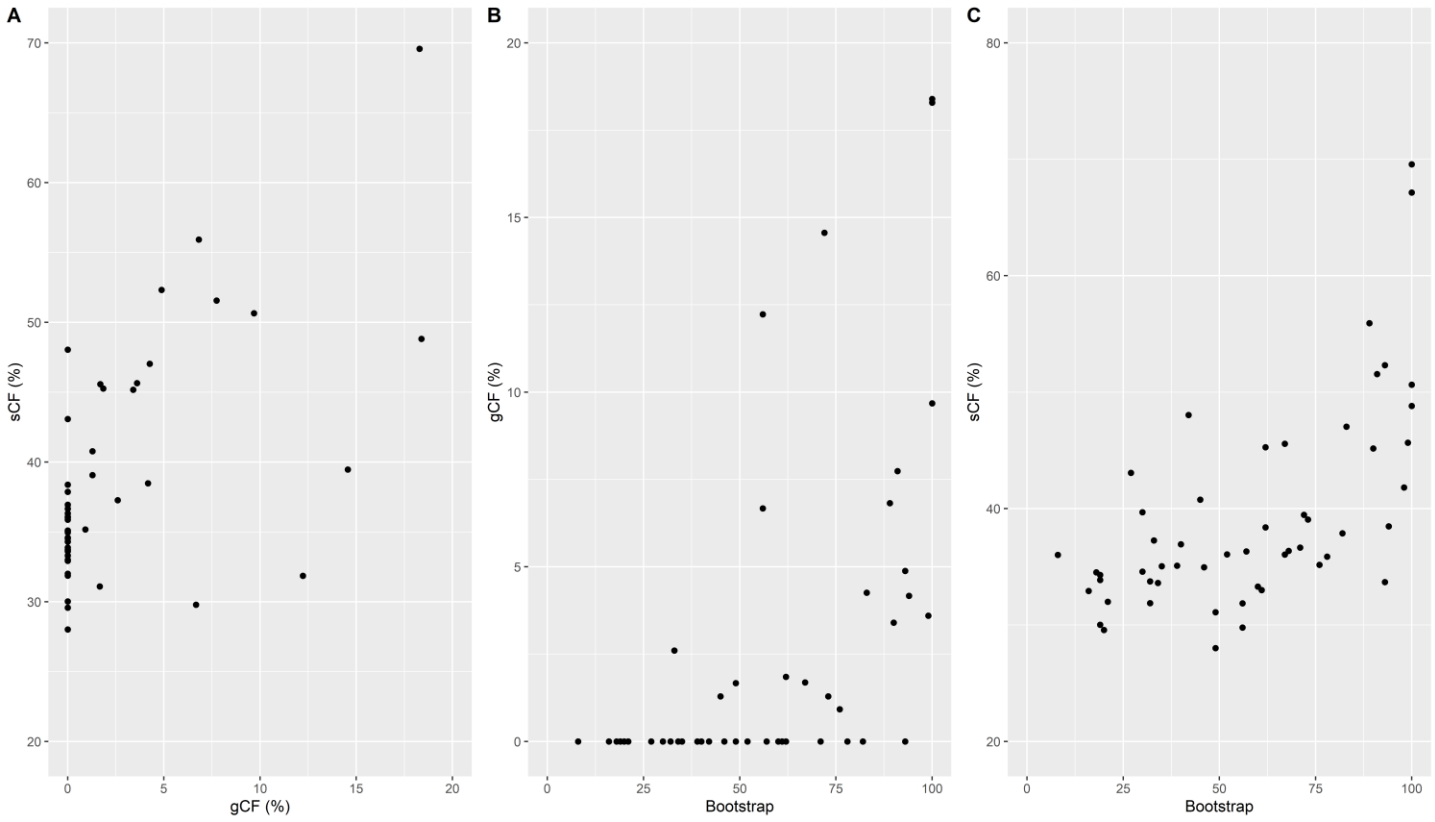
Q. laeta	trans_SPI4	Mexico: San Luis Potosi: San José Gallinas	22° 4'15.60"	100°13'15.60"
Q. laeta	trans_SPI6	Mexico: San Luis Potosi: San José Gallinas	22° 4'15.60"	100°13'15.60"
Q. laeta	trans_VER1	Mexico: Guanajuato: Vergel de Bernalejo	21°32'16.80"	100°16'44.40"
Q. laeta	trans_VER9	Mexico: Guanajuato: Vergel de Bernalejo	21°32'16.80"	100°16'44.40"
Q. laeta	trans_0929	Mexico: Durango: Mezquital	23°27'16.2"	104°21'16.3"
Q. laeta	trans_0951	Mexico: Nayarit: El Nayar	22°19'03.5"	104°24'42.1"
Q. arizonica	arizo _S06	Mexico: Durango: Tepehuanes	25°15'29.9"	105°52'12.0"
Q. arizonica	arizo _S19	Mexico: Durango: Tepehuanes	25°15'29.9"	105°52'12.0"
Q. chihuahuensis	chihua_S27	Mexico: Durango: Durango	23°56'19.7"	104°41'31.5"
Q. depressipes	depress_S24	Mexico: Durango: Mezquital	23°27'16.2"	104°21'16.3"
Q. magnoliifolia	magnol_S20	Mexico: Durango: Canelas	25°07'33.6"	106°30'03.6"
Q. oblongifolia	oblong_S23	Mexico: Durango: Santiago Papasquiaro	25°03'33.9"	105°17'39.3"
Q. obtusata	obtus_TUL1	Mexico: Puebla: waterfalls of Tuliman	19°52'8.40"	97°58'30.00"
Q. obtusata	obtus_TUL2	Mexico: Puebla: waterfalls of Tuliman	19°52'8.40"	97°58'30.00"
Q. peduncularis	pedun_COY1	Mexico: Oaxaca: Coyula	17°55'19.20"	96°55'51.60"
Q. peduncularis	pedun_COY2	Mexico: Oaxaca: Coyula	17°55'19.20"	96°55'51.60"
Q. praeco	praeco_S22	Mexico: Durango: Canelas	25°07'33.6"	106°30'03.6"
Q. praeco	praeco_S23	Mexico: Durango: Canelas	25°07'33.6"	106°30'03.6"
Q. resinosa	resino_S01	Mexico: Guanajuato: Vergel de Bernalejo	21°32'16.80"	100°16'44.40"
Q. resinosa	resino_S35	Mexico: Guanajuato: Vergel de Bernalejo	21°32'16.80"	100°16'44.40"
Q. rugosa	rugosa_S58	Mexico: Michoacán: Contepec	19°59'45.60"	100°9'28.80"
Q. rugosa	rugosa_S64	Mexico: Durango: Santiago Papasquiaro	25°03'33.9"	105°17'39.3"
Q. sebifera	sebife_S16	Mexico: Puebla: Tetela de Ocampo	19°48'0.00"	97°48'36.00"
Q. sebifera	sebife_S28	Mexico: Puebla: Tetela de Ocampo	19°48'0.00"	97°48'36.00"
Q. striatula	striatu_S30	Mexico: Zacatecas: Sombrerete	23°47'19.6"	103°48'33.1"
Q. subspathulata	subspat_S21	Mexico: Durango: Canelas	25°07'33.6"	106°30'03.6"

Online Appendix 2a Values of bootstrap / Gene concordance factor (gCF) / site concordance factor (sCF) calculated for ML concatenated tree.

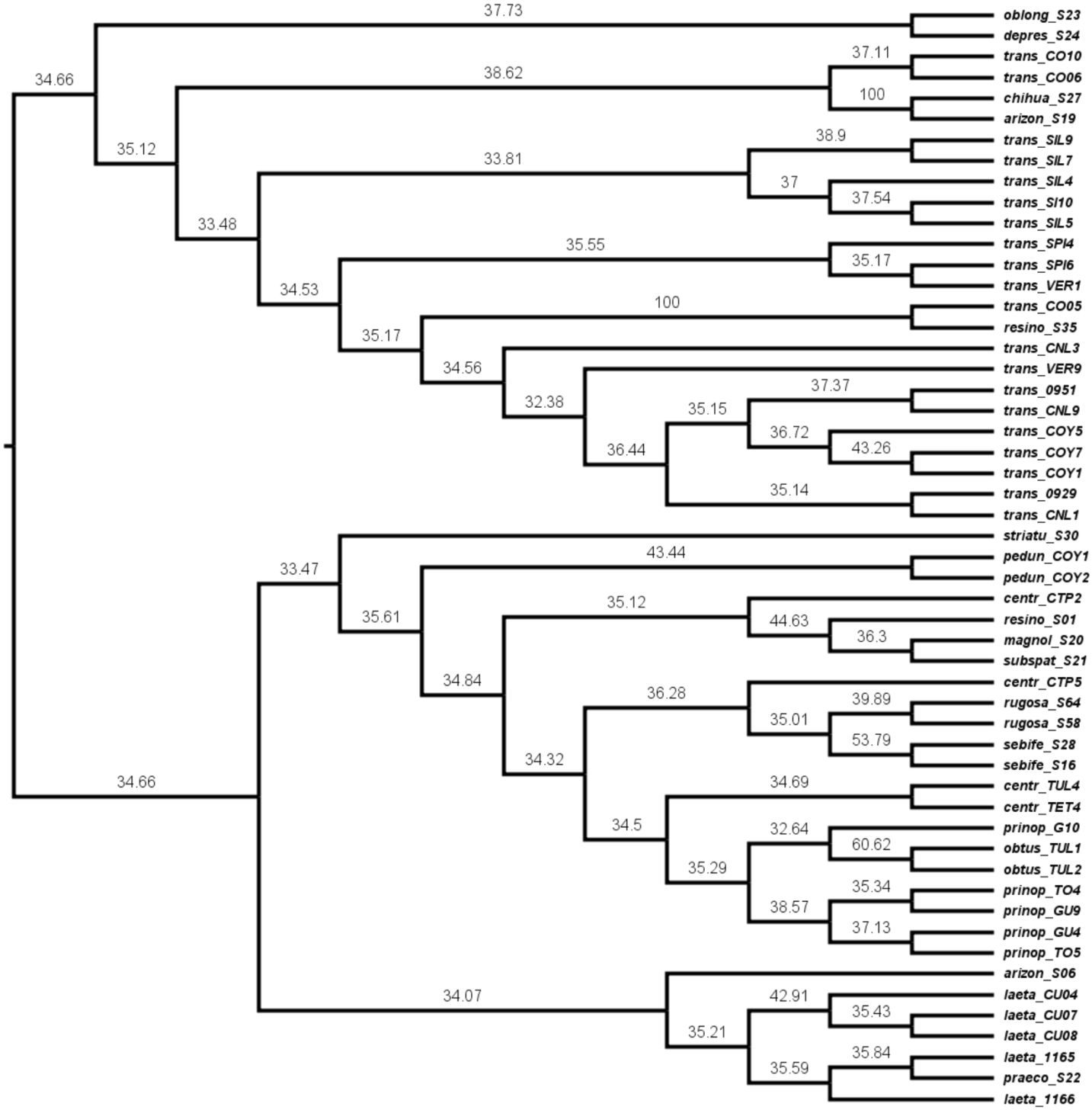


.

Online Appendix 2b A) Relation among site concordance factor (sCF) and gene concordance factor (gCF); B) Relation among gene concordance factor (gCF) and bootstrap support; C) Relation among site concordance factor (sCF) and bootstrap support. In all cases are not correlated, but BS is high for some low concordance factor and site concordance factor.



Online Appendix 2c. Species tree inferred from the ASTRAL analyses based on 155 loci. Quartet scores are displayed on the branches.

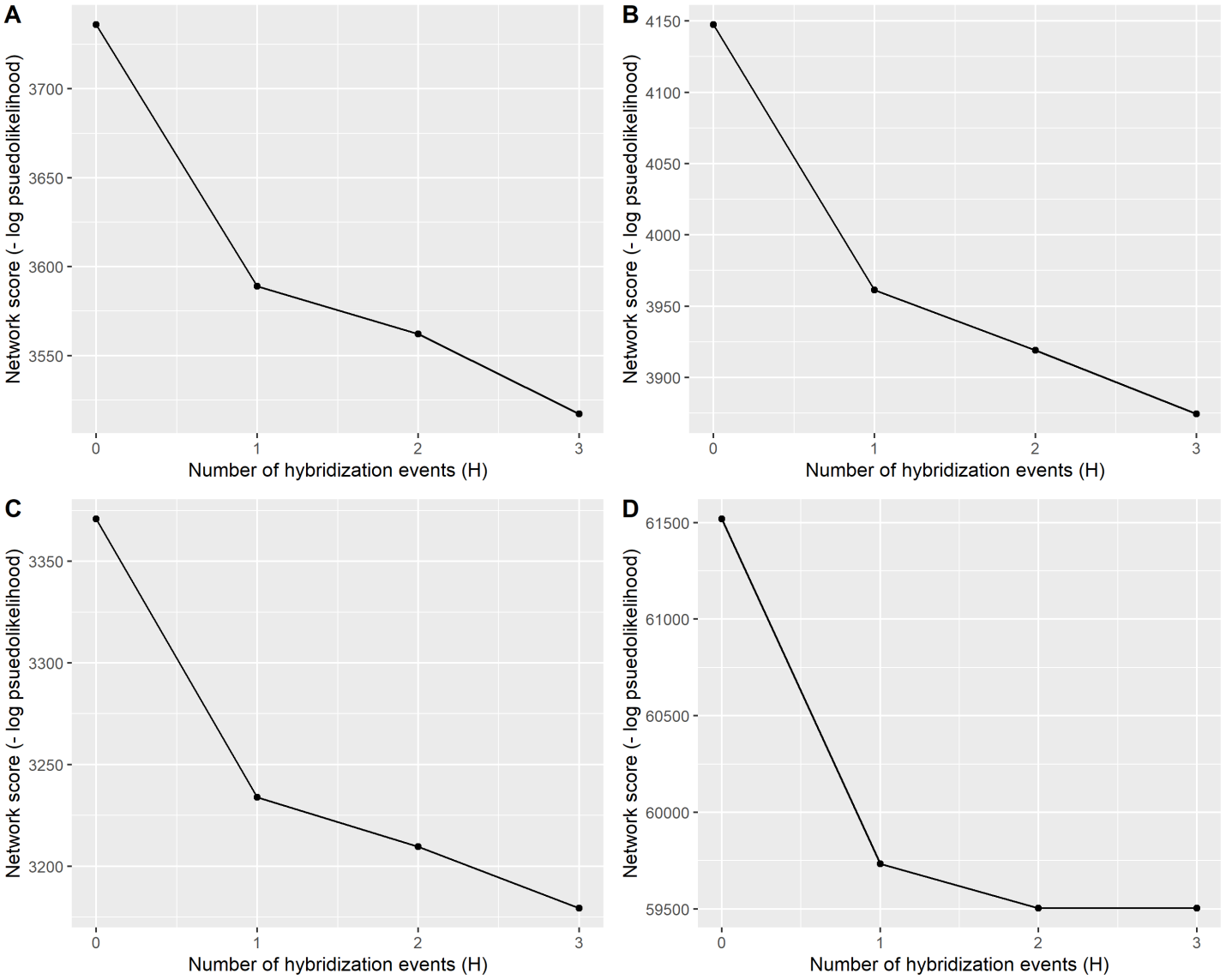


Online Appendix 3 The negative log psuedolikelihood for phylogenetic network estimation

h Group	0	1	2	3
Dataset1	3736.18	**3588.92**	3562.07	3517.19
Dataset2	4147.48	**3961.34**	3918.96	3874.48
Dataset3	3370.85	**3234.00**	3209.63	3179.40
Dataset4	61518.86	**59733.89**	59504.22	59504.22

Notes: h = number of hybridization events.

- Negative log pseudolikelihood profile for each number of hybridization events for each dataset inferred using *SNaQ*. The best-fitting number of hybridization events (*h*) is displayed as the value at which the rate of change in -log pseudolikelihood plateaus. A: Dataset1; B: Dataset2; C: Dataset3 and D: Dataset4.



online Appendix 4b. 1. Metrics of performance of selected model settings regarding regularization multiplier (RM), feature classes (F; l = linear, q = quadratic and p = product,), and sets of environmental variables, per each identified lineage within *Quercus laeta* complex

Model	Best models	p-value P. ROC	Mean AUC ratio	Omission rate (5%)	AICc	ΔAICc	Parameters
***centralis TMVBc* lineage**	RM_0.3_F_lq_set_03	0	1.726	0	181.798	1.044	2
	RM_0.4_F_lq_set_03	0	1.708	0	181.862	1.108	2
	RM_0.5_F_lq_set_03	0	1.699	0	181.943	1.189	2
	RM_0.6_F_lq_set_03	0	1.678	0	182.041	1.287	2
	RM_0.7_F_lq_set_03	0	1.658	0	182.157	1.403	2
	RM_0.8_F_lq_set_03	0	1.639	0	182.291	1.537	2
	RM_0.9_F_lq_set_03	0	1.646	0	182.443	1.689	2
	RM_1.1_F_lqh_set_03	0	1.642	0	181.033	0.279	2
	RM_1.3_F_lqh_set_03	0	1.651	0	181.658	0.904	2
	RM_1_F_lq_set_03	0	1.648	0	182.615	1.860	2
	RM_1_F_lqh_set_03	0	1.652	0	180.754	0	2
***centralis TMVBe* lineage**	RM_0.4_F_l_set_01	0	1.772	0	183.323	1.606	3
	RM_0.5_F_l_set_01	0	1.777	0	183.587	1.870	3
	RM_0.5_F_lq_set_01	0	1.791	0	181.717	0.000	3
	RM_0.6_F_lq_set_01	0	1.795	0	182.019	0.302	3
	RM_0.7_F_lq_set_01	0	1.793	0	182.241	0.524	3
	RM_0.8_F_lq_set_01	0	1.785	0	182.323	0.606	3
***laeta SMS* lineage**	RM_0.1_F_l_set_01	0	1.896	0	153.922	0	3
	RM_0.2_F_l_set_01	0	1.896	0	154.301	0.378	3
	RM_0.3_F_l_set_01	0	1.895	0	154.696	0.773	3
	RM_0.4_F_l_set_01	0	1.894	0	155.100	1.178	3
	RM_0.5_F_l_set_01	0	1.891	0	155.509	1.587	3
	RM_0.6_F_l_set_01	0	1.888	0	155.920	1.997	3
***laeta SMOc* lineage**	RM_0.1_F_l_set_02	0	1.932	0	252.462	0	6
	RM_0.2_F_l_set_02	0	1.926	0	252.994	0.531	6
	RM_0.3_F_l_set_02	0	1.931	0	253.806	1.343	6
***prinopsis* lineage**	RM_0.3_F_lq_set_02	0	1.946	0	380.668	0	6

	RM_0.2_F_lq_set_02	0	1.949	0	382.547	1.878	7
***transmontana* lineage**	RM_0.8_F_lqp_set_03	0	1.922	0	224.598	0.000	2
	RM_0.9_F_lqp_set_03	0	1.919	0	224.841	0.243	2
	RM_1.1_F_l_set_03	0	1.920	0	225.927	1.329	2
	RM_1.2_F_l_set_03	0	1.921	0	226.038	1.440	2
	RM_1.3_F_l_set_03	0	1.919	0	226.159	1.561	3
	RM_1.3_F_lqp_set_02	0	1.933	0	225.980	1.382	2
	RM_1_F_lqp_set_03	0	1.920	0	225.116	0.518	2

Online Appendix 4a. Set of Worldclim variables used for the construction of the ecological niche modelling for *Quercus laeta* lineages.

