UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS
INSTITUTO DE BIOLOGÍA

ANÁLISIS FILOGENÉTICO DE *PINUS* SUBSECCIÓN *CEMBROIDES* ENGELM. A PARTIR DE DATOS MULTI LOCUS

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

JOSÉ RUBÉN MONTES MONTIEL

TUTOR PRINCIPAL DE TESIS: DR. DAVID SEBASTIAN GERNANDT
INSTITUTO DE BIOLOGÍA, UNAM
COMITÉ TUTOR: DRA. CLAUDIA PATRICIA ORNELAS GARCÍA
INSTITUTO DE BIOLOGÍA, UNAM
COMITÉ TUTOR: DR. DANIEL IGNACIO PIÑERO DALMAU
INSTITUTO DE ECOLOGÍA, UNAM

Ciudad Universitaria, CD. MX. **2022**

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS
INSTITUTO DE BIOLOGÍA

ANÁLISIS FILOGENÉTICO DE *PINUS* SUBSECCIÓN *CEMBROIDES* ENGELM. A PARTIR DE DATOS MULTI LOCUS

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

JOSÉ RUBÉN MONTES MONTIEL

TUTOR PRINCIPAL DE TESIS: DR. DAVID SEBASTIAN GERNANDT
INSTITUTO DE BIOLOGÍA, UNAM
COMITÉ TUTOR: DRA. CLAUDIA PATRICIA ORNELAS GARCÍA
INSTITUTO DE BIOLOGÍA, UNAM
COMITÉ TUTOR: DR. DANIEL IGNACIO PIÑERO DALMAU
INSTITUTO DE ECOLOGÍA, UNAM

Ciudad Universitaria, CD. MX. **2022**

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE BIOLOGÍA

OFICIO CPCB/772/2022

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 1**6 de mayo de 2022** se aprobó el siguiente jurado para el examen de grado de **DOCTOR EN CIENCIAS** del estudiante **MONTES MONTIEL JOSÉ RUBÉN** con número de cuenta **517007806** con la tesis titulada **"ANÁLISIS FILOGENÉTICO DE *PINUS* SUBSECCIÓN *CEMBROIDES* ENGELM. A PARTIR DE DATOS MULTI LOCUS",** realizada bajo la dirección del **DR. DAVID SEBASTIAN GERNANDT,** quedando integrado de la siguiente manera:

Presidente:	DRA. ALEJANDRA CITLALLI MORENO LETELIER
Vocal:	DRA. ALEJANDRA VÁZQUEZ LOBO YURÉN
Vocal:	DR. JORGE ALBERTO PÉREZ DE LA ROSA
Vocal:	DRA. CAROLINA GRANADOS MENDOZA
Secretario:	DRA. CLAUDIA PATRICIA ORNELAS GARCÍA

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 26 de agosto de 2022

COORDINADOR DEL PROGRAMA

DR. ADOLFO GERARDO NAVARRO SIGÜENZA

POSGRADO EN CIENCIAS BIOLOGICAS
COORDINACIÓN

Agradecimientos

En primer lugar, emito un enorme agradecimiento al Posgrado en Ciencias Biológicas de la Universidad Nacional Autónoma de México por el constante e incalculable apoyo durante mi estancia doctoral desde el primer día como estudiante hasta el día de hoy.

En seguida, quiero agradecer a la máxima autoridad de fomento a la investigación en México, el Consejo Nacional de Ciencia y Tecnología por la beca doctoral recibida para realizar mis estudios de posgrado e investigación (beca CONACyT 230807). Del mismo modo, quiero hacer un especial agradecimiento a PAPIIT-DGAPA, UNAM (IN209816) por el financiamiento otorgado para la realización de este proyecto.

Finalmente, al estimado Dr. David Sebastian Gernandt por la oportunidad de trabajar juntos y por todo su apoyo incondicional tanto académico, económico y moral. ¡Mil gracias, Dave! Del mismo modo, agradezco infinitamente a la Dra. Patricia Ornelas García y al Dr. Daniel Piñero Dalmau por sus excelentes observaciones, correcciones y atinados comentarios hacia el proyecto en cada uno de mis comités tutorales y durante la revisión de esta tesis.

Agradecimientos a título personal

Considero que una de las primeras personas que merece un enorme agradecimiento es mi amigo, tutor y maestro Arturo Estrada Torres, quien siempre confío en mí, me apoyó en todo momento y me incursionó en el ámbito de la investigación. Mil gracias Arturo por compartir tu conocimiento y tu tiempo. Tu has sido una pieza importante en mi vida académica.