Lineage	Sample size	Variable set	Code
centralis TMVBc	7	Set 1	Bio 1, Bio 2, Bio 12, Bio14, Bio 17, Bio 19
		Set 2	Bio 4, Bio 8, Bio 13, Bio 15, Bio 18
		Set 3	Bio 3, Bio 7, Bio 9
centralis TMVBe	7	Set 1	Bio 1, Bio 2, Bio 6, Bio 12, Bio 14, Bio 15
		Set 2	Bio 4, Bio 9, Bio 13, Bio 15, Bio 18
		Set 3	Bio 4, Bio 7, Bio 9
laeta SMS	6	Set 1	Bio 1, Bio 14, Bio 15, Bio 19
		Set 2	Bio 3, Bio 4, Bio 9, Bio 10, Bio 16, Bio 19
		Set 3	Bio 7, Bio 09
laeta SMOc	10	Set 1	Bio 1, Bio 2, Bio 3, Bio 4, Bio 12, Bio 14, Bio 19
		Set 2	Bio 6, Bio 14, Bio 15, Bio 16, Bio 17, Bio 19
		Set 3	Bio 2, Bio 3, Bio 9, Bio 18, Bio 19
prinopsis	17	Set 1	Bio 1, Bio 2, Bio 4, Bio 7, Bio 12, Bio 14
		Set 2	Bio 2, Bio 4, Bio 7, Bio 14
		Set 3	Bio 3, Bio 4, Bio 7, Bio 9, Bio 14, Bio 18
transmontana	7	Set 1	Bio 1, Bio 2, Bio 3, Bio 4, Bio 12, Bio 15
		Set 2	Bio 2, Bio 3, Bio 4, Bio 6, Bio 15, Bio 18
		Set 3	Bio 3, Bio 7, Bio 8, Bio 19

IV. **CAPÍTULO III**

Micromorphological characterization as a taxonomical tool for identification of phylogenomic oak lineages: A case study of the *Quercus laeta* complex (*In Prep*.)

Micromorphological characterization as a taxonomical tool for identification of phylogenomic oak lineages: A case study of the *Quercus laeta* complex

Saddan Morales-Saldaña[1], Susana Valencia-Ávalos[2], Gonzalo Contreras-Negrete[3] and Antonio González-Rodríguez[1*]

[1]*Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Morelia, 58190 Michoacán, México*

[2]*Herbario de la Facultad de Ciencias, Departamento de Biología Comparada, Universidad Nacional Autónoma de México, Circuito Exterior, s.n., Ciudad Universitaria, Coyoacán, México City CP 04510, México*

[3]*Ciencias Agrogenómicas, Escuela Nacional de Estudios Superiores, Universidad Nacional Autónoma de México, León 37684, Mexico*

[*] Author for correspondence (agrodrig@cieco.unam.mx)

Abstract

Foliar micromorphological analysis has recurrently proven to be a valuable tool to assess relationships at different taxonomical levels in the genus *Quercus*. However, *Quercus* also is characterized by showing complicated taxonomical patterns resulting from frequent convergent evolution of vegetative characters, low differentiation among closely related species, and hybridization. Therefore, a purely qualitative approach may not be enough to document the micromorphological intraspecific and interspecific variation. In the present study, we used scanning electron microscopy to perform a qualitative and quantitative analysis of trichomes and foliar waxes for 48 individuals belonging to four species identified within *Quercus laeta* complex, with the goal of determining the value of these micromorphological traits in the taxonomy of the group. Type and density of trichomes, and trichome arm number and length significantly differentiated to *Q. centralis*, *Q. laeta*, *Q. prinopsis*, and *Q. transmontana*. A taxonomic key was prepared for the identification of these species. Our results provided useful information to correctly identify and discriminate among species of the *Quercus laeta* complex, by highlighting quantitative (continuous) rather than qualitative (categorical) differences, contrary to what has historically been used to discriminate among Mexican oak species.

KEYWORDS

North America, scanning electron microscopy, species discovery, trichome morphology.

Introduction

The considerable variability in foliar micromorphological characters present in flowering plants makes these traits ideal for research in fields such as taxonomy, ecology, physiology, and evolution (Adebowale et al. 2014; Deng et al. 2014; Mannethody and Purayidathkandy 2018; Stephan et al. 2018; Zafar et al. 2020). Particularly, micromorphological characters have been successfully employed as a tool to solve taxonomic problems across different plant lineages since the 20th century (Payne 1978; Stace 1984; Deng et al. 2014; Chen et al. 2015; Ecevit-Genç et al. 2017; Woodenberg et al. 2019; Siadati et al. 2020; Ullah et al. 2020).

Throughout history, the interspecific delimitation of oaks (*Quercus* L.) has challenged botanist (Canon and Petit 2019; Valencia-Á. 2021). In these efforts, foliar micromorphological analysis has recurrently proven to be a valuable tool to analyze relationships at different taxonomical levels (Tschan and Denk 2012; Deng et al. 2013; Deng et al. 2014) and, particularly, the characterization of trichome morphological features has been one of the most significant tools for the delimitation of oak species (Vázquez 2006; Scareli-Santos et al. 2007; Tschan and Denk 2012). Since Dyal (1936) and Camus (1936–38), a functional classification based on glandular and non-glandular trichomes was established. Subsequently, with the introduction of the scanning electron microscope (SEM), several studies have contributed to the classification of different and new types of trichomes (Hardin 1976, 1979; Stace 1984; Jones 1986; Vázquez 2006; Tschan and Denk 2012), as well as to the delimitation of Asian (Kim et al. 2011; Panahi et al. 2012; Deng et al. 2015; Deng et al. 2017), European (Bussotti and Grossoni 1997; Tschan and Denk 2012; Fortini et al. 2015), and North American oak species (Dyal 1936; Hardin 1979; Manos 1993; Scareli-Santos et al. 2007; Scareli-Santos et al. 2013). However, although trichomes

alone provide relatively stable character sets, an integral analysis of micromorphological leaf traits including qualitative character states of trichomes, waxes and stomata, as well as quantitative measurements of trichome arm length and stomata width and length, has provided increased resolution for species delimitation (Fortini et al. 2009; Panahi et al. 2012; Tschan and Denk 2012; Lopez-Camaal et al. 2017).

For Mexican oaks, important taxonomic confusion still exists mainly within the section *Quercus* (Valencia-Á. 2004) because intergradation in macromorphological characters is common in this group, resulting in species complexes. Therefore, the effort to establish clearer limits among taxa requires the inclusion of additional traits and complementary approaches. An example of this situation is *Quercus laeta*, a taxon with a long history of controversy (McVaugh1974; Romero et al. 2002; Valencia-Á. 2004; Hipp et al. 2018; Morales-Saldaña et al. 2022). *Quercus laeta* is endemic to Mexico but with a wide geographic distribution in different habitat types. Even so, according to previous authors, *Q. laeta* individuals are considered members of the same based on characters such as the lack of glandular trichomes and the presence of papillae and sessile slightly contortous trichomes on the leaf underside. However, there is considerable variation in pubescence degree on the underside of leaves, lamina shape and pubescence of twigs, being one of the oak species with a large polymorphism (Valencia-Á. 2004; Morales-Saldaña et al. 2022).

Recently, it has been suggested that *Quercus laeta* is not a highly polymorphic species but a complex of morphologically similar species and with weak reproductive barriers among its members (Morales et al., 2022; Morale et al., In. prep). This is supported by several phylogenetics studies, which show evidence that *Q. laeta* populations are not monophyletic, recognizing the existence at least four species within the *Q. laeta* complex (called *Q. centralis*, *Q. laeta*, *Q. prinopsis*, and *Q. transmontana*), each restricted

geographically and ecologically (Hipp et al. 2018, 2020; Morales-Saldaña et al. in prep.). Although these previous studies have demonstrated the existence of distinct biological entities within the *Q. laeta* complex, these species have not been thoroughly characterized morphologically and, in particular, a detailed analysis of micromorphological differentiation is required so their taxonomic status can be better understood.

In this context, we present a detailed quantitative study of trichome morphology and waxes across *Quercus laeta* complex. The objectives of this study were to: 1) assess qualitative and quantitative trichome variability among species of the *Q. laeta* complex using scanning electron microscopy; and 2) test the applicability of micromorphological characters and determine the most useful combination of micromorphological characters for the identification of species within *Q. laeta* complex (type and density of trichomes, variation in the length of trichome arms, and wax types)

Material and methods

Plant material and scanning electron microscopy (SEM)

Leaf samples were collected from natural populations covering almost the complete geographical distribution of the *Q. laeta* complex (Fig. 1). A total of 48 individuals including 13 samples representing to *Q. centralis*, six of the *Q. laeta*, 13 of the *Q. prinopsis* and 16 of the *Q. transmontana* were analyzed. In all cases, mature leaves were collected during the fruiting period (August-November) with no visible damage. After collection, leaves were dried at room temperature and conserved in plastic bags. The voucher specimens were deposited at MEXU and FCME herbaria.

For scanning electron microscopy (SEM), the middle region of the leaves was cut between the midrib and the margin into pieces of approximately 1 cm^2. Subsequently, the pieces were fixed on metal stubs using double sided adhesive without further treatment.

Micromorphological features were observed on the abaxial surface and Scanning Electron Microscopy (SEM) micrographs were taken using a JEOL JSM-IT300 electron microscope (Microscopy Laboratory ENES UNAM MORELIA) .

Description of types of trichomes and wax types

A brief description was made for the different trichome types observed on the abaxial surface using the terminology and classification of Hardin (1976, 1979). The nomenclature proposed by Barthlott et al. (1998) was followed for description of wax types.

Micromorphological quantitative comparison among species

Length of trichome arms (µm), number of arms, and trichome density were analyzed on SEM images using the ImageJ software (Rasband 1997–2018). For compound trichomes (fasciculate and stellate), a maximum of eight arms per trichome were measured from their attachment point (excluding the pedestal in fasciculate trichomes) or the center of the central disc (in stellate trichomes) to the distal arm end. Also, we counted the number of arms for both fasciculate and stellate trichomes. To determine trichome density on the abaxial leaf surface, we divided the micrographs into grids of 0.05 mm^2 and counted the number of glandular and non-glandular trichomes in each grid (five grids per sample). A total of 978 trichome arms and 573 grids were analyzed.

To compare the distribution of the numeric variables among species, first we visualized the data with violin graphs using ggplot2 R package version 4.2.1 (R Development Core Team 20202). Violin graphs have the advantage of allowing visualization of the probability density of the data at different values by a kernel density estimator. Second, conformance to a normal distribution was assessed for all variables using the Shapiro-Wilk goodness of fit test. Once the normality of the data was confirmed, we conducted an analysis of variance (ANOVA). Subsequently a Tukey post hoc test was

used to compare mean values of trichome density, number of arms, and length of arms among the four species identified within *Q. laeta* complex. These analyses were performed using R version 4.2.1 (R Development Core Team 2022). To identify the traits with higher contribution to the separation of species, a principal components analysis (PCA) and a canonical discriminant analysis (CDA) were carried out, using the individual mean values of each trait evaluated (i.e., length of trichome arms, number of trichome arms, and density of glandular and non-glandular trichomes). Only individuals that had metrics for all traits were included. A multivariate analysis of variance (MANOVA) was carried out to test the differences of the set of traits at the lineage level; a $p < 0.05$ was considered to be significant. The analysis was performed using the 'stat', "ggbiplot", "MASS', and "candisc" libraries in R version 4.2.1 (R Development Core Team 2022). Finally, on the basis of all the qualitative and qualitative traits evaluated, a taxonomic key was proposed for the species identification.

Results

Qualitative description of trichome and wax types

We identified three different trichome types, simple uniseriate (glandular), fasciculate and stellate (non-glandular), in the four species of the *Q. laeta* complex (Table 1; Fig. 2) following Hardin (1976, 1979). We observed three different wax types among the samples (crystalloids, wax crusts and wax layer) (Table 1; Fig. 3) *sensu* Barthlott et al. (1998).

Glandular trichomes

Simple uniseriate (Fig. 2; A- D). Trichomes of this type consist of a single column of two or more thin-walled cells, turgescent in the juvenile state, in herbarium specimens usually collapsed (because of loss of the cellular content).

Occurrence. This trichome type was found rarely on the abaxial leaf surfaces of *Q.*

prinopsis and *Q. laeta* and commonly on the abaxial leaf surfaces of *Q. centralis* and *Q. transmontana*.

Non-glandular trichomes

Trichomes of this group are generally thick-walled, multicellular, and compound. The function of these trichomes appears to be related to physical protection (Jones 1986). Within this category we can find the fasciculate and stellate trichomes.

Fasciculate (Fig. 2; D-H). Multicellular and compound trichomes, arranged in tufts of (2) 6 to 8 (12) thick-walled arms that originate from a common point. Fasciculate tufts occur in a sessile form with arms slightly curled down and then twisted, individual arms acicular or subulate; in the basal part groups of arms or all arms fused, in the distal part splaying out like a bouquet of flowers, those with very long elements often intertwined; surface of the arms is nearly smooth.

Occurrence. Fasciculate trichomes occur on the abaxial surface of all investigated species. This trichome type occurs both on veins and in intercostal areas.

Stellate (Fig. 2; I-L). Multicellular, compound trichomes, usually with 4–8 (15) subulate to acicular arms radiating from a common origin adpressed to the epidermis, hair base absent, but attachment point slightly rising above the surrounding epidermis, stellate trichomes are usually persistent.

Occurrence. Stellate trichomes occur on the abaxial side of the leaf in *Q. centralis, Q. laeta*, and *Q. prinopsis*. They are found predominantly in the intercostal regions, only occasionally on the veins. In *Q. prinopsis*, the stellate trichomes are usually hidden below the fasciculate trichomes, while in *Q. centralis* and *Q. laeta* they are not hidden below the fasciculate trichomes.

Wax types. Wax of the crust type was observed on fasciculate trichomes of all species (Fig. 3A-B). Wax layers, which are defined as continuous coverings usually less than 1 μm thick without a prominent surface sculpturing, were also observed on the abaxial surface of all species (Figure 3E-F). Crystalloid waxes were observed exclusively on the abaxial surfaces of the *Q. transmontana* and *Q. laeta* (Table 1; Fig. 3C-D).

Quantitative morphological comparison

A summary of quantitative leaf epidermal micromorphological features for all species is presented in Table 2.

Length variation of trichome arms. Arm length of fasciculate and stellate trichomes differed, being longer in fasciculate than in stellate trichomes (Fig. 4A). *Quercus prinopsis* showed the largest variability for the arm length of stellate trichomes, with most arms ranging between 104.19 – 168.55 μm with a median of 135.01 and mean of 139.09 μm. In contrast, *Q. centralis* (92.79 – 135.33 μm with a median of 117.42 and mean of 116.37 μm) and *Q. laeta* (100.36 – 133.86 μm with a median of 114.43 and mean of 118.30 μm) showed less variability (Fig. 4B; Table 2). The ANOVA and the Tuckey post hoc test revealed significant differences ($F_{2, 429} = 18.80$; $p < 0.005$) in the length of stellate trichome arms between *Q. centralis* and *Q. prinopsis* but not between the *Q. centralis* and *Q. laeta* (Table 3).

For the arm length of fasciculate trichomes, violin plots indicated a great dispersion of the data (Fig. 4C). *Quercus prinopsis* showed generally larger values than the rest of the species, with arms ranging in length from 225.84 – 395.23 μm with a median of 318.26 and mean of 313.20 μm. *Quercus transmontana* also showed large values although with a bimodal distribution, with arm length between 236.4 – 371.9 μm, a median of 300.3 μm, and mean of 307.2 μm. In contrast, *Q. centralis* and *Q. laeta* showed a tendency to shorter

trichomes. In *Q. centralis* these ranged between 171.74 – 306.92 μm with a median of 234.91 and mean of 250.68 μm, although with outliers above 500 μm. For *Q. laeta*, most of the values ranged between 154.78 – 241.80 μm with a median of 188.01 and mean of 205.41 μm, with outliers between 400 – 500 μm (Table 2; Fig. 4C). The ANOVA test revealed significant differences ($F_{3,514} = 12.07$; $p < 0.00001$) in the length variation of the arms of fasciculate trichome among species and the Tukey post hoc test revealed significant differences ($p < 0.005$) between pairwise comparisons of *Q. laeta* – *Q. centralis*, *Q. centralis* – *Q. prinopsi*s, *Q. centralis* – *Q. transmontana, Q. prinopsis* – *Q. laeta* and *Q. laeta* – *Q. transmontana*, but not between *Q. prinopsis* – *Q. transmontana* (Table 3).

Variation in the number of trichome arms. Violin plots indicated a tendency for stellate trichomes to present a greater number of arms compared to fasciculate trichomes (Fig. 5A). In the *Q. laeta*, stellate trichomes can have up to 15 arms, but are more likely to have between 7 and 8 arms (Fig. 5B). Both *Q. centralis* and *Q. prinopsis* had a high frequency of trichomes with 7 and 8 arms, though *Q. prinopsis* had stellate trichomes with up to 11 arms. In relation to fasciculate trichomes, both *Q. centralis* and *Q. prinopsis* had a higher frequency of trichomes with 7 or 8 arms. In *Q. laeta*, though most of the trichomes had 8 arms, some had up to 12. Finally, *Q. transmontana* showed the greatest variability with fasciculate trichomes having between 2 – 9 arms, though with a higher frequency of trichomes with 7 or 8 arms (Figure 5C). The ANOVA test revealed significant differences ($F_{2,197} = 31.03$; $p < 0.00001$) in the number of arms of stellate trichomes among species and the Tukey post hoc test revealed significant differences ($p < 0.005$) between *Q. centralis* – *Q. laeta* and *Q. laeta* – *Q. prinopsis*, but not between *Q. centralis* – *Q. prinopsis* ($p > 0.50$) (Table 4). For fasciculate trichomes, the ANOVA test revealed significant differences ($F_{2,197} = 31.03$; $p < 0.00001$) and the Tukey post hoc test showed significant differences (*p*

< 0.005) in all pairwise comparisons except between *Q. transmontana* – *Q. prinopsis* (Table 4).

Glandular and non-glandular trichome density. The violin plots indicated high non-glandular trichome density in *Q. laeta* and *Q. prinopsis*, although in both cases with a multimodal distribution, while *Q. centralis* and *Q. transmontana* showed a low density of non-glandular trichomes (Fig. 5A). The ANOVA test revealed significant differences ($F_{3,534}$ = 85.3; $p < 0.00001$) and the Tukey post hoc test revealed significant differences ($p <$ 0.005) in non-glandular trichome density in all pairwise comparisons, except between *Q. transmontana* and *Q. centralis* (Table 5). In contrast, the violin plot for glandular trichome density showed higher values in *Q. centralis* and *Q. transmontana*, with a larger spread of data in *Q. centralis* (Fig. 5B). In the case of *Q. laeta* and *Q. prinopsis*, glandular trichome density was lower but with some outliers. The ANOVA test showed significant differences ($F_{3,534}$ = 149.4; $p < 0.0001$) and the Tukey post hoc test indicated significant differences (p < 0.0001) in glandular trichome density between all pairwise comparisons (Table 5).

The first two axes of the PCA explained 63% and 19.1% of the variation among traits for the four species (Fig. 7). For the PC1 a marked separation among the species was found, where an increase in loads of variation in the number of trichome arms (NTA) and trichome non-glandular density (TNGD) grouped the *Q. laeta* and *Q. prinopsis* on the positive side of the axis, as well as an increase in trichome glandular density (TGD) and length variation of trichome arms (LTA) loads segregated the *Q. centralis* and *Q. transmontana* on the negative side of the axis.

The CDA resulted in a reliable classification of specimens of the differents species (Fig. 8), in agreement with the results obtained in the PCA. The first two axes accounted for 95% of the variation. The variables that contributed most to the first canonical axis

(85.7%) were the trichome glandular density (TGD) and trichome non-glandular density (TNGD), while characters that best discriminated along canonical axis two (10.2%) were number of trichome arms (NTA) and length variation of trichome arms (LTA) (Pillai´s trace = 1.4, $p < 0.001$, 85.8% of variation explained) (Table 6). The confusion matrix for the CDA showed for *Q. centralis* a rate of 83% of cases correctly classified, but 16% misclassified as *Q. transmontana*. In the case of *Q. laeta* showed a predicted rate of 100%, while *Q. prinopsis* reached a 90% of correctly classified cases, but 9% of cases were assigned as *Q. transmontana*. Finally, for *Q. transmontana* 69% of the cases were correctly assigned, while 23% and 7% were misclassified as *Q. centralis* and *Q. prinopsis*, respectively (Supplementary table 1).

Discussion

The value of micromorphological characters, especially trichomes, for taxonomy and systematics across different plant groups have been emphasized by many authors (Gharemaninejada et al. 2012; Adebowale et al. 2014; Ali et al. 2020; Arabameri et al. 2020; Gissi et al. 2022). For the genus *Quercus*, nearly all authors consider characters of the indumentum to be important for species delimitation (Hardin 1976, 1979; Valencia and Delgado 2003; Vázquez 2006; Schand and Denk 2012; Scareli-Santos et al. 2013). However, *Quercus* also is characterized for showing complicated taxonomical patterns (Manos et al. 1999; Valencia-Á. 2004; Petit et al. 2004; Valencia 2021; Morales-Saldaña et al. 2022,) resulting from frequent parallel or convergent evolution of vegetative characters, pronounced intraspecific variation, low differentiation among closely related species, and hybridization (Tucker 1974; Baquedano et al. 2008; McVay et al. 2017; Kim et al. 2018; Crowl et al. 2020; Valencia 2021). Furthermore, if we consider that one or a few specimens are generally used for the characterization of species, a purely qualitative approach could be

not enough to document intraspecific micromorphological variation, leading to potential taxa misidentification. For these reasons a quantitative complementary approach may be useful, even in cases of species that are difficult to distinguish based on macromorphological characters (Vázquez 2006; Scareli-Santos et al. 2013). In this context, the present SEM study has provided, useful information to discriminate among different species within the *Quercus laeta* complex, by highlighting quantitative (continuous) rather than qualitative (categorical) differences, contrary to what has historically been used to discriminate among Mexican oak species (Liebmann 1854; McVaugh 1974; Romero et al. 2002).

Hardin's studies (1976; 1979) to characterize the variety of trichomes in North American oaks has undoubtedly been a pillar in the field of oak micromorphology. However, despite the effort to characterize the micromorphology of Mexican oak species (Spellenberg and Bacon 1996; Valencia and Salinas 2003; Vázquez 2006; Scareli-Santos et al. 2013), there have been few studies focused on evaluating inter- and intraspecific variation under a statistical framework (Scareli-Santos et al. 2007; López-Caamal et al. 2017; Sánchez-Acevedo et al. 2022). This lack of statistical rigor could cause taxonomic difficulties in certain species complexes, as is the case for *Quercus laeta* complex. In this context, the original description by Liebmann (1854) originally described a densely tomentose abaxial surface with the presence of stellate trichomes as diagnostic features to distinguish *Q. laeta*. However, several regional floras have used different character combinations to identify *Q. laeta* (McVaugh 1974; González-Villarreal 1986; Romero 2002, Romero et al. 2014; Pérez and Valencia-Á. 2017). Our results indicate that all species share several micromorphological traits, such as the presence of papillae, fasciculate and glandular trichomes, as well as different types of wax on abaxial surface (Table 1).

However, it was also possible to detect certain consistent micromorphological traits capable of discriminating among species that are independent evolutionary units according to previous population genetics and phylogenetic studies (Morales-Saldaña et al. 2022, Morales-Saldaña *in. prep.*). In this sense, the presence or absence of stellate trichomes and the density of both glandular and non-glandular trichomes should be considered key features to discriminate among species. Despite it having been shown that trichome density can also be strongly influenced by the environment (Hernandez and Park 2022; Sánchez-Acevedo et al. 2022), our results showed a widespread association between patterns of differentiation of micromorphological traits and phylogenetic assignment, indicating that there is a consistent pattern.

Since Liebmann (1854), the presence of stellate trichomes has only been commented by Romero (2002) as one of the features to consider in the identification of *Q. laeta*, but specimens examined by her were collected in the Trans-Mexican Volcanic Belt, so probably belonged to what we have recognized as *Q. centralis* (Morales-Saldaña et al. in prep). In contrast, other taxonomists, using samples from other geographic regions in Mexico, have not reported the presence of stellate trichomes (McVaugh 1974; González-Villarreal 1986; Romero et al. 2014). In that sense, according to our results, the absence of stellate trichomes on the abaxial leaf surface immediately discriminates to *Q. transmontana* from the other species. Also, it was possible to identify other distinctive traits, such as the presence of crystalloid waxes, significantly longer arms of the fasciculate trichomes (236.4 – 371.9 μm) in comparison to those observed in *Q. centralis* and *Q. laeta*, and a low density of fasciculate trichomes in comparison to *Q. laeta* and *Q. prinopsis*. This variation, together with the macromorphological morphometric differentiation presented by Morales-Saldaña et al. (2022), confirms that this is the most distinct lineage, supporting the hypothesis that

Q. transmontana should be considered as an independent species (Morales-Saldaña et al. 2022, in prep.).

Although glandular, stellate, and fasciculate trichomes were found in *Q. centralis*, *Q. laeta*, and *Q. prinopsis*, *Q. centralis* was characterized by an abaxial surface with a low density of fasciculate and stellate trichomes, but with a high density of simple uniseriate trichomes (glandular trichomes), distinguishing it from the other two species. Furthermore, arms of the fasciculate trichomes were significantly longer (171.94 – 306.92 μm) in *Q. centralis* than in *Q. laeta* (154.78 –241.80 μm) but significantly shorter than in *Q. prinopsis* (225.84 – 395.23 μm). For the recognition of *Q. prinopsis* and *Q. laeta*, PCA and CDA analysis suggested that glandular trichome density and the arm number of the trichomes are the two most important variables to differentiate between species. In such a way *Q. prinopsis* is characterized by a high density of non-glandular trichomes as well as stellate and fasciculate trichomes with significantly longer arms than *Q. laeta*, which showed a higher arm number. Another difference is the presence of crystalloid waxes in *Q. laeta* which are absent in *Q. prinopsis*. Therefore, quantitative differences are key to discriminate between these species as occurs in other oak groups (Bussotti and Grossoni 1997). Although Tschan and Denk (2012) considered that quantitative data should be used with care for taxon delimitation, especially in cases of overlapping variability, as is the case in some species investigated here, the multiple samples representing different populations and the results of statistical tests suggest that micromorphology is capable of differentiating among species and, therefore, should serve as a tool in the correct identification of these species.

In this sense, our results allow us to interpret phenotypic variation as a way to buttress the results obtained from phylogenetic analysis (Morales-Saldaña et al. in prep.),

since ignoring phenotypic differentiation may leave us with units that are little more than historical constructs without meaning for biodiversity (Zapata and Jiménez 2012; Freudenstein et al. 2017). In this framework, our study emphasizes pairwise comparisons of quantitative characters among putative species (lineages) under the hypothesis that a bimodal (or polymodal) distribution of morphological variation suggests the existence of more than one species (Futuyma 1998; Zapata and Jiménez 2012). The probability density function illustrated by violin plots allowed us to observe multiple modes in the distribution of the quantitative characters, suggesting a certain degree of micromorphological divergence among species in the analyzed features, which was statistically supported by pairwise comparisons and multivariate analyses. However, violin plots also showed bimodal distribution within some species, suggesting that even within the species there is an important micromorphological variation.

Since phenotypic discontinuities can result from geographic differentiation within a single species (Stephan et al. 2018; Maya-García et al. 2020), description of new species based only on morphological variation could result in taxonomic inflation; so that, an evolutionary framework is necessary to explain certain patterns found in this work. On the one hand, according to phylogenetic studies carried out by Hipp et al. (2018; 2020) and Morales-Saldaña et al. (in prep.), *Q. laeta* complex are a non-monophyletic group, which suggests that the features they share, such as the presence of papillae, fasciculate and glandular trichomes as well as the presence of wax crusts and layers on the abaxial surface do not necessarily indicate close taxonomic affinity, and that they could be parallelisms (Tucker 1974; Struck et al. 2018), so they should not be considered to diagnose different species. On the other hand, micromorphological differences found in this study provide a morphological identity that these evolutionarily independent lineages have acquired

throughout the speciation process.

Conclusions

The use of SEM to qualitatively and quantitatively analyze micromorphological characteristics helped to correctly identify and discriminate among species of the *Quercus laeta* complex previously proposed by phylogenetic data. Our study stresses the importance of stellate trichomes and trichomes density as key features that in conjunction with the arm length and arm number variation provide reliable taxonomic characters that can serve as an additional tool in the correct identification of these species. In addition to this, we considered that the characteristics proposed here to recognize the different species are consistent across multiple geographical samples suggesting that they not change with environmental variation that in combination with approaches taken up in this work could serve to reduce the ambiguity among other oaks species complexes.

Taxonomic key based on leaf micromorphological characters to species identification

1. Abaxial surface with stellate trichomes absent……..... *Q. transmontana*

1. Abaxial surface with stellate trichomes present

2. Abaxial surface with a low density of fasciculate trichomes, but with abundant (high density) simple uniseriate trichomes (glandular trichomes) …… ……...…. *Q. centralis*

2. Abaxial surface with a high density of fasciculate trichomes but with a low density of simple uniseriate trichomes (glandular trichomes)

3.- Stellate and fasciculate trichomes significantly longer (225.84 – 395.23 μm), wax platelets absent on the adaxial surface ...……..……..... *Q. prinopsis*

3 Stellate and fasciculate trichomes significantly shorter (154.78 – 241.80 μm), wax platelets present on the adaxial surface ……...……..……..… *Q. laeta*

Acknowledgments

Saddan Morales-Saldaña thanks Posgrado en Ciencias Biológicas—UNAM and Consejo Nacional de Ciencia y Tecnología (CONACYT) for the PhD scholarship and financial support (scholarship/CVU: 483720/624511). This study constitutes a partial fulfillment of the Graduate Program in Biological Sciences of the National Autonomous University of Mexico (UNAM)..

Declaration of interest statement

The authors reported no potential conflict of interest.

ORCID

Saddan Morales-Saldaña https://orcid.org/0000-0001-6844-356X

Susana Valencia-Ávalos

Gonzalo Contreras-Negrete https://orcid.org/0000-0003-3713-4638

Antonio González-Rodríguez https://orcid.org/0000-0002-6196-7288

References

Adebowale A, Naidoo Y, Lamb J. Nicholas A. 2014. Comparative foliar epidermal micromorphology of Southern African *Strychnos* L. (Loganiaceae): taxonomic, ecological, and cytological considerations. Plant Syst. Evol. 300: 127-138.

Ali M, Bahadur S, Hussain A, Saeed S, Khuram I, Ullah M, ... & Akhtar N. 2020. Foliar epidermal micromorphology and its taxonomic significance in *Polygonatum* (Asparagaceae) using scanning electron microscopy. Microsc. Res. Tech. 83(11): 1381-1390.

Arabameri M, Khodayari H, Zarre S. 2020. Trichome micromorphology in *Alcea* L. and allied genera (Malvaceae) and its systematic implication. Nord. J. Bot. 38(6): 1-16.

Barthlott W, Neinhuis C, Cutler D, Ditsch F, Meusel I, Theisen I, Wilhelmi H. 1998. Classification and terminology of plant epicuticular waxes. Bot. J. Linn. Soc. 126(3): 237-260.

Baquedano FJ, Valladares F, Castillo FJ. 2008. Phenotypic plasticity blurs ecotypic divergence in the response of *Quercus coccifera* and *Pinus halepensis* to water stress. Eur. J. For. Res. 127(6): 495-506.

Bussottii FP, Grossoni P. 1997. European and Mediterranean oaks (*Quercus* L.; Fagaceae): SEM characterization of the micromorphology of the abaxial leaf surface. Bot. J. Linn. Soc. 124 :183-199.

Camus A. 1934-1938. Monographie du genre *Quercus*.Vol 1. Sous-genre Cyclobalanopsis. Sous-genre Euquercus, sections Cerris et Mesobalanus. Academie des Sciences. Lechevalier, Paris.

Cannon CH, Petit RJ. 2020. The oak syngameon: more than the sum of its parts. New Phytol., 226(4): 978-983.

Crowl AA, Manos PS, McVay JD, Lemmon AR, Lemmon EM, Hipp AL. 2020. Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*). New Phytol. 226(4): 1158-1170.

Chen J, Tan HTW, Wong KM. 2015. All that shimmers is not the same: silvered leaflets and micromorphological diversity in *Timonius* DC. (Rubiaceae). Plant Syst. Evol. 301:841–848.

Deng M, Li Q, Yang S, Liu Y, Xu J. 2013. Comparative morphology of leaf epidermis in the genus *Lithocarpus* and its implication in leaf epidermal feature evolution in Fagaceae. Plant Syst. Evol. 299:659–681.

Deng M, Hipp A, Song YG, Li QS, Coombes A, Cotton A. 2014. Leaf epidermal features of *Quercus* subgenus *Cyclobalanopsis* (Fagaceae) and their systematic significance. Bot. J. Linn. Soc.176: 224-259.

Deng M, Song Y, Li QS, Coombes A, Li QJ. 2015. Clarification of identity of *Quercus patkoiensis* and *Q. semiserratoides* (Fagaceae) using leaf epidermal features. Plant Syst. Evol. 301: 849-860.

Deng M, Xiang LX, Song GY, Coombes A, Yang RX, Xiong SY, and Li SQ. 2017. Leaf epidermal features of *Quercus* group Ilex (Fagaceae) and their application to species identification. Rev. Palaeobot. Palynol. 237: 10-36.

Dyal SA. 1936. A key to species of oaks of eastern North America based on foliage and twig characters. Rhodora. 38:53-63

Ecevit-Genç G, Özcan T, Dirmenci T. 2017. Nutlet and leaf micromorphology in some Turkish species of *Teucrium* L. (Lamiaceae). Phytotaxa. 312:71–82.

Fortini P, Antonecchia G, Di Marzio P, Maiuro L, Viscosi V. 2015. Role of micromorphological leaf traits and molecular data in taxonomy of three sympatric

white oak species and their hybrids (*Quercus* L.). Plant Biosyst. 149: 546-558.

Fortini P, Viscosi V, Maiuro L, Fineschi S, Vendramin GG. 2009. Comparative leaf surface morphology and molecular data of five oaks of the subgenus *Quercus* Oerst (Fagaceae). Plant Biosyst. 143(3): 543-554.

Freudenstein JV, Broe MB, Folk RA, Sinn BT. 2017. Biodiversity and the species concept—lineages are not enough. Syst. Biol. 66(4): 644-656.

Futuyma DJ. 1998. Evolutionary Biology. Sunderland (MA): Sinauer. Gabriel K.R., Sokal R.R. 1969. A new statistical approach to geographic variation analysis. Syst. Zool. 18:259–278.

Gharemaninejada F, Khalili Z, Maassoumi AA, Mirzaie-Nodoushan H, Riahi M. 2012. Leaf epidermal features of *Salix* species (Salicaceae) and their systematic significance. Am. J. Bot. 99: 769-777.

Gissi DS, Seixas DP, Fortuna-Perez AP, Torke BM, Simon MF, Souza G, ... Rodrigues TM. 2022. Leaf and stem anatomy of the *Stylosanthes guianensis* complex (Aubl.) Sw. (Leguminosae, Papilionoideae, Dalbergieae) and its systematic significance. Flora. 287: 151992.