Papá, Mamá, Mirla, Nadia y Mario, agradezco encarecidamente su confianza y apoyo. Gracias por los consejos y la paciencia que me han brindado todos estos años. Gracias por ser una motivación en formación académica. Siéntase felices y orgullosos de lo que hemos logrado juntos.

Querida Angélica, mil gracias por acompañarme en gran parte de este proyecto. Atesoré que estuvieras ahí desde el inicio de esta etapa Doctoral hasta la culminación de la misma. No tengo palabras para agradecer todo lo que hiciste por mí en su momento. Aunque ya no estés más en mi vida debes saber que este logro académico también es tuyo. Donde quiera que te encuentras te deseo lo mejor. Por siempre RIA.

Agradezco infinitamente a la familia Castolo Pérez por todo el apoyo, el cariño, las atenciones y la hospitalidad que me brindaron. En especial a la Sra. Angélica Pérez Reyes, quien para mi siempre será un ejemplar mujer.

Sin lugar a duda, quiero agradecer a cada uno de mis maestros y maestras del posgrado en ciencias biológicas por compartir su conocimiento y su amistad. En especial a la Dra Alejandra Moreno Letelier, Dra. Alicia Mastretta Yanes, Dr. Ricardo García Sandoval, Dra. Susana Magallón Puebla, Dr. Gerardo Salazar Chávez y al Dr. Jorge Pérez de la Rosa.

Agradezco encarecidamente a la Dra. Xitlalli Aguirre-Dugua y a todo el equipo de trabajo del Dr. David S. Gernandt por su experiencia y apoyo para la realización de mi proyecto.

Finalmente, quiero agradecer a todas aquellas personas que de alguna manera formaron parte de esta etapa en mi vida.

IN MEMORIAM

Marcelina Cruz Ofelia Reyes Velazco
(1940 – 2018)

María Hernández y Barrientos
(1932 – 2020)

Josefina Irene Aguilera Tapia
(1936 – 2020)

Rosalía Amalia Ortega Arellano
(1964 – 2021)

José Rubén Montes

ÍNDICE

LISTA DE FIGURAS

LISTA DE TABLAS

RESUMEN

ANÁLISIS FILOGENÉTICO DE *PINUS* SUBSECCIÓN *CEMBROIDES*

ENGELM. A PARTIR DE DATOS MULTI LOCUS

Resumen

Pinus subsección *Cembroides* es un grupo taxonómico de árboles y arbustos con aproximadamente 15 taxa restringidos a ambientes áridos o semiáridos desde el suroeste de Estados Unidos, hasta el sureste de Puebla en México. *Pinus* subsección *Cembroides* ha sido objeto de estudio en múltiples trabajos taxonómicos, bioquímicos, evolutivos y ecológicos. Las especies de *Pinus* subsección *Cembroides* representan un modelo para explorar causas de incongruencia filogenética como sorteo incompleto de linajes (ILS, por sus siglas en inglés) y evolución reticulada, debido a sus largos tiempos generacionales, grandes tamaños efectivos poblacionales, polinización por viento y barreras reproductivas débiles. Este grupo también ofrece una buena oportunidad para poner a prueba la delimitación de especies debido a la persistente incertidumbre taxonómica y la falta de límites claros entre algunas especies dentro de la subsección. *Pinus* subsección *Cembroides* también presenta la oportunidad de estimar los tiempos de divergencia y calcular las tasas de diversificación en un grupo donde aún no se comprende bien su dinámica de diversificación. Afortunadamente, la creciente disponibilidad de datos moleculares y nuevos enfoques ha permitido re-estimar los tiempos de divergencia aún inciertos en varios grupos taxonómicos. Por la misma razón, el desarrollo de métodos coalescentes ha aumentado el interés por estudiar delimitación de linajes e incongruencia filogenética en pinos y otros linajes de plantas. Esta era de datos genómicos y métodos coalescentes hacen posible alcanzar el objetivo central de este estudio que fue inferir las relaciones evolutivas, así como explorar eventos de incongruencia filogenética como ILS y evolución reticulada e inferir los límites entre especies en *Pinus* subsección *Cembroides*, a partir de secuencias de 308 genes nucleares de baja copia generados por el método Hyb-Seq.