González-Villarreal LM. 1986. Contribución al Conocimiento del Género *Quercus* (Fagaceae) en el Estado de Jalisco. Instituto de Botánica, Colección Flora de Jalisco, Universidad de Guadalajara, Guadalajara.

Hardin JW. 1976.Terminology and classification of *Quercus* trichomes. J. Elisha Mitchell Sci. Soc. 92: 151-161.

Hardin JW. 1979. Patterns of variation in foliar trichomes of eastern North American *Quercus*. Am. J. Bot.66: 576-585.

Hernandez JO, Park BB. 2022. The Leaf Trichome, Venation, and Mesophyll Structural

Traits Play Important Roles in the Physiological Responses of Oak Seedlings to Water-Deficit Stress. Int. J. Mol. Sci. 23(15): 8640.

Hipp AL, Manos PS, González-Rodríguez A, Hahn M, Kaproth M, McVay JD, Avalos SV, Cavender-Bares J. 2018. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. New Phytol. 217: 439–452.

Hipp AL, Manos PS, Hahn M, Avishai M, Bodénes C, Cavender-Bares J, Crowl A, Deng M, Denk T, Fitz-Gibbon S, Gailing O, González-Elizondo MS, González-Rodríguez A, Grimm GW, Jiang XL, Kremer A, Lesur I, McVay JD, Plomion C, Rodríguez-Correa H, Schulze ED, Simeone MC, Sork VL, Valencia-Avalos S. 2020. Genomic landscape of the global oak phylogeny. New Phytol. 1-15.

Jones JH. 1986. Evolution of the Fagaceae: the implications of foliar features. Ann. Missouri Bot. Gard. 73: 228–275.

Kim WK, Cho HD, Kim PG. 2011. Morphology of Foliar Trichomes of the Chinese Cork Oak *Quercus variabilis* by Electron Microscopy and Three-Dimensional Surface Profiling. Microsc. Microanal. 17: 461-468.

Kim BY, Wei X, Fitz-Gibbon S, Lohmueller KE, Ortego J, Gugger PF, Sork VL. 2018. RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: Fagaceae). Mol. Ecol. 27(22): 4556-4571.

Liebmann FM. 1984. Oversigt over det kongelige danske videnskabernes selskabs forhandlinger og dets medlemmers arbeider. 179

López-Caamal A, Ruiz-Amaro LDC, Zepeda-Rodríguez A, Mussali-Galante P, Tovar-Sánchez E. 2017. Micromorphological character expression of the hybrid *Quercus* ×

dysophylla and its parental species (*Q. crassifolia* and *Q. crassipes*). Bot. Sci. 95(3): 375-389.

Mannethody S, Purayidathkandy S. 2018. Trichome micromorphology and its systematic significance in Asian *Leucas* (Lamiaceae). Flora. 242: 70-78.

Manos PS. 1993. Foliar trichome variation in *Quercus* section *Protobalanus* (Fagaceae). Sida. 15: 391-403.

Manos SP, Doyle JJ, Nixon KC. 1999. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). Mol. Phylogenet. Evol. 12: 333–349.

Maya-García R, Torres-Miranda A, Cuevas-Reyes P, Oyama K. 2020. Morphological differentiation among populations of *Quercus elliptica* Née (Fagaceae) along an environmental gradient in Mexico and Central America. Bot. Sci. 98(1): 50-66.

McVaugh R. 1974. Fagaceae. In: McVaugh R. ed. Flora Novo-galiciana. Controbutions from the University of Michigan herbarium. 12: 50-30.

McVay JD, Hauser D, Hipp AL, Manos PS. 2017. Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in western North America. Genome. 60(9): 733-742.

Morales-Saldaña S, Valencia-Ávalos S, Oyama K, Tovar-Sánchez E, Hipp AL, González-Rodríguez A. 2022. Even more oak species in Mexico? Genetic structure and morphological differentiation support the presence of at least two specific entities within *Quercus laeta*. J Syst Evol. 60: 1124-1139.

Panahi P, Jamzad Z, Pourmajidian MR, Fallah A, Pourhashemi M. 2012. Foliar epidermis morphology in *Quercus* (subgenus *Quercus*, section *Quercus*) in Iran. Acta Bot. Croat. 71(1): 95-113.

Payne WW. 1978. A glossary of plant hair terminology. Brittonia. 30: 239–255

Pérez ME, Valencia-Á S. 2017. Estudio preliminar del género Quercus (Fagaceae) en Tamaulipas, México. Acta Bot. Mex. (120) : 59-111.

Petit RJ, Bodénès C, Ducousso A, Roussel G, Kremer A. 2004. Hybridization as a mechanism of invasion in oaks. New Phytol. 161(1): 151-164.

R Core Team 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rasband WS. 1997-2018. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, https://imagej.nih.gov/ij/,

Romero SR, Carlos E, Zenteno R, de Lourdes M, Enríquez A. 2002. El género *Quercus* (Fagaceae) en el estado de México. Ann. Missouri Bot. Gard. 82: 551–593.

Romero SR, Rojas ZCE, and Rubio LEL. 2014. Flora del Bajío y regiones adyacentes. Familia Fagaceae. Fascículo, 181.México: Instituto de Ecología, A.C.

Sánchez-Acevedo V, Torres-Miranda A, Gutiérrez-Arellano C, Hernández-Esquivel KB, Oyama K. 2022. Variability in leaf morphological traits of an endemic Mexican oak (*Quercus mexicana 131exican131* Bonpl.) along an environmental gradient. Bot. Sci. 100(3): 579-599.

Scareli-Santos C, Herrera-Arroyo ML, Sánchez-Mondragon ML, González-Rodríguez A, Bacon J, Oyama K. 2007. Comparative analysis of micromorphological characters in two distantly related Mexican oaks, *Quercus conzattii* and *Q. eduardii* (Fagaceae), and their hybrids. Brittonia. 59: 37-48.

Scareli-Santos C, Sánchez-Mondragón ML, González-Rodríguez A, Oyama K. 2013. Foliar micromorphology of mexican oaks (*Quercus*: Fagaceae). Acta Bot. Mex. (104): 31-52.

Siadati S, Salmaki Y, Bräuchler C. 2020. Trichome morphology provides phylogenetically informative signal for generic delimitation in tribe Marrubieae (Lamiaceae). Flora. 273: 151720.

Spellenberg R, Bacon JR. 1996. Taxonomy and distribution of a natural group of black oaks of Mexico (*Quercus*, section Lobatae, subsection Racemiflorae). Syst. Bot. 85-99.

Stace CA. 1984. The taxonomic importance of the leaf surface. In: Heywood V H. Moore D H, eds. Current concepts in plant taxonomy. London: Academic Press.

Stephan JM, Teeny PW, Vessella F, Schirone B. 2018. Oak morphological traits: between taxa and environmental variability. Flora. 243: 32-44.

Struck TH, Feder JL, Bendiksby M, Birkeland S, Cerca J, Gusarov VI., ... & Dimitrov D. 2018. Finding evolutionary processes hidden in cryptic species. Trends Ecol. Evol. 33(3): 153-163.

Tschan FG, Denk T. 2012. Trichome types, foliar indumentum and epicuticular wax in the Mediterranean gall oaks, *Quercus* subsection Galliferae (Fagaceae): implications for taxonomy, ecology, and evolution. Bot. J. Linn. Soc. 169: 611-644.

Tucker JM. 1974. Patterns of parallel evolution of leaf form in new world oaks. Taxon. 23(1): 129-154.

Ullah F, Ayaz A, Saqib S, Parmar G, Bahadur S, Zaman W. 2021. Taxonomic implication of leaf epidermal anatomy of selected taxa of Scrophulariaceae from Pakistan. Microsc. Res. Tech. 84(3): 521-530.

Valencia-Á. S. 2004. Diversidad del género *Quercus* (Fagaceae) en México. Bot. Sci. 75: 33-53.

Valencia-A, S. 2021. Species delimitation in the genus *Quercus* (Fagaceae). Bot. Sci., 99: 1-12.

Valencia-Ávalos S., Delgado-Salinas A. 2003. Los tricomas foliares en la caracterización de un grupo de especies del género *Quercus*, sección *Lobatae* (Fagaceae). Anales del Instituto de Biología. Serie Botánica, 74(1), 5-15.

Vázquez ML. 2006. Trichome morphology in selected mexican red oaks species (*Quercus* section *Lobatae*). Sida. 22: 1091-1110

Woodenberg W, Govender J, Murugan N, Ramdhani SS. 2019. Cycad forensics: leaflet micromorphology as a taxonomic tool for South African cycads. Plant Syst. Evol. 305: 445-457.

Zafar M, Ahmad M, Shaheen S, Sultana S, Rehman SU, Amina H. 2019. Micromorphological investigation of leaf epidermis and seeds of Vitaceae from Pakistan using light microscopy and scanning electron microscopy. Microsc. Res. Tech. 82(4): 335-344.

Zapata F, Jiménez I. 2012. Species delimitation: inferring gaps in morphology across geography. Syst. Biol., 61(2): 179.

.

Figures

Figure 1. Localities sampled for *Quercus laeta* complex.

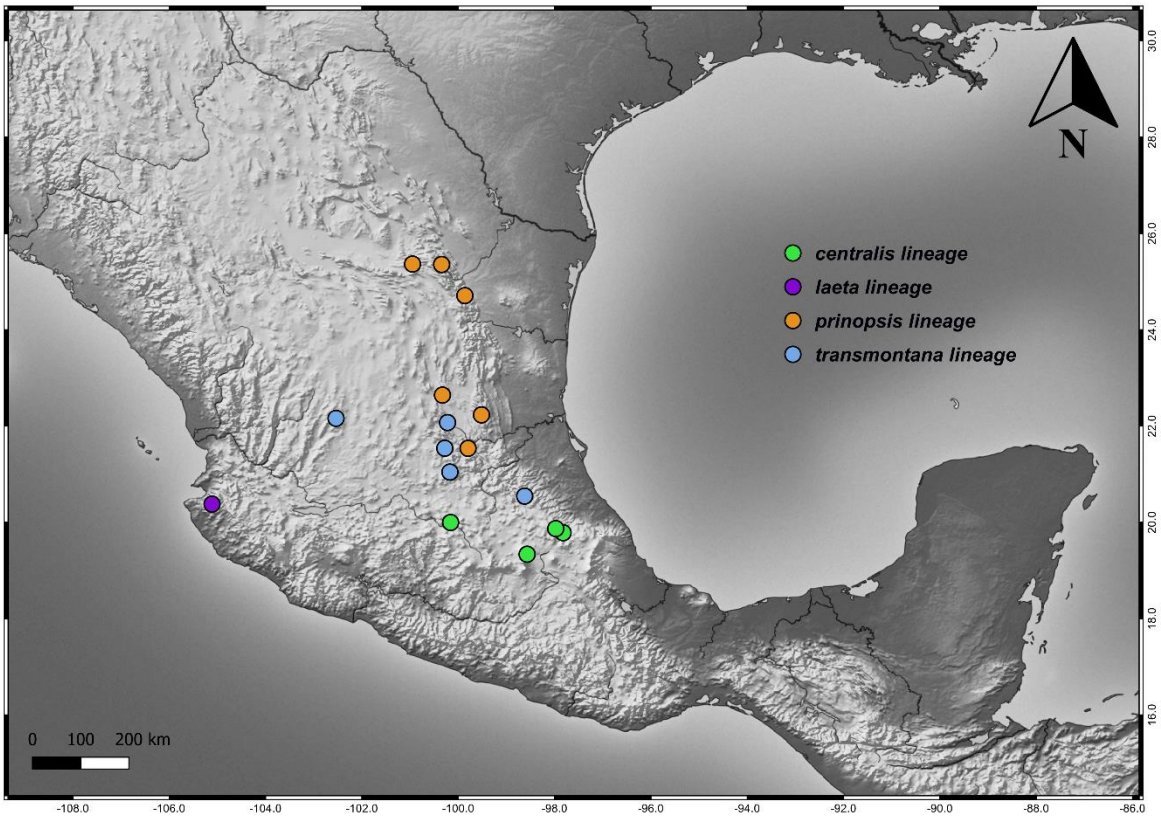


Figure 2. SEM micrographs for the trichomes types. Figs. A-D. Simple uniseriate trichomes; (A) *Q. centralis*, (B) *Q. centralis*, (C) *Q. transmontana*, (D) *Q. transmontana*. Figs. D-H. Fasciculate trichomes; (D) *Q. transmontana* (E) *Q. centralis* (F) *Q. centralis* (G) *Q. transmontana* (H) *Q. prinopsis*. Figs I-L Stellate trichomes; (I) *Q. laeta*, (J) *Q. centralis*, (K) *Q. centralis*, (L).

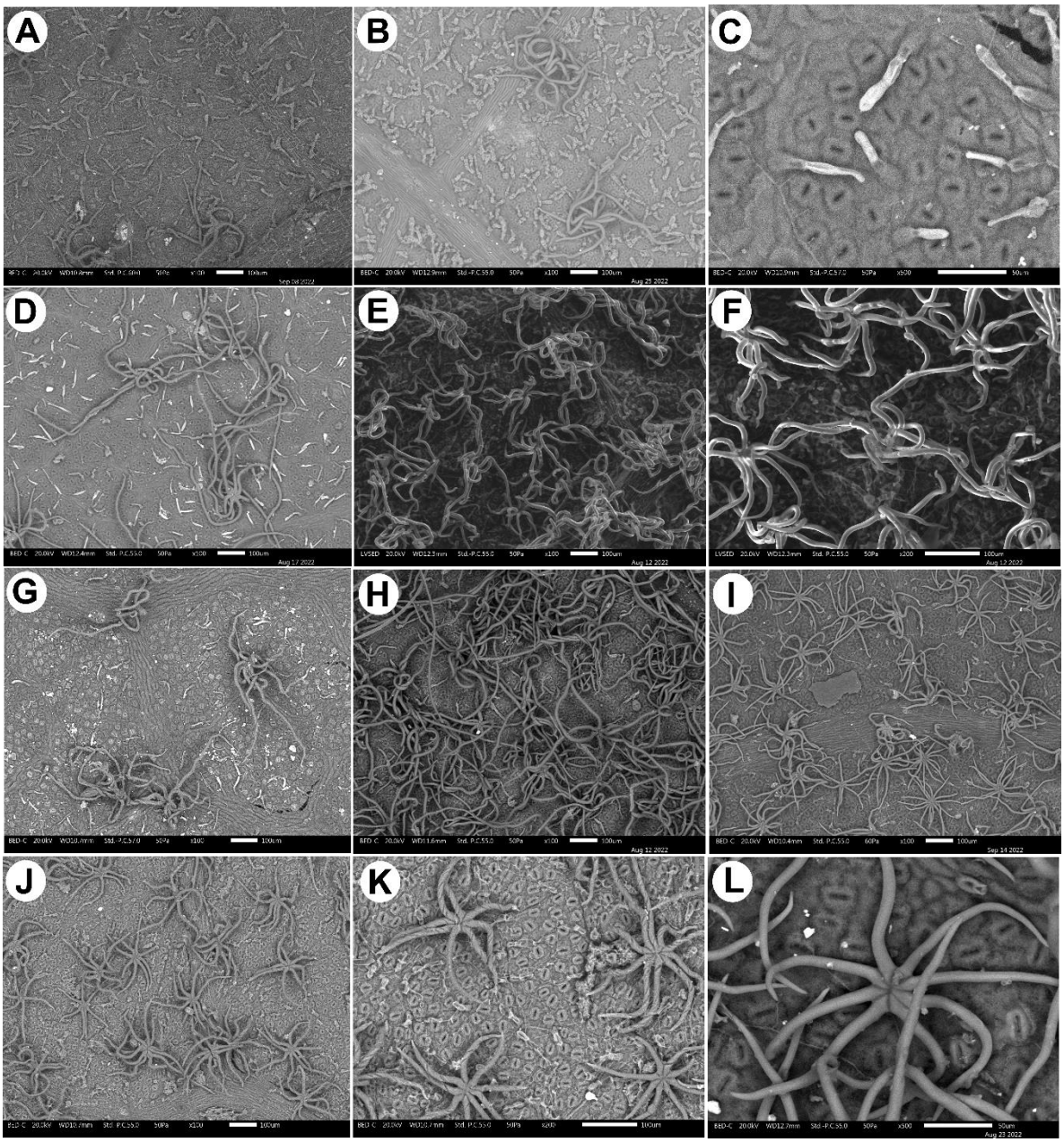


Figure 3. Different types of waxes present in *Quercus laeta* complex. Figs A-B. *Q. transmontana* and *Q. centralis*: Wax crusts. Figs C-D. *Q. transmontana* and *Q. laeta*: Crystalloids. Figs E-F. *Q. transmontana* and *Q. prinopsis*: Wax layers.

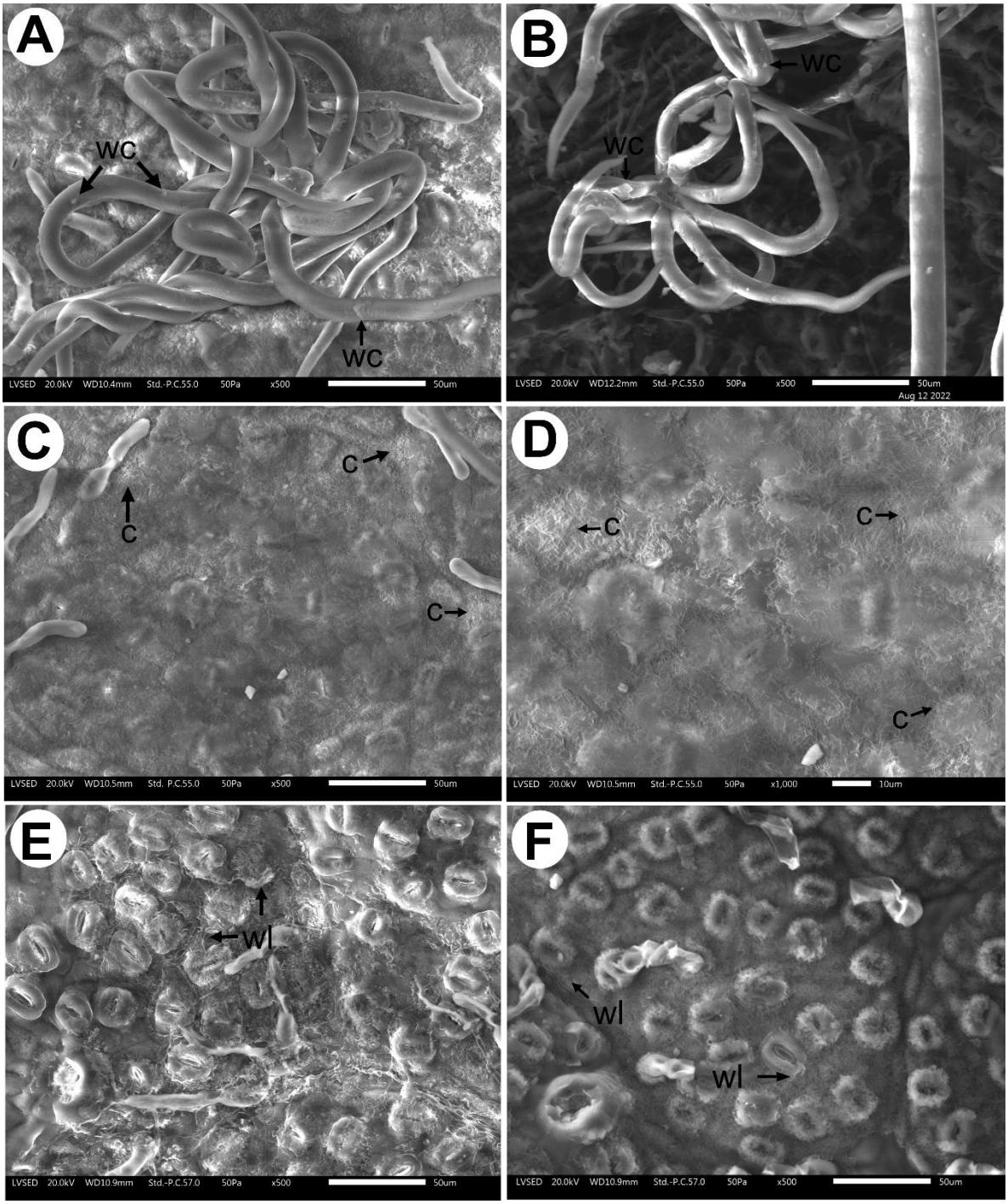


Figure 4. Length variation of trichome´s arms for each species and trichome type. A. Length variation of trichome´s arms between fasciculate and stellate trichomes. B Length variation of stellate trichomes arms. C. Length variation of fasciculate trichomes arms. Wider sections of the violin plot represent a higher probability of observations taking a given value, the thinner sections correspond to a lower probability

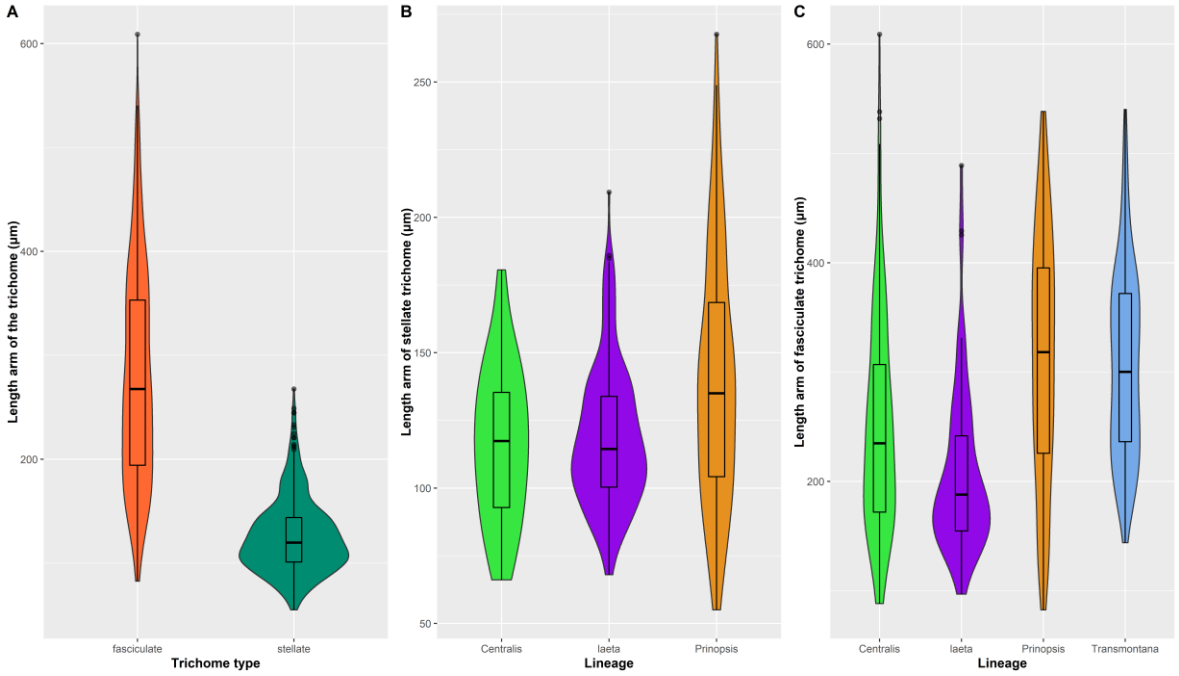


Figure 5 Arm number variation for each lineage and trichome type. A. Arm number variation between fasciculate and stellate trichomes. B Arms number for stellate trichomes by species C. Arms number for fasciculate trichomes by species. Wider sections of the violin plot represent a higher probability of observations taking a given value, the thinner sections correspond to a lower probability

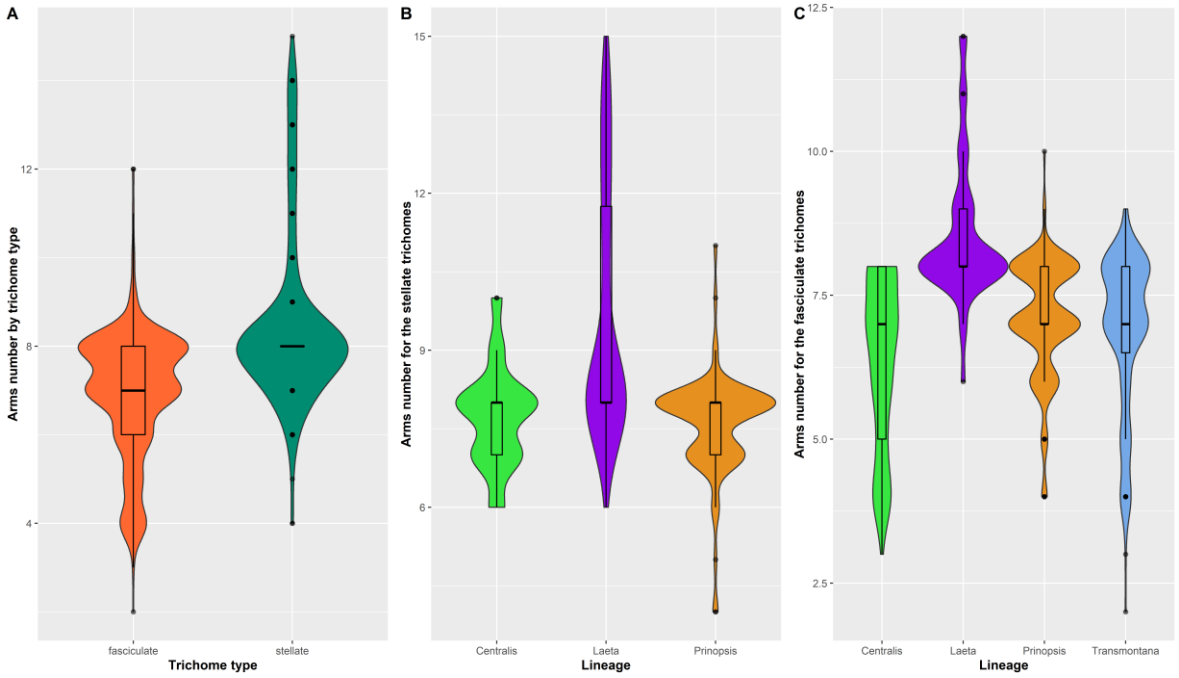


Figure 6 Trichome density variation among species. A. Number of non-glandular trichomes by unit area sampled. B. Number of glandular trichomes by unit area sampled. Wider sections of the violin plot represent a higher probability of observations taking a given value, the thinner sections correspond to a lower probability

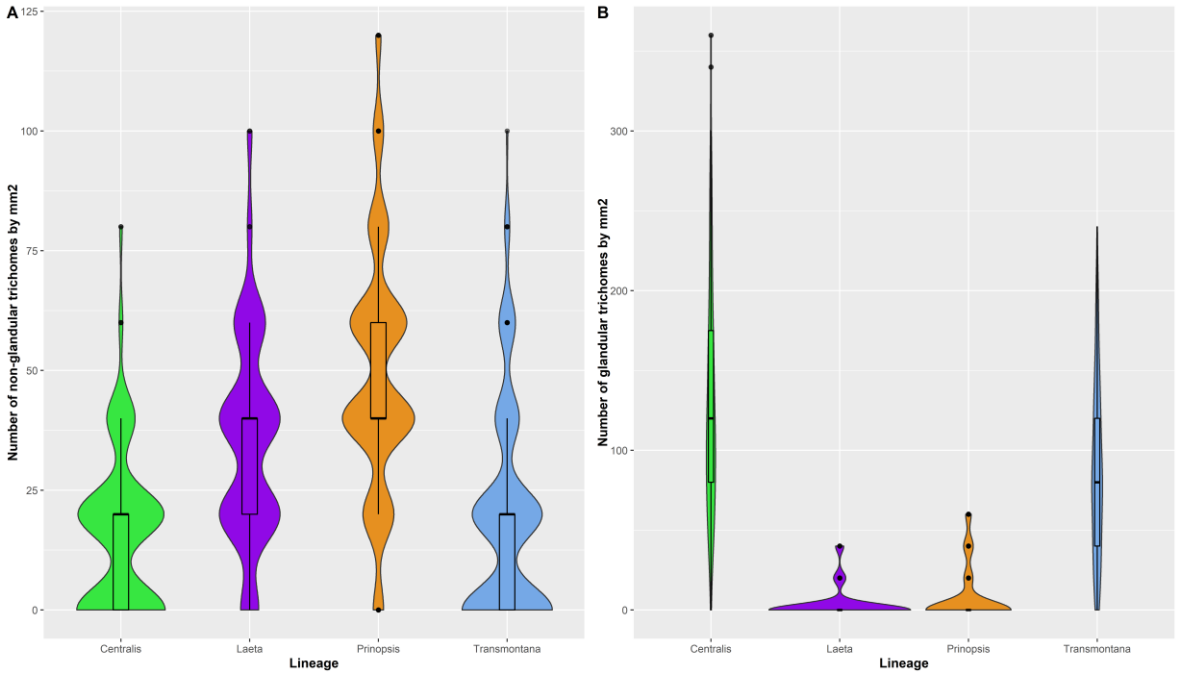


Figure 7 Principal Component Analysis (PCA) among the studied species. LTA: length variation of trichome arms; TGD: trichome glandular density; TNGD: trichome non-glandular density; and NTA: variation in the number of trichome arms

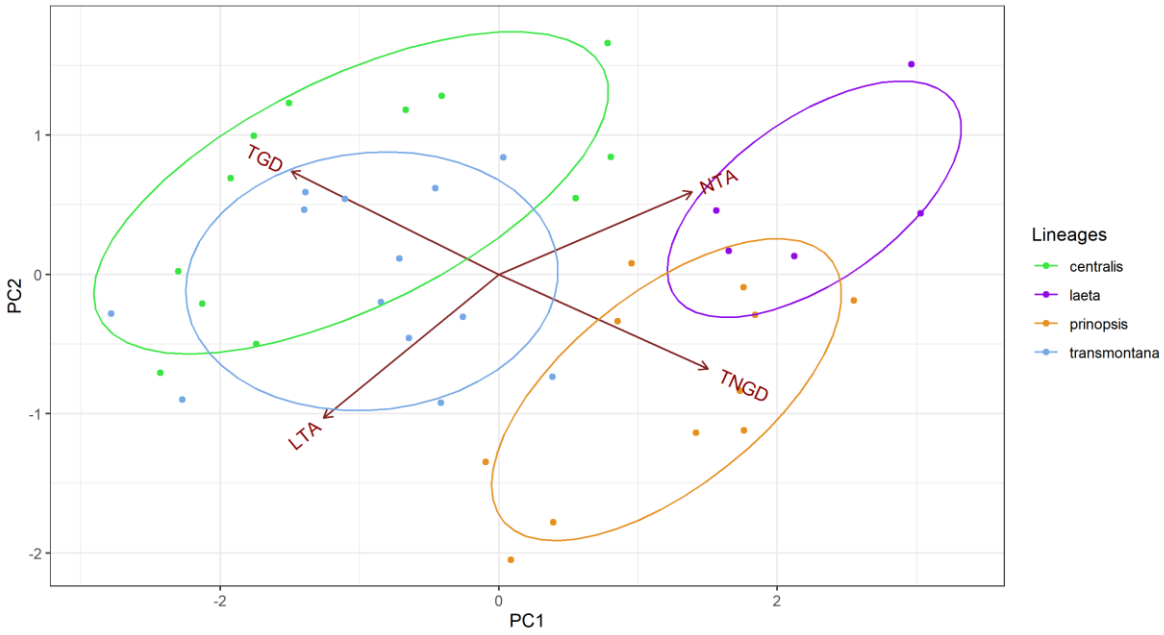
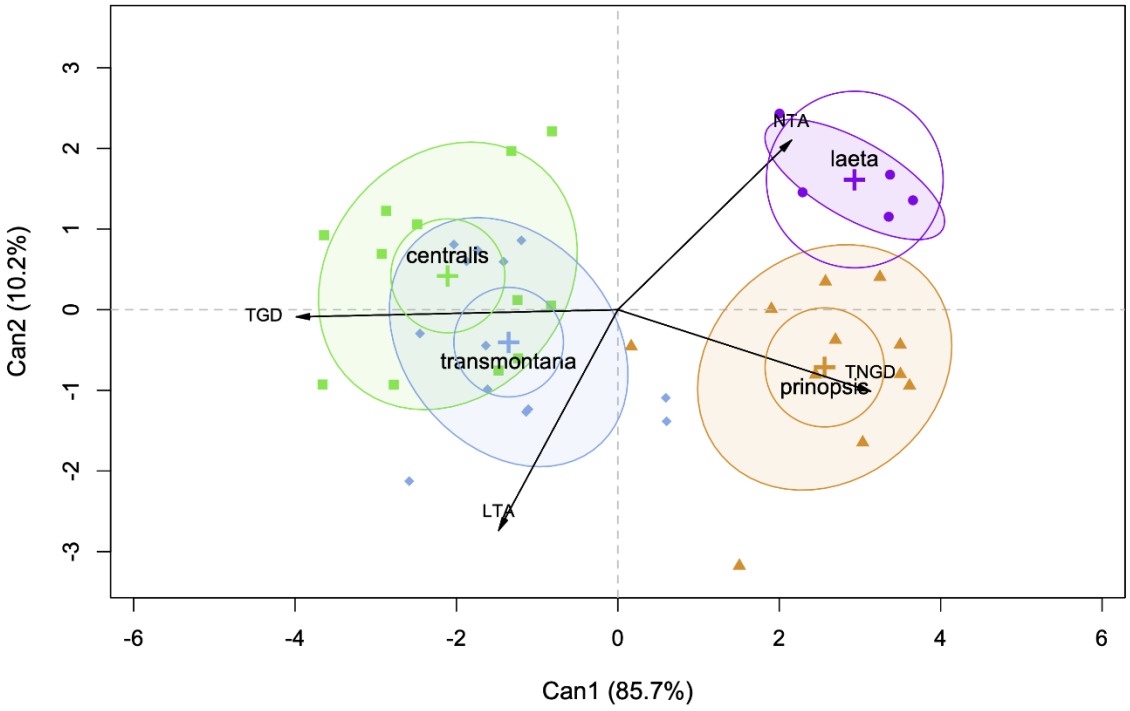


Figure 8 Scatterplot of scores derived from discriminant functions Axis 1 vs. Axis 2 produced by discriminant analysis applied to micromorphological characters for four species of the *Quercus laeta* complex. LTA: length variation of trichome arms; TGD: trichome glandular density; TNGD: trichome non-glandular density; and NTA: variation in the number of trichome arms



Tables.

Table 1. Details of surface abaxial and trichome morphological characteristics in the four species of the *Quercus laeta* complex

Species	Density	Surface	Wax types			Glandular trichomes	Non-glandular trichomes	
			Crystalloids	Wax crusts	Wax layer	Simple uniseriate	Fasciculate	Stellate
Q. centralis	Dense	Papillate	-	+	+	+	+	+
Q. laeta	Dense	Papillate	+	+	+	+	+	+
Q. prinopsis	Very dense	Papillate	-	+	+	+	+	+
Q. transmontana	Sparse	Papillate	+	+	+	+	+	-

Table 2. Quantitative summary leaf epidermal features on abaxial surface of the species within *Quercus laeta* complex.