Los resultados del primer capítulo incluyen información de los dos primeros objetivos específicos de esta tesis: 1) Inferir las relaciones evolutivas de las especies de *Pinus* subsección *Cembroides* mediante análisis de coalescencia y 2) Explorar la importancia relativa del sorteo incompleto de linajes y evolución reticulada como causas de discordancia filogenética en *Pinus* subsección *Cembroides.* Los análisis basados en coalescencia (ASTRAL, Phylonet, SVDquartets) coincidieron en recuperar la monofilia de *Pinus* subsección

Cembroides y soportaron la mayoría de las relaciones evolutivas dentro de la subsección que han sido previamente reportadas con ADN de cloroplasto. El análisis de ASTRAL-III fue consistente con la presencia de niveles altos de ILS en el grupo de pinos piñoneros con conos pequeños. Finalmente, los análisis que toman en cuenta tanto ILS, como evolución reticulada, identificaron algunos escenarios de introgresión inesperados que no se habían reportado en la literatura.

El segundo capítulo contiene información sobre el tercer objetivo específico del proyecto: 3) Inferir los límites entre especies de *Pinus* subsección *Cembroides* a partir de métodos coalescentes multiespecie. Los resultados obtenidos demostraron inconsistencias entre los métodos coalescentes aplicados (i.e. GMYC, PTP y Tr2) debido a eventos de sobreestimación y agrupación, donde algunos individuos de una especie morfológica se dividen en varios linajes independientes o individuos de una especie morfológica se fusionan en un linaje independiente. De tal forma que el análisis GMYC basado con ADN del cloroplasto recuperó un mayor número de especies que PTP y que Tr2. Por el contrario, PTP y GMYC fueron incapaces de delimitar algunas especies que se identifican fácilmente con morfología. Finalmente, los resultados del método Tr2 coincidieron mejor con delimitaciones previas basadas en morfología, ADN, geografía y caracteres bioquímicos en reconocer 13 especies.

El último capítulo contiene información de los dos últimos objetivos específicos de la tesis: 4) Re-estimar los tiempos de divergencia dentro de *Pinus* subsección *Cembroides* utilizando genes nucleares de baja copia y todas las especies existentes y 5) Estimar las tasas de diversificación en *Pinus* subsección *Cembroides*. Se realizó un análisis de evidencia total para evaluar la posición filogenética de los fósiles y se calcularon relojes moleculares Bayesianos bajo el enfoque de fechamiento de nodos. El árbol de reloj molecular estimó que el origen de *Pinus* subsección *Cembroides* ocurrió durante el Oligoceno tardío (~25.5 millones de años) mientras que su diversificación ocurrió durante el Mioceno medio (~14.0 millones de años), coincidiendo con los resultados generados previamente con datos de un pequeño número de loci. El análisis macroevolutivo detectó un incremento en la tasa de especiación en *Pinus* subsección *Cembroides* y una desaceleración en las tasas de especiación en sus

grupos hermanos. En términos generales, las tasas de extinción se han mantenido constantes a lo largo del tiempo en *Pinus* subsección *Cembroides*.

Abstract

Pinus subsection *Cembroides* is a group of small to medium-sized trees and shrubs with approximately 15 taxa restricted to arid or semi-arid environments extending from the southwestern United States to the Mexican state of Puebla. *Pinus* subsection *Cembroides* has been in the subject of multiple taxonomic, biochemical, evolutionary, and ecological studies. The group offers an opportunity to explore the relative importance of incomplete lineage sorting (ILS) and reticulation as causes of phylogenetic discordance due to longer generation times, large population sizes, wind-dispersed pollen, and weak barriers to interspecific gene flow. *Pinus* subsection *Cembroides* also offers a good opportunity to study the boundaries among species. *Pinus* subsection *Cembroides* also offers the opportunity to estimate the divergence times and diversification rates in a group where the dynamic diversification is still unclear. Fortunately, the increasing availability of molecular data and new approaches has permitted re-estimate the divergence time in several taxonomic groups. Similarly, the development of coalescent methods has increased interest in studying species delimitation and exploring the causes that promote the phylogenetic discordance in pines and other botanical groups. This era of both genomic data and coalescent methods make it possible to achieve the central objective of this study, which was to infer the evolutionary relationships, as well as study phylogenetic incongruence events and clarify species delimitation in *Pinus* subsection *Cembroides* from low-coy nuclear genes generated for the Hyb-Seq method.