Species	Non-glandular trichome density (mm^2)	Glandular trichome density (mm^2)	Arm trichome fasciculate length (µm)	Arm trichome stellate length (µm)	Arm number for the stellate trichomes	Arm number for fasciculate trichomes
Q. centralis	0 – 20 (80)	(0) 80 – 175 (360)	(88) 171.94 – 306.92 (608.93) (250.67 ± 99.77)	(66) 92.79 – 135.33 (180) (116.36 ± 28.43)	6 – 10 (7.73 ± 0.96)	3 –8 (6.36 ± 1.45)
Q. laeta	(0) 20 – 40 (100)	0 – 40 (3.5)	(96) 154.78 –241.80 (489) 205.48 ± 74.22	(68) 100.36 – 133.86 (209) (116.36 ± 28.43)	6 – 16 (9.53 ± 2.31)	6 – 12 (8.53 ± 1.24)
Q. prinopsis	(0) 40 – 60 (120)	0 – 60	(82) 225.84 – 395.23 (538) (313.20 ± 111.96)	(55) 104.19 – 168.54 (267) (139.35 ± 45.47)	4 – 11 (7.58 ± 0.98)	4 – 10 (7.11 ± 1.02)
Q. transmontana	0 – 20 (100)	40 – 120 (240)	(144) 236.4 – 371.9 (540.2) (307.22 ± 85.97)	–	2 – 9 (6.86 ± 1.38)	–

Table 3. Pairwise Tukey test among species for arm length of trichomes. p-values for the pairwise comparison among species for stellate trichomes (above), p-values for the pairwise comparison among species for fasciculate trichomes (below). Significance. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'

	Q. centralis	***Q. laeta***	***Q. prinopsis***	***Q. transmontana***
Q. centralis	----------	0.93	0.0001***	–
Q. laeta	0.004*	----------	0.0000**	–
Q. prinopsis	0.0000***	0.0000***	----------	–
Q. transmontana	0.0000***	0.0000***	0.94	----------

Table 4. Pairwise Tukey test among species for the number of arms of the trichomes. p-values for the pairwise comparison among species for stellate trichomes (above), p-values for the pairwise comparison among species for fasciculate trichomes (below). Significance. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'

	Q. centralis	***Q. laeta***	***Q. prinopsis***	***Q. transmontana***
Q. centralis	----------	0.0000***	0.90	–
Q. laeta	0.0000***	----------	0.0000**	–
Q. prinopsis	0.0000***	0.0000***	----------	–
Q. transmontana	0.005*	0.0000***	0.37	----------

Table 5. Pairwise Tukey test among species for the density trichomes. p-values for the pairwise comparison among species for non-glandular trichomes (above), p-values for the pairwise comparison among species for glandular trichomes (below). Significance. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'

	Q. centralis	***Q. laeta***	***Q. prinopsis***	***Q. transmontana***
Q. centralis	----------	0.0000***	0.0000***	0.5470
Q. laeta	0.0000***	----------	0.0000**	0.0000***
Q. prinopsis	0.0000***	0.9451	----------	0.0000***
Q. transmontana	0.0000***	0.0000***	0.0000 ***	----------

Table 6. CDA. Standardized coefficients for canonical variables derived from discriminant function analysis of the *Quercus laeta* complex. Numbers in bold font indicate the higher values.

Variables	**Axis 1**	**Axis 2**
number of trichome arms (NTA)	-0.13533	**0.46711**
trichome non-glandular density (TNGD)	**-0.16944**	-0.80755
trichome glandular density (TGD)	**0.92138**	-0.10573
length variation of trichome arms (LTA)	-0.09507	**-0.77656**

Supplementary table 1 Confusion matrix for summary of classification of the four species of the *Quercus laeta* complex traits

Species	***Q. centralis***	***Q. laeta***	***Q. prinopsis***	***Q. transmontana***
Q. centralis	0.833	0	0	0.167
Q. laeta	0	1	0	0
Q. prinopsis	0	0	0.909	0.091
Q. transmontana	0.231	0	0.077	0.692

V. **Discusión general**

La especiación es uno de los procesos responsables del origen de la diversidad biológica (Gavrilets, 2003; Nosil et al., 2021), y típicamente comienza con una barrera al flujo génico, la cual en consecuencia podría promover una mayor divergencia genética y fenotípica (Nosil y Feder, 2012). Este modelo sugiere una división de linajes a nivel poblacional, donde divergen y permanecen independientes. Sin embargo, esta visión puede en muchas ocasiones simplificar la historia evolutiva de los taxa. Ahora se conoce que la hibridación es, de hecho, frecuente y generalizada y ha jugado un papel importante en la diversificación a través de diversos linajes del árbol de la vida (Abbott et al., 2013; Mallet et al., 2016; Solís-Lemus y Ané, 2016; Gernandt et al., 2018; Crowl et al., 2020; Kleinkopf et al., 2019; Taylor y Larson, 2019; Rancilhac et al., 2021). En este contexto y bajo la visión génica de la especiación, la divergencia de especies ocurre a través de un continuo de diferenciación genética (Wu, 2001), con especies incipientes que pasan por una fase en la que solo están parcialmente aisladas reproductivamente (Kopp y Frank, 2005; Stankowski y Ravinet, 2021). Esta idea de un continuo de especiación ha permitido que estos puntos de vista sobre la especiación aparentemente paradójicos coexistan como parte del mismo marco conceptual (Stankowski y Ravinet, 2021). Además, la reconstrucción conceptual del concepto de especie, mediante una propuesta unificadora que identifica un elemento en común entre múltiples conceptos de especie, al equiparar a la especie con segmentos de linajes de meta-poblaciones que evolucionan de manera conjunta (De Queiroz, 2007), ha permitido eliminar los conflictos entre diferentes conceptos de especie sin negar la importancia de las propiedades secundarias que subyacen a dichos conceptos.

Estos avances teóricos-prácticos han dirigido este trabajo, permitiendo implementar estrategias donde convergen disciplinas como la genética de poblaciones, la morfología, ecología y la filogenética para abordar problemas taxonómicos bajo un marco evolutivo en grupos donde ha sido persistente la incertidumbre con respecto a la divergencia y diferenciación de taxones. En este sentido, el género *Quercus* resultó ser un sistema idóneo para investigar patrones de especiación desde múltiples perspectivas. De modo que, este estudio representa uno de los primeros intentos de integrar evidencia a nivel ecológico, genómico y morfológico para solucionar primero a un nivel taxonómico, la existencia de múltiples especies dentro de *Quercus laeta* (aquí llamado el complejo *Quercus laeta*) y segundo, hacer inferencias bajo un marco evolutivo sobre los procesos que han llevado a la diversificación del complejo, arrojando luz sobre la historia evolutiva y la diversificación dentro del clado *Leucomexicana.*

Evidencia de una divergencia a múltiples niveles dentro del complejo *Quercus laeta.*

De Queiroz (2007) propone un cambio fundamental en la forma en que se conceptualizan las especies. Por un lado, mantiene el elemento común a todos los conceptos contemporáneos de especie, por otro lado, elimina los conflictos entre conceptos antagonistas sin negar la importancia de las propiedades que subyacen a dichos conceptos. De modo que estas propiedades subyacentes son consideradas propiedades contingentes, características que las especies pueden llegar a tener o no durante el curso de su existencia. En este contexto, los linajes identificados dentro del complejo *Q. laeta* muestran diferentes grados de diferenciación con base en diversas propiedades contingentes; desde muy divergentes, con diferencias morfológicas conspicuas y recíprocamente monofiléticas, hasta especies con bajos niveles de diferenciación genética y morfológica, aunque ecológicamente divergentes

(Tabla 1). Además, nuestros resultados sugieren que los diferentes grados de diferenciación que ocurren dentro del complejo se deben a que los linajes de *Q. laeta* representan un grupo polifilético, con diferentes trayectorias evolutivas, donde procesos como la introgresión histórica, el sorteo incompleto de linajes (ILS) y la divergencia de nicho han dado forma a la diversificación.

Recientemente, Valencia-Á. (2020) consideró relevante evaluar la divergencia en las especies de encinos desde diferentes perspectivas, por ejemplo, señaló la importancia de valorar, además del concepto taxonómico, el concepto ecológico y genético para la delimitación de especies. Bajo esta premisa, uno de nuestros resultados más notables fue que utilizando pocos loci de microsatélites nucleares y analizándolos con diferentes métodos basados en designaciones tanto *a priori* como *de novo,* fue posible identificar un patrón de diferenciación genética jerárquica dentro del complejo. Si bien hubo una diferencia en el número de grupos genéticos detectados mediante los diferentes métodos, fue posible identificar una importante y constante discontinuidad genética entre dos entidades. Nuestros resultados genéticos sugieren que los individuos etiquetados inicialmente como morfotipo *transmontana* y *transmontana coyula* podrían considerarse como una misma entidad genética (denominada *Q. transmontana;* Capítulo I; Figs. 2, 3, 4). Mientras que los individuos categorizados inicialmente en los morfotipos *centralis*, *laeta* y *prinopsis* formaron un segundo grupo genético (*Q. laeta;* Capítulo I; Figs. 2, 3, 4), aunque con un importante patrón de subestructura genética discutido más a fondo en el capítulo I. Paralelamente, este patrón de diferenciación genética estuvo acompañado de una diferenciación a nivel morfométrico, en donde fue posible identificar una divergencia entre las hojas lanceoladas y oblongas de *Q. transmontana* y las hojas obovadas y elípticas de *Q. laeta*.

En el capítulo II, combinando enfoques filogenéticos, basados en secuenciación de enriquecimiento dirigido, con la evaluación de la diferenciación de nichos ecológicos permitió identificar hasta seis linajes dentro del complejo, contrastando con las dos entidades genético-morfométricas identificadas en el capítulo I. Si bien *Q. transmontana,* identificado mediante datos genéticos y morfométricos, fue soportado como un linaje independiente utilizando datos multilocus y ecológicamente divergente mediante análisis de divergencia de nicho, *Q. laeta* (también identificado utilizando datos genéticos y morfométricos) fue dividido hasta en cinco linajes mediante el uso de datos multilocus. Además, los resultados de análisis filogenéticos sugieren que los linajes identificados dentro del complejo *Q. laeta* son polifiléticos estando más cercanamente relacionados con otras especies del clado *Leucomexicana* que entre sí mismos. Considerando estos resultados, es pertinente mencionar que la asignación de poblaciones basada en la agrupación genética (realizada en el capítulo I), si bien podría establecer límites entre grupos con un importante aislamiento genético, debería considerarse un enfoque a seguir en las primeras etapas de la delimitación de especies, y no como un método de delimitación por sí mismo, ya que si bien estos enfoques agrupan individuos que probablemente intercambien genes o tienen frecuencias alélicas similares debido a la ascendencia compartida (Pritchard et al., 2000; Rittmeyer y Austin, 2012; Carstens et al., 2013), estudios recientes han demostrado que el nivel de divergencia genética intraespecífica no predice la tasa de especiación, lo que indica que las poblaciones estructuradas a menudo no logran convertirse en especies (Huang y Knowles, 2016; Singhal et al., 2018). De tal manera, resulta necesario contrastar cualquier hipótesis genética de especie utilizando múltiples e independientes líneas de evidencia para robustecer y validar su estatus.

Una de las principales dificultades, no solo dentro del complejo, sino de los encinos en general, es la identificación de caracteres diagnósticos morfológicos estables. Particularmente para *Q. laeta*, la interpretación morfológica ha sido ambivalente. Por un lado, representa una de las especies de roble con mayor polimorfismo poblacional (Valencia-Á, 2004; Morales-Saldaña et al., 2022), cuya variación se ha observado principalmente en el grado de pubescencia sobre el envés de las hojas, la forma de la lámina foliar y la pubescencia de las ramillas. Por otro lado, el hecho de que estas poblaciones compartan ciertos caracteres micromorfológicos considerados como caracteres diagnósticos, como la presencia de papilas, tricomas sésiles contortos y la ausencia de tricomas glandulares en el envés de la hoja, ha llevado a diversos taxónomos a considerar que las poblaciones polimórficas presentes en *Q. laeta* en realidad representan una sola especie (McVaugh, 1974; Rangel et al., 2002), y que la gran variación morfológica podría ser consecuencia de la plasticidad fenotípica y eventos de hibridación local con diversas especies de encinos. Sin embargo, ningún estudio había evaluado anteriormente la variación morfológica poblacional bajo una hipótesis filogenómica y por consiguiente presentado discontinuidades morfológicas dentro del complejo.

Los análisis morfométricos revelaron que la forma de la lámina foliar es importante para identificar a *Q. transmontana* del resto de especies; sin embargo, el traslape morfométrico foliar es prevalente dentro del complejo. De tal manera que el mayor grado de diferenciación morfológica ocurre a un nivel micro morfológico y únicamente bajo un enfoque cuantitativo-comparativo fue posible detectar ciertas combinaciones de caracteres (p. ej. la presencia/ausencia de tricomas estrellados, longitud de los radios y la densidad de los tricomas), que permiten distinguir desde una perspectiva morfológica a estos linajes

detectados mediante datos genómicos. De este modo, los resultados obtenidos en este trabajo resaltan que i) los caracteres considerados históricamente como diagnósticos no serían los adecuados para discriminar entre linajes, y que la presencia de estos caracteres en diversos linajes no relacionados podría ser consecuencia de una convergencia morfológica, y ii) se proponen una serie de caracteres micromorfológicos que pueden ser evaluados y utilizados como caracteres diagnósticos para identificar y soportar desde una perspectiva morfológica a los linajes detectados mediante datos multilocus. Además, considerando que el enfoque cualitativo claramente no puede distinguir entre los linajes presentes en el complejo, la propuesta del uso de un enfoque cuantitativo para evaluar discontinuidades morfológicas podría ser un enfoque que utilizar para resolver problemas de discriminación en otros complejos de encinos taxonómicamente controversiales cuando no es posible el uso de datos multilocus.

Finalmente, este estudio mostró que los linajes identificados mediante datos genómicos y respaldados por datos micromorfológicos están restringidos a determinadas regiones geográficas y con diferentes afinidades climáticas, sugiriendo que estos linajes son ecológicamente divergentes entre sí. Particularmente para México, se ha propuesto que las altas tasas de diversificación de linajes están asociadas con altas tasas de evolución a lo largo de gradientes de humedad (Hipp et al., 2018). Por lo tanto, nuestros resultados sugieren que las barreras montañosas pueden haber jugado un papel clave en la especiación y la diversificación a través de los efectos de la complejidad topográfica en la estratificación ecológica, la heterogeneidad ambiental y la limitación del flujo de genes (Rodríguez-Correa et al., 2015; Hipp et al., 2018; Barret et al., 2019).

Implicaciones taxonómicas para los encinos mexicanos

Bastante se ha hablado sobre la relevancia de la especie como unidad central en la biología, y que la asociación de nombres científicos inequívocamente a las especies es fundamental para una referencia confiable en el sistema de información biológica (Wheeler 2004; Padial y De la Riva, 2021). En este contexto, los nombres taxonómicos tienen consecuencias en el mundo real; pueden redefinir los programas de conservación, impactar el comercio internacional e informar sobre el manejo de plagas (Garnett y Christidis, 2017; Thomson et al., 2018). Por consiguiente, las revisiones taxonómicas, así como la identificación y discriminación de especies deben basarse en medidas sólidas y altamente replicables, así como en hipótesis comprobables debido a que son la base de la investigación en la biodiversidad, incluyendo estudios de evolución, conservación y biogeografía (Hebert et al., 2003; De Queiroz, 2007; Fujita et al., 2012; Luo et al., 2018).

A pesar del desarrollo de nuevos enfoques y marcos conceptuales en el campo de la delimitación de especies, históricamente, el proceso de identificación y delimitación de taxones dentro del género *Quercus* se ha basado en el concepto taxonómico de especie, por lo cual los rasgos macro y micro morfológicos han tenido mayor relevancia (Valencia-Á., 2020). Sin embargo, frecuentemente estos rasgos morfológicos son altamente homoplásicos, por lo cual en muchos casos el uso exclusivo de evidencia morfológica podría no ser suficiente para distinguir entre especies (Deng et al., 2017; Valencia-Á., 2020). En este contexto, el complejo *Q. laeta* ha sido un grupo donde la controversia con respecto a la divergencia y diferenciación de taxones ha sido constante, ya que ha sufrido múltiples ajustes en su clasificación intraespecífica (McVaugh, 1974; Rangel et al., 2002), debido principalmente a la presencia de caracteres morfológicos ambivalentes que han sido un desafío clasificar. No obstante, estudios filogenómicos previos habían sugerido que *Q. laeta*

no es un grupo monofilético (Hipp et al., 2018, Hipp et al., 2020), aunque el muestreo geográfico en estos estudios fue limitado. Considerando lo anterior, los resultados obtenidos a partir de ampliar el muestro a nivel poblacional a lo largo de la distribución geográfica de *Q. laeta*, y analizarlas bajo múltiples perspectivas, apoyan firmemente que las poblaciones del complejo *Q. laeta* son genética, morfológica y ecológicamente divergentes. De tal modo que es posible reconocer la existencia de al menos cuatro linajes principalmente alopátricos dentro del complejo, designados como *centralis*, *laeta*, *prinopsis* y *transmontana*, los cuales podrían considerarse como especies independientes bajo diferentes conceptos de especies. De tal manera, estos resultados sugieren que la diversidad de especies de robles en México sigue estando subestimada, a pesar de descripciones y reevaluaciones taxonómicas recientes (Valencia et al., 2016; González-Villarreal, 2018; McCauley et al., 2019; McCauley y Oyama, 2020; Valencia y Coombes, 2020; Morales-Saldaña et al., 2022) por lo que probablemente, escenarios similares ocurran al interior de otros complejos de encinos no estudiados.

Ahora bien, aunque la designación de nombres para las especies putativas aquí propuestas se realizó con base en los nombres propuestos por Trelease (1924), podrían existir nombres más antiguos que tengan prioridad nomenclatural, por lo que de manera complementaria a la caracterización micromorfológica presentada en este trabajo es necesaria una diagnosis morfológica detallada y, una investigación nomenclatural para los linajes aquí propuestos antes de que puedan ser formalmente redescritos.

Tabla 1. Caracteres sugeridos para distinguir entre especies identificadas dentro del complejo *Quercus laeta*.

Caracter	***Q. centralis***	***Q. laeta***	***Q. prinopsis***	***Q. transmontana***
Ramillas	Glabrescentes	Puberulentas	Tomentosas amarillo dorado	Pubescente
Forma de la hoja	Elíptica	Obovado	Obovada	Lanceolada, oblongo
Tipo de margen	Crenado	Crenado	Dentado	Entero, ocasionalmente crenado
Ancho de la hoja (mm)	20 – 57	10 – 30 (38)	19 – 51	9 – 30
Largo de la lámina (mm)	60 – 136	42 – 95 (133)	50 – 125	33- 89 (96)
Forma del ápice	Mucronado	Mucronado	Mucronado	Mucronado
Forma de la base	Redondeada, Oblicua	Redondeada	Oblicua, cordada	Redondeada
Forma de las yemas	Ovoides con escamas glabrescentes	Ovoides con escamas pubescentes	Cónicas con escamas pubescentes	Globosas/Ovoides con escama pubescentes
Forma de las estípulas	Acuminadas	Acuminadas pubescentes	Lineares	Lineares
Tricomas estrellados	Presente	Presente	Presente	Ausente
Densidad de tricomas glandulares (mm^2)	(0) 80 – 175 (360)	0 – 40 (3.5)	0 – 60	40 – 120 (240)
Densidad de tricomas no glandulares (mm^2)	0 – 20 (80)	(0) 20 – 40 (100)	(0) 40 – 60 (120)	0 – 20 (100)
Longitud de los brazos en tricomas fasciculados (µm)	(88) 171.94 – 306.92 (608.93) (250.67 ± 99.77)	(96) 154.78 –241.80 (489) 205.48 ± 74.22	(82) 225.84 – 395.23 (538) (313.20 ± 111.96)	(144) 236.4 – 371.9 (540.2) (307.22 ± 85.97)
Longitud de los brazos en tricomas estrellados (µm)	(66) 92.79 – 135.33 (180) (116.36 ± 28.43)	(68) 100.36 – 133.86 (209) (116.36 ± 28.43)	(55) 104.19 – 168.54 (267) (139.35 ± 45.47)	—
Número de brazo en tricomas estrellados	6 – 10 (7.73 ± 0.96)	6 – 16 (9.53 ± 2.31)	4 – 11 (7.58 ± 0.98)	—
Número de brazo en tricomas fasciculados	3 –8 (6.36 ± 1.45)	6 – 12 (8.53 ± 1.24)	4 – 10 (7.11 ± 1.02)	2 – 9 (6.86 ± 1.38)
Patrón filogenético	Polifilético	Polifilético	Monofilético	Monofilético
Hábitat	Bosque de pino-encino	Bosque de pino-encino	Bosque de encino seco	Bosque de encino, Ecotono entre bosque de encino y matorral xerófilo
Distribución geográfica	Región centro-este de la Faja Volcánica Transmexicana	Región norte de la Sierra Madre del Sur y región sur de la Sierra Madre Occidental	Sierra Madre Oriental	Región centro-sur de la Sierra Madre Oriental, Valle de Tehuacán-Cuicatlán y regiones montañosas del Altiplano Mexicano

CONCLUSIONES

La delimitación de especies necesita pruebas refutables para ser objetiva y empírica, de tal manera que establecer límites y proponer hipótesis de especies en grupos taxonómicamente conflictivos utilizando una sola línea de evidencia ocasionaría sesgos metodológicos y conceptuales que podrían requerir un monitoreo constante sobre la validez del taxon, especialmente en aquellos grupos donde procesos como la hibridación e introgresión son constantes. Asimismo, se debe insistir en la importancia de utilizar enfoques y herramientas que faciliten estudiar procesos de divergencia a nivel micro y macroevolutivo, y así obtener información que permita esclarecer las relaciones a un nivel taxonómico, pero bajo un claro contexto evolutivo.

Particularmente, este enfoque permitió aumentar el rigor en la delimitación de especies dentro del género *Quercus*, remarcando que el uso de marcadores de cloroplasto podría no ser de los más indicados al momento de establecer límites dentro de complejos de especies de encinos. En contraste, el uso de marcadores de microsatélites en conjunto con una evaluación morfométrica podría servir como un primer paso en el descubrimiento de especies de encinos a pesar de su propensión a la hibridación, sin embargo, es importante señalar que la evaluación de cualquier hipótesis de especie debe realizarse bajo un marco filogenético. En este contexto, el uso del enriquecimiento dirigido permitió detectar genes ortólogos útiles para la estimación filogenómica, que en conjunto con evidencia de diferenciación de nicho climático soportan la hipótesis sobre la existencia de múltiples linajes dentro del complejo *Quercus laeta*, los cuales están restringidos a ciertas regiones geográficas y con diferentes afinidades climáticas. Finalmente, mediante una evaluación cuantitativa a nivel micromorfológico sugerimos caracteres cuantitativos y cualitativos que podrían tener un valor taxonómico para diferenciar entre especies, de modo que proponemos retomar la importancia de los tricomas estrellados y considerar la densidad de tricomas como características clave que en conjunto con la longitud del brazo y la variación en el número de brazos de los tricomas brindan caracteres taxonómicos confiables que pueden servir como una herramienta adicional en la correcta identificación de estos linajes. Considerando toda la evidencia obtenida a partir de múltiples disciplinas proponemos que el complejo *Quercus*

laeta en realidad está compuesto por cuatro diferentes especies: *Quercus centralis*, *Quercus prinopsis*, *Quercus laeta* y *Quercus transmontana*, cada una con sus propias características micro morfológicas, afinidades climáticas y tendencias evolutivas.

El presente estudio es un esfuerzo por contribuir al paradigma de que el rigor en la delimitación de especies debe residir en determinar qué parte de la variación observada en poblaciones (a nivel genómico, ecológico, fenotípico, funcional o, cualquier otra evidencia empírica), se explica mejor por la filogenia que por cualquier otro proceso. Este enfoque permitirá que la clasificación sea cada vez menos arbitraria y cada vez más replicable, permitiendo poner a prueba los límites de las especies, lo que proporciona un marco justo y riguroso para la investigación y la conservación.

Literatura Citada

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS., Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Möst M, Mullen S, Nichols R, Nolte AW, Parisod C, Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura, JM, Väinolä R, Wolf JBW, Zinner D. 2013. Hybridization and speciation. *J. Evol. Biol.* 26: 229-246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Albarrán-Lara A, Petit R, Kremer A, Caron H, Peñaloza-Ramírez JM, Gugger PF, Dávila-Aranda PD, Oyama K. 2019. Low genetic differentiation between two morphologically and ecologically distinct giant-leaved Mexican oaks. *Plant Syst. Evol*. 305: 89 – 101. https://doi.org/10.1007/s00606-018-1554-8

Andrew RL, Rieseberg LH. 2013. Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution* 67: 2468 – 2482. https://doi.org/10.1111/evo.12106

Backs JR, Ashley MV. 2016. Evolutionary history and gene flow of an endemic island oak: *Quercus pacifica*. *Am. J. Bot*. 103: 2115 – 2125. https://doi.org/10.3732/ajb.1600259

Barrett, CF, Sinn BT, King LT, Medina JC, Bacon CD, Lahmeyer SC, Hodel DR. 2019. Phylogenomics, biogeography and evolution in the American genus *Brahea* (Arecaceae). *Bot. J. Linn. Soc.* 190: 242-259. https://doi.org/10.1093/botlinnean/boz015

Bello A, Labat N. 1987. Los encinos (*Quercus*) del Estado de Michoacán. Centre de Études Mexicaines et Centroaméricaines. Cuadernos de Estudios Michoacanos 4, Secretaría de Agricultura y Recursos Hidráulicos e Instituto Nacional de Investigaciones Forestales y Agropecuarias, México, D.F.

Brandvain Y, Kenney AM, Flagel L. 2014. Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genet.* 10: e1004410. https://doi.org/10.1371/journal.pgen.1004410

Buell CR. 2009. Poaceae genomes: going from unattainable to becoming a model clade for comparative plant genomics. *Plant Physiol*. 149: 111 – 116.

https://doi.org/10.1104/pp.108.128926

Camus, A. 1936-1954. Les Chênes: Monographie du genre Quercus (et Lithocarpus). Tome I-II-III. Paul Lechevalier & Fils éd, Paris.

Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013. How to fail at species delimitation. *Mol. Ecol.* 22: 4369-4383. https://doi.org/10.1111/mec.12413

Cavender-Bares J. 2006. Diversity, distribution, and ecosystem services of the North American oaks. International Oaks 27: 37 – 48.

Cavender-Bares J. 2019. Diversification, adaptation, and community assembly of the American oaks (*Quercus*), a model clade for integrating ecology and evolution. *New Phytol*. 221: 669 – 692. https://doi.org/10.1111/nph.15450

Cavender-Bares J, González-Rodríguez A, Eaton DA, Hipp A, Beulke A, Manos P. 2015. Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): a genomic and population genetics approach. *Mol. Ecol*. 24: 3668 – 3687. https://doi.org/10.1111/mec.13269

Chapman MA, Hiscock SJ, Filatov DA. 2013. Genomic divergence during speciation driven by adaptation to altitude. *Mol. Biol. Evol*. 30: 2553 – 2567. https://doi.org/10.1093/molbev/mst168

Chávez FZ. 1998. Observaciones sobre la distribución de encinos en México. *Polibotánica* 8: 47 – 64.

Conner WR, Delaney EK, Bronski MJ, Ginsberg PS, Wheeler TB, Richardson KM, Peckenpaugh B, Kim JK, Watada M, Hoffmann AA, Eisen BM, Kopp A, Cooper SB,Turelli M. 2021. A phylogeny for the *Drosophila montium* species group: A model clade for comparative analyses. *Mol. Phylogenet. Evol*. 158: 107061. https://doi.org/10.1016/j.ympev.2020.107061

Coyner BS, Murphy PJ, Matocq MD. 2015. Hybridization and asymmetric introgression across a narrow zone of contact between *Neotoma fuscipes* and *N. macrotis* (Rodentia: Cricetidae). *Biol. J. Linn. Soc*. 115: 162 – 172. https://doi.org/10.1111/bij.12487

Craft KJ, Ashley MV, Koenig WD. 2002. Limited hybridization between *Quercus lobata*

and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. *Am. J. Bot.* 89: 1792–1798. DOI: 10.3732/ajb.89.11.1792

Crowl AA, Manos PS, McVay JD, Lemmon AR, Lemmon EM, Hipp AL. 2020. Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*). *New Phytol.*, 226: 1158-1170. https://doi.org/10.1111/nph.15842

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23: 3133 – 3157. https://doi.org/10.1111/mec.12796

Curtu AL, Gailing O, Finkeldey R. 2007. Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evol. Biol.* 7: 1 – 15. https://doi.org/10.1186/1471-2148-7-218

Dayrat B. 2005. Towards integrative taxonomy. *Biol. J. Linn. Soc* 85: 407 – 417. https://doi.org/10.1111/j.1095-8312.2005.00503.x

De Beaulieu AIH, Lamant T. 2010. Guide Illustre des Chenes. Geer, Belgium: Edilens.

Degnan J, Rosenberg N. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 332 – 340. https://doi.org/10.1016/j.tree.2009.01.009

Denk T, Grimm GW, Manos PS, Deng M, Hipp AL. 2017. An updated infrageneric classification of the oaks: review of previous taxonomic schemes and synthesis of evolutionary patterns. In: Gil-Pelegrín E, Peguero-Pina JJ, Sancho Knapik D, eds. Tree physiology. Oaks physiological ecology. Exploring the functional diversity of genus Quercus L. Cham, Switzerland: Springer, 13–38.

Deng M, Jiang XL, Song YG, Coombes A., Yang, XR, Xiong YS, Li QS. 2017. Leaf epidermal features of *Quercus* Group Ilex (Fagaceae) and their application to species identification. *Rev. Palaeobot. Palynol.* 237: 10-36. https://doi.org/10.1016/j.revpalbo.2016.11.006

De La Torre AR, Roberts DR, Aitken SN. 2014. Genome-wide admixture and ecological niche modelling reveal the maintenance of species boundaries despite long history of interspecific gene flow. *Mol. Ecol.* 23: 2046 – 2059. https://doi.org/10.1111/mec.12710

De Queiroz K. 1998. The general lineage concept of species, species criteria, and the process

of speciation: A conceptual unification and terminological recommendations. Pages 57–75 in Endless forms: Species and speciation (D. J. Howard, and S. H. Berlocher, eds.). Oxford University Press, New York.

De Queiroz K. 2005. Different species problems and their resolution. *BioEssays* 27: 1263 – 1269. https://doi.org/10.1002/bies.20325

De Queiroz K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56: 879- 886. https://doi.org/10.1080/10635150701701083

Dobzhansky T. 1937. Genetic nature of species differences. *Am. Nat.* 71: 404 – 420.

Duranton M, Allal F, Fraïsse C, Bierne N, Bonhomme F, Gagnaire PA. 2018. The origin and remolding of genomic islands of differentiation in the European sea bass. *Nat. Commun*. 9: 1 – 11. https://doi.org/10.1038/s41467-018-04963-6

Eaton DA, Hipp AL, González-Rodríguez A, Cavender-Bares J. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution* 69: 2587 – 2601. https://doi.org/10.1111/evo.12758

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecol. Evol.* 27: 480-488. https://doi.org/10.1016/j.tree.2012.04.012

Gailing O, Curtu AL. 2014. Interspecific gene flow and maintenance of species integrity in oaks. *Ann. For. Res.* 57: 5 – 18. DOI: 10.15287/afr.2014.171

Gailing O, Lind J, Lilleskov E. 2012. Leaf morphological and genetic differentiation between *Quercus rubra* L. and *Q. ellipsoidalis* EJ Hill populations in contrasting environments. *Plant Syst. Evol*. 298: 1533 – 1545. https://doi.org/10.1007/s00606-012-0656-y

Garnett ST, Christidis L. 2017. Taxonomy anarchy hampers conservation. *Nature,* 546: 25-27. https://doi.org/10.1038/546025a

Gavrilets S. 2003. Perspective: models of speciation: what have we learned in 40 years? *Evolution*, 57: 2197-2215. https://doi.org/10.1111/j.0014-3820.2003.tb00233.x

Gernandt DS, Aguirre Dugua X, Vázquez-Lobo A, Willyard A, Moreno Letelier A, Pérez de la Rosa JA, Piñero D, Liston A. 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection Australes. *Am. J. Bot.* 105: 711-725. DOI:

10.1002/ajb2.1052

Goicoechea PG, Herrán A, Durand J, Bodénès C, Plomion C, Kremer A. 2015. A linkage disequilibrium perspective on the genetic mosaic of speciation in two hybridizing Mediterranean white oaks. *Heredity* 114: 373 – 386. https://doi.org/10.1038/hdy.2014.113

González R. 1993. La diversidad de los encinos mexicanos. *Rev. Soc. Mex. Hist. Nat*. 44: 125 – 142

González-Rodríguez A, Bain FJ, Golden L, Oyama K. 2004a. Chloroplast DNA variation in the *Quercus affinis* – *Q. laurina* complex in Mexico: geographical structure and association with nuclear and morphological variation. *Mol. Ecol*. 13: 3467 – 3476. https://doi.org/10.1111/j.1365-294X.2004.02344.x

González-Rodríguez A, Arias DM, Valencia S, Oyama K. 2004b. Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks. *Am. J. Bot*. 91: 401 – 409. https://doi.org/10.3732/ajb.91.3.401

González-Villarreal LM. 1986. Contribución al conocimiento del género *Quercus* (Fagaceae) en el estado de Jalisco. Flora de Jalisco. Instituto de Botánica, Universidad de Guadalajara. Guadalajara, Jalisco, México.

González-Villarreal LM. 2003a. Two new species of oak (Fagaceae, *Quercus* sect. *Lobatae*) from the Sierra Madre del Sur, Mexico. *Brittonia* 55: 49 – 60. https://doi.org/10.1663/0007-196X(2003)055[0049:TNSOOF]2.0.CO;2

González-Villarreal LM. 2003b. *Quercus tuitensis* (Fagaceae, *Quercus* sect. *Lobatae*) a new deciduous oak from western Jalisco, Mexico. *Brittonia* 55: 42 – 48. https://doi.org/10.1663/0007-196X(2003)055[0042:QTFQSL]2.0.CO;2

González-Villarreal LM. 2018. Dos nuevas especies de encinos (*Quercus*: Fagaceae), adicionales para la flora de Jalisco y áreas colindantes, en el Occidente de México. *Ibugana*, 9: 47-71.

Govaerts R, Frodin DG. 1998. World checklist and bibliography of Fagales (Betulaceae, Corylaceae, Fagaceae, and Ticodendraceae). Royal Botanical Garden, Kew.

Hardin J. 1975. Hybridization and Introgression in *Quercus alba*. *J. Arnold Arbor*. 56: 336 – 363.

Hart, MW. 2011. The species concept as an emergent property of population biology. *Evolution*. 65: 613-616.

Hausdorf, B. 2011. Progress toward a general species concept. *Evolution*, 65: 923-931.

Hauser DA, Keuter A, McVay JD, Hipp AL, Manos PS. 2017. The evolution and diversification of the red oaks of the California Floristic Province (*Quercus* section *Lobatae*, series Agrifoliae). *Am. J. Bot.* 104: 1581 – 1595. https://doi.org/10.3732/ajb.1700291

Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond.. B Biol Sci.* 270:313–321. DOI: 10.1098/rspb.2002.2218

Hey J. 2001. The mind of the species problem. *Trends Ecol. Evol*. 16: 326-329. https://doi.org/10.1016/S0169-5347(01)02145-0

Hey J. 2006. On the failure of modern species concepts. *Trends* Ecol. *Evol*. 21: 447 – 450. https://doi.org/10.1016/j.tree.2006.05.011

Hipp LA, Eaton AD, Cavender-Bares J, Fitzek E, Nipper R y Manos SP. 2014. A Framework phylogeny of the American oak clade based on sequenced RAD data. *PloS ONE* 9: 1-12. https://doi.org/10.1371/journal.pone.0102272

Hipp AL, Manos PS, González-Rodríguez A, Hahn M, Kaproth M, McVay JD, Valencia AS, Cavender-Bares, J. 2018. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. *New Phytol.* 217: 439-452. https://doi.org/10.1111/nph.14773

Hipp AL, Manos PS, Hahn M, Avishai M, Bodénes C, Cavender-Bares J, Crowl A, Deng M, Denk T, Fitz-Gibbon S, Gailing O, González- Elizondo MS, González-Rodríguez A, Grimm GW, Jiang XL, Kremer A, Lesur I, McVay JD, Plomion C, Rodríguez-Correa H, Schulze ED, Simeone MC, Sork VL, Valencia-Avalos S. 2020. Genomic landscape of the global oak phylogeny. *New Phytol.* 226: 1–15. 2019 https://doi.org/10.1111/nph.16162

Hong DY. 2020. Gen-morph species concept—A new and integrative species concept for outbreeding organisms. *J. Syst. Evol*. 58: 725-742.

Huang JP, Knowles LL. 2016. The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single Bayesian approach

in Hercules beetles. *Syst. Biol.* 65: 685-699. DOI: 10.1093/sysbio/syv119

Howard D, Preszler R, Williams J, Fenchel S, Boecklen W. 1997. How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelli*. *Evolution* 51: 747 – 755. https://doi.org/10.1111/j.1558-5646.1997.tb03658.x

Ishida TA, Hattori K, Sato H, Kimura MT. 2003. Differentiation and hybridization between *Quercus crispula* and *Q. dentata* (Fagaceae): insights from morphological traits, amplified fragment length polymorphism markers, and leafminer composition. *Am. J. Bot*. 90: 769 – 76. https://doi.org/10.3732/ajb.90.5.769

Jones JH. 1986. Evolution of the Fagaceae: the implications of foliar features. *Ann. Mo. Bot. Gard*. 73: 228 – 275. https://doi.org/10.2307/2399112

Kim BY, Wei X, Fitz-Gibbon S, Lohmueller KE, Ortego J, Gugger PF, Sork VL. 2018. RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: Fagaceae). *Mol. Ecol*. 27: 4556 – 4571. https://doi.org/10.1111/mec.14869

Kleinkopf JA, Roberts WR, Wagner WL, Roalson EH. 2019. Diversification of Hawaiian *Cyrtandra* (Gesneriaceae) under the influence of incomplete lineage sorting and hybridization. *J. Syst. Evol.* 57: 561-578. https://doi.org/10.1093/sysbio/syv119

Knapp S, Bohs L, Nee M, Spooner DM. 2004. Solanaceae – a model for linking genomics with biodiversity. *Compar. Funct. Genom*. 5: 285 – 291. https://doi.org/10.1002/cfg.393

Kopp A, Frank AK. 2005. Speciation in progress? A continuum of reproductive isolation in *Drosophila bipectinata*. *Genetica*, 125: 55-68. DOI: 10.1007/s10709-005-4787-8

Kremer A, Abbott A, Carlson J, Manos P, Plomion C. 2012. Genomics of Fagaceae. *Tree Genet. Genomes* 8: 583 – 610.

Kusi J, Karsai I. 2020. Plastic leaf morphology in three species of *Quercus*: The more exposed leaves are smaller, more lobated and denser. *Plant Species Biol*. 35: 24 – 37. https://doi.org/10.1111/1442-1984.12253

Lee MS. 2003. Species concepts and species reality: salvaging a Linnaean rank. *J. Evol. Biol*. 16: 179 –188. https://doi.org/10.1046/j.1420-9101.2003.00520.x

Liebmann FM. 1854. Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger og dets Medlemmers Arbeider. Copenhagen, 173.

Liu Z, Pagani M, Zinniker D, DeConto R, Huber M, Brinkhuis H, Shah SR, Leckie RM, Pearson A. 2009. Global cooling during the Eocene-Oligocene climate transition. *Science* 323: 1187 – 1190. DOI: 10.1126/science.1166368

Luo A, Ling C, Ho SY, Zhu CD. 2018. Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Syst. Biol.* 67: 830-846.

Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? *BioEssays*, 38: 140-149. https://doi.org/10.1002/bies.201500149

Manos PS, Doyle JJ, Nixon KC. 1999. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Mol. Phylogenet. Evol*. 12: 333 – 349. https://doi.org/10.1006/mpev.1999.0614

Manos PS, Hipp AL. 2021. An updated infrageneric classification of the North American Oaks (*Quercus* Subgenus *Quercus*): Review of the contribution of phylogenomic data to biogeography and species diversity. *Forests* 12: 786. https://doi.org/10.3390/f12060786

Manos PS, Meireles JE. 2015. Biogeographic analysis of the woody plants of the Southern Appalachians: implications for the origins of a regional flora. *Am. J. Bot*. 102: 780 – 804. https://doi.org/10.3732/ajb.1400530

Manos PS, Zhou ZK, Cannon HC. 2001. Systematics of Fagaceae: Phylogenetic tests of reproductive trait evolution. *Int. J. Plant Sci*. 162: 1361 – 1379. https://doi.org/10.1086/322949

Manos PS y Stanford AM. 2001. The historical biogeography of Fagaceae: tracking the Tertiary history of temperate and subtropical forest of the northern hemisphere. *Int. J. Plant Sci*. 162: 77 – 93. https://doi.org/10.1086/323280

Manos PS, Cannon HC y Oh SH. 2008. Phylogenetic relationships and taxonomic status of the paleoendemic Fagaceae of western North America: recognition of a new genus, Notholithocarpus. *Madroño* 55: 181 – 190. https://doi.org/10.3120/0024-9637-55.3.181

Martin SH, Dasmahapatra KK, Nadeau NJ. 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Res*. 23: 1817 – 28. doi: 10.1101/gr.159426.113

Maya-García R, Torres-Miranda A, Cuevas-Reyes P, Oyama K. 2020. Morphological

differentiation among populations of Quercus elliptica Née (Fagaceae) along an environmental gradient in Mexico and Central America. *Bot. Sci*. 98: 50 – 66. https://doi.org/10.17129/botsci.2395

Mayden RL, Claridge MF, Dawah HA, Wilson MR. A hierarchy of species concepts: The denouement in the saga of the species problem, Species: The units of biodiversity, 1997 London Chapman and Hall (pg. 381-424) Mayr E. 1982. Speciation and macroevolution. *Evolution* 36: 1119 – 1132.

McCauley RA, Cortés-Palomec AC, Oyama K. 2019. Species diversification in a lineage of Mexican red oak (*Quercus* section *Lobatae* subsection *Racemiflorae*)—the interplay between distance, habitat, and hybridization. *Tree Genet. Genomes* 15: 1-23.

Mccauley RA, Oyama K. 2020. A re-evaluation of taxonomy in Quercus section Lobatae subsection Racemiflorae (Fagaceae), resurrection of the name Q. pennivenia and description of a new taxon, *Q. huicholensis*. *Phytotaxa* 471: 247 – 257. https://doi.org/10.11646/phytotaxa.471.3.6

McIntyre DJ. 1991. Pollen and spore flora of an Eocene forest, eastern Axel Heiberg Island. N.W.T. Geol. Surv. Can. 403: 83 – 97. DOI:10.4095/131949

McVaugh R. 1974. Fagaceae. In: McVaugh R. ed. Flora Novo-galiciana. Contributions from the University of Michigan herbarium. 12: 50-30.

McVay DJ, Hipp LA, Manos PS. 2017a. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proc. R. Soc. B:Biol*. 284: 1 – 8. https://doi.org/10.1098/rspb.2017.0300

McVay JD, Hauser D, Hipp AL, Manos PS. 2017b. Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in western North America. *Genome* 60: 733 – 742. doi: 10.1139/gen-2016-0206

Mir C, Toumi L, Jarne P, Sarda V, Di Giusto F, Lumaret R. 2006. Endemic North African *Quercus afares* Pomel originates from hybridization between two genetically very distant oak species (*Q. suber* L. and *Q. canariensis* Willd.): evidence from nuclear and cytoplasmic markers. *Heredity* 96: 175 – 184. doi: 10.1038/sj.hdy.6800782.

Morales-Saldaña S, Valencia-Ávalos S, Oyama K, Tovar-Sánchez E, Hipp AL, González-Rodríguez A. 2022. Even more oak species in Mexico? Genetic structure and

morphological differentiation support the presence of at least two specific entities within *Quercus laeta*. *J. Syst. Evol. 60: 1124 – 1139*. https://doi.org/10.1111/jse.12818

Morii Y, Yokoyama J, Kawata M, Davison A, Chiba S. 2015. Evidence of introgressive hybridization between the morphologically divergent land snails *Ainohelix* and *Ezohelix*. *Biol. J. Linn. Soc*. 115: 77 – 95. https://doi.org/10.1111/bij.12466

Müller CH. 1942. The Central American species of *Quercus*. U.S. Dept. Agr. Misc. Publ. 477: 1 – 92.

Nixon KC. 1993a. Infrageneric classification of *Quercus* (Fagaceae) and typification of sectional names. *Ann. For. Sci*. 50: 25 – 34. DOI: 10.1051/forest:19930701

Nixon KC. 1993b. The genus *Quercus* in Mexico. En: Biological diversity of Mexico: origins and distribution (eds. Ramamoorthy, T.P., Bye R., Lot A. y Fa J.) Oxford University Press. New York USA, pp 447-458.

Nixon KC, Muller CH. 1993. The *Quercus hypoxantha* complex (Fagaceae) in northeastern Mexico. *Sida* 15: 57 – 69. https://doi.org/10.2307/2807497

Nixon KC, Jensen RJ, Manos P, Muller CH. 1997. Flora of North America, North of Mexico. Vol. 3 Magnoliophyta: Magnoliidae and Hamamelidae. Oxford University Press, Nueva York.

Nixon KC. 2006. Global and Neotropical distribution and diversity of oak (genus *Quercus*) and oak forests. In: Ecology and conservation of Neotropical montane oak forests. Ecological Studies vol 185 (M. Kappelle, ed), pp 3-13. Springer, Berlin.

Nixon KC, Barrie FR. 2017. Three previously undescribed species of *Quercus* (Fagaceae) from Mesoamerica and the designation of a lectotype for *Q. acutifolia*. *Novon* 25: 444 – 450. https://doi.org/10.3417/D-16-00014

Nosil P, Feder JL, Gompert Z. 2021. How many genetic changes create new species? *Science*, 371: 777-779. DOI: 10.1126/science.abf6671

O'Grady P, DeSalle R. 2018. Hawaiian *Drosophila* as an evolutionary model clade: Days of future past. *Bioessays* 40: 1700246. DOI: 10.1002/bies.201700246

Oh SH y Manos PS. 2008. Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear CRABS CLAW sequences. *Taxon* 57: 434 – 451.

Ortego J, Gugger PF, Riordan EC, Sork VL. 2014. Influence of climatic niche suitability and

geographical overlap on hybridization patterns among southern Californian oaks. *J. Biogeogr*. 41: 1895 – 1908. https://doi.org/10.1111/jbi.12334

Osada N, Wu CI. 2005. Inferring the mode of speciation from genomic data: a study of the great apes. *Genetics* 169: 259 – 264. DOI: 10.1534/genetics.104.029231

Padial JM, Miralles A, De la Riva I. et al. 2010. The integrative future of taxonomy. *Front. Zool*. 7: 16.

Pennisi E. 2016. Shaking up the tree of life. *Science* 354: 817. DOI: 10.1126/science.354.6314.817

Peñaloza-Ramírez JM, 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. Ann. Bot. 105: 389 – 399. doi: 10.1093/aob/mcp301

Pritchard KJ, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Rancilhac L, Irisarri I, Angelini C, Arntzen JW, Babik W, Bossuyt F, Künsel S, Lüddecke T, Pasmans F, Sanchez E, Weisrock D, Veith M, Wielstra B, Steinfartz S, Hofreiter M, Philippe H, Vences M. 2021. Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders. *Mol. Phylogenet. Evol.* 155: 106967.

Rangel SR, Carlos E, Zenteno R, de Lourdes M, Enríquez A. 2002. El género *Quercus* (Fagaceae) en el estado de México. *Ann. Mo. Bot. Gard.* 82: 551–593.

Rannala B, Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164: 1645 – 1656. DOI: 10.1093/genetics/164.4.1645

Renaut S, Grassa CJ, Yeaman S. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Commun*. 4: 1827. https://doi.org/10.1038/ncomms2833

Rittmeyer EN, Austin CC. 2012. The effects of sampling on delimiting species from multi-locus sequence data. *Mol. Phylogenet. Evol.* 65: 451-463.

Rodríguez-Correa H, Oyama K, MacGregor-Fors I, González-Rodríguez A. 2015. How are oaks distributed in the Neotropics? A perspective from species turnover, areas of endemism, and climatic niches. *Int. J. Plant Sci.* 176: 222-231.

Roe AD, Sperling FH. 2007. Population structure and species boundary delimitation of

cryptic *Dioryctria* moths: an integrative approach. *Mol. Ecol.* 16: 3617 – 3633. DOI: 10.1111/j.1365-294X.2007.03412.x

Romero RS. 2001. Estudio taxonómico de la serie Acutifoliae (*Quercus*, Fagaceae). Tesis doctoral. Facultad de Ciencias, Universidad Nacional Autónoma de México. México, D.F. 174 pp.

Rushton BS. 1993. Natural hybridization within the genus *Quercus* L. In Annales des sciences forestières (Vol. 50, No. Supplement, pp. 73s-90s). EDP Sciences.

Sabás-Rosales JL, Siqueiros-Delgado ME, Valencia-Ávalos S, Enríquez-Enríquez, ED. 2017. Reconocimiento taxonómico de seis especies arbustivas de encinos (*Quercus* secc. *Quercus*: Fagaceae). *Polibotánica* 44: 11 – 38.

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Reich. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 507: 354 – 357. https://doi.org/10.1038/nature12961

Scareli-Santos C, Sánchez-Mondragón ML, González-Rodríguez A, Oyama K. 2013. Foliar micromorphology of mexican oaks (*Quercus*: Fagaceae). *Acta Bot. Mex.* 104: 31 – 52.

Schumer M, Xu C, Powell DL. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* 360. DOI: 10.1126/science.aar3684

Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annu. Rev. Entomol*. 55: 421 – 438. doi: 10.1146/annurev-ento-112408-085432

Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodénes C, ... y Kremer A. 2004. Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* 168:1615 – 1626.

Shaffer HB, Thomson RC. 2007. Delimiting species in recent radiations. *Syst. Biol.* 56: 896 – 906. DOI: 10.1080/10635150701772563

Solís-Lemus C, Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet* 12: e1005896.

Spellenberg R. 1998. *Quercus lesueuri*, un miembro meridional del complejo de Q. X undulata (Fagaceae, subgenera *Quercus*). *Acta Bot. Mex.* 42: 25 – 33.

https://doi.org/10.21829/abm42.1998.795

Spellenberg R y Bacon JR. 1996. Taxonomy and distribution of a natural groups of black oaks of Mexico (*Quercus*, section *Lobatae*, subsection Racemiflorae). Syst. Bot. 21: 85 – 99. https://doi.org/10.2307/2419565

Stankowski S, Ravinet M. 2021. Defining the speciation continuum. *Evolution*, 75: 1256-1273.

Singhal S, Huang H, Grundler MR, Marchán-Rivadeneira MR, Holmes I, Title PO, Donnellan CS, Rabosky DL. 2018. Does population structure predict the rate of speciation? A comparative test across Australia's most diverse vertebrate radiation. *Am. Nat.* 192: 432-447.

Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat. Ecol. Evol.* 3: 170-177.

Thomson SA, Pyle RL, Ahyong ST, Alonso-Zarazaga M, Ammirati J, Araya JF,… Segers H. 2018. Taxonomy based on science is necessary for global conservation. *PLoS Biol.*, 16: e2005075. https://doi.org/10.1371/journal.pbio.2005075

Tovar-Sánchez E, Oyama K. 2004. Natural Hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: morphological and molecular evidence. *Am. J. Bot*. 91: 1352 – 1363. DOI: 10.3732/ajb.91.9.1352

Trelease W. 1924. A model for chloroplast capture. *Mem. Nat. Acad. Sci.* 20: 89–1782.

Turner TL, Hahn MW, Nuzhdin SV. 2005. Genomic islands of speciation in Anopheles gambiae. *PLoS Biol*. 3: 1572 – 8. https://doi.org/10.1371/journal.pbio.0030285

Valencia AS. 1995. Contribución al conocimiento del género *Quercus* (Fagaceae) en el Estado de Guerrero, México. Contribuciones del Herbario de la Facultad de Ciencias No. 1, Universidad Nacional Autónoma de México, México, D.F.

Valencia-A S. 2004. Diversidad del género *Quercus* (Fagaceae) en México. *Bol. Soc. Bot. Mex.* 75: 33–53.

Valencia-A S. 2020. Species delimitation in the genus *Quercus* (Fagaceae). *Bot. Sci.* 99: 1-12.

Valencia S, Coombes AJ. 2020. Nomenclatural revision and lectotypification of five species of Mexican oaks: *Quercus* (Fagaceae). *Phytotaxa*, 428: 81-92.

Valencia-Cuevas L, Mussali-Galante P, Piñero D, Castillo-Mendoza E, Rangel-Altamirano

G, Tovar-Sánchez E. 2015. Hybridization of Quercus castanea (Fagaceae) across a red oak species gradient in Mexico. *Plant Syst. Evol.* 301:1085 – 1097. https://doi.org/10.1007/s00606-014-1151-4

Valencia-A S, Flores-Franco G y Jiménez-Ramírez J. 2015. A nomenclatural revision of *Quercus acutifolia*, *Q. conspersa* and *Q. grahamii* (Lobatae, Fagaceae) *Phytotaxa*, 218: 289 – 294. DOI: https://doi.org/10.11646/phytotaxa.218.3.7

Valencia AS y Lozada LP. 2003. *Quercus nixoniana* (Fagaceae), una nueva especie de la Sección *Lobatae*, de la Sierra Madre del Sur, México. *Novon* 261-264. https://doi.org/10.2307/3393528

Valencia S, Rosales JLS, Arellano OJS. 2016. A new species of *Quercus*, section *Lobatae* (Fagaceae) from the Sierra Madre Oriental, Mexico. *Phytotaxa*, 269: 120-126.

Vázquez ML, Nixon KC. 2013. Taxonomy of *Quercus crassifolia* (Fagaceae) and morphologically similar species in Mexico. *Brittonia* 65: 209 – 227.

Wang X, He Z, Shi S, Wu C I. 2020. Genes and speciation: is it time to abandon the biological species concept? Nat. Sci. Rev. 7: 1387 – 1397. https://doi.org/10.1093/nsr/nwz220

Wheeler QD. 2004. Taxonomic triage and the poverty of phylogeny. *Philos. Trans. R. Soc. B. Biol. Sci.* 359: 571-583.

Wu CI. 2001. The genic view of the process of speciation. *J. Evol. Biol.* 14: 851-865.

Wu CI, Ting CT. 2004. Genes and speciation. Nat. Rev. Genet. 5: 114 – 122. https://doi.org/10.1038/nrg1269

Yang J, Di X, Meng X, Feng L, Liu Z, Zhao G. 2016. Phylogeography and evolution of two closely oak species (Quercus) from north and northeast China. *Tree Genet. Genomics* 12: 89. https://doi.org/10.1007/s11295-016-1044-5

Zachos FE. 2016. Species concepts in biology. Historical development, theoretical foundations and practical relevance. Springer, Cham.

Zachos FE.2018. (New) Species concepts, species delimitation and the inherent limitations of taxonomy. *J. Genet.* 97, 811-815.

Zhang X, Li Y, Liu C, Xia T, Zhang Q, Fang Y. 2015. Phylogeography of the temperate tree species Quercus acutissima in China: Inferences frm chloroplast DNA variations. *Biochem. Syst. Ecol*. 63: 190-197. https://doi.org/10.1016/j.bse.2015.10.010