The results of the first chapter include information on the firsts two specific goals of this thesis: 1) To infer the evolutionary relationships of species in *Pinus* subsection *Cembroides* from coalescence analyses and 2) to explore the relative importance of the incomplete lineages sorting and reticulation as consequences of phylogenetic discordance in *Pinus* subsection *Cembroides*. Coalescence-based analyses (ASTRAL, Phylonet, and SVDquartets) agreed in recovering *Pinus* subsection *Cembroides* as monophyletic and in recovering similar relationships among species as in previous plastid DNA-based studies. The

ASTRAL-III tree was consistent with the presence of very high levels of ILS in the group of pinyon pines with small cones. Analyses that account for both incomplete lineage sorting and reticulation identify some unexpected hybridization scenarios that were not reported in the literature.

The second chapter contains information on the third specific objective of the project: 3) to infer the species boundaries in *Pinus* subsection *Cembroides* using coalescent-based methods. The results show inconsistencies among GMYC, PTP, and Tr2 due to oversplitting and lumping events, where some individuals of one species are divided into more independent lineages or individuals of one species are fused into one independent lineage. GMYC analysis based on plastid DNA recovered a higher number of species than PTP and the multi-locus coalescent approach. In contrast, both PTP and GMYC were incapable of identifying some species that are readily identified morphologically. Finally, the results of the Tr2 method coincided well with previous delimitations based on morphology, DNA, geography, and secondary chemistry, supporting the recognition of 13 species.

The last chapter contains information on the last two specific objectives of the thesis: 4) to re-estimate divergence times within *Pinus* subsection *Cembroides* using low-copy nuclear genes and all extant species, and 5) to estimate the diversification rates in *Pinus* subsection *Cembroides*. A total evidence analysis was performed to assess the phylogenetic position of the three fossils and Bayesian molecular clocks were calculated under the node dating approach. The clock-tree estimated that the origin of *Pinus* subsection *Cembroides* was dated to the Late Oligocene (about 25.5 million years ago), whereas their diversification occurred during the Middle Miocene (about 14.0 million years ago), supporting previously generated results with data from a small number of loci. Macroevolutionary analysis detected an increment in the rate of speciation in *Pinus* subsection *Cembroides* and a deceleration in speciation rates in close relatives. In general, the extinction rates have remained constant throughout the time of the phylogeny.

1. INTRODUCCIÓN

1.1. Generalidades de *Pinus* subsección *Cembroides*

Las especies del género *Pinus* L (Pinaceae) son árboles ampliamente distribuidos en el hemisferio norte con excepción de *P. merkusii* Jungh. & de Vriese, el cual se distribuye al sur del ecuador en la región de Sumatra (Critchfield y Little, 1966; Mirov, 1967). Las especies de *Pinus* en su mayoría se caracterizan por presentar hojas secundarias aciculares arregladas en fascículos de 2 hasta 8 hojas, excepto para *Pinus californiarum* Bailey, *P. monophylla* Torr. & Frém. y *P. edulis* var. *fallax* (Little) Businský, que poseen acículas solitarias. *Pinus* es un grupo monofilético que comprende dos subgéneros taxonómicos (*Pinus* y *Strobus* Lemmon) y cada subgénero con dos secciones: *Pinus* y *Trifoliae* Duhamel (*Pinus*) y *Parrya* Mayr y *Quinquefoliae* Duhamel (*Strobus*). La sección *Parrya* está dividida en tres subsecciones distribuídas en Norte América: *Balfourianae* Engelm. presente en el Oeste de Estados Unidos, *Nelsoniae* Van Der Burgh restringida a México y *Cembroides* Engelm. distribuída desde el suroeste de Estados Unidos hasta el estado de Puebla (Gernandt et al., 2005). *Pinus* subsección *Cembroides* comprende a los pinos piñoneros que son un grupo bien definido con aproximadamente 13 especies reconocidas (Fig. 1) (Farjon y Filer, 2013), siendo la segunda subsección con más especies reconocidas dentro del subgénero *Strobus*, solo después de la subsección *Strobus* con 21 especies (Price et al., 1998).

Las especies de *Pinus* subsección *Cembroides* son árboles de pequeño a mediano tamaño (1-30 metros) que se caracterizan por poseer conos ovulados pequeños en su mayoría (21 hasta 159 mm de longitud), con pedúnculo corto, escamas con umbo dorsal, apófisis gruesa y piramidal; semillas sin alas, comestibles, de 5 a 20 mm denominadas piñones de testa delgada o gruesa en algunas especies como *P. maximartinezii* Rzed. (Malusa, 1992). Excepto para *P. rzedowskii* Madrigal & M. Caball., donde las semillas son alargadas, con una testa delgada y alas funcionales (20-30 mm de longitud) (Madrigal y Caballero, 1969).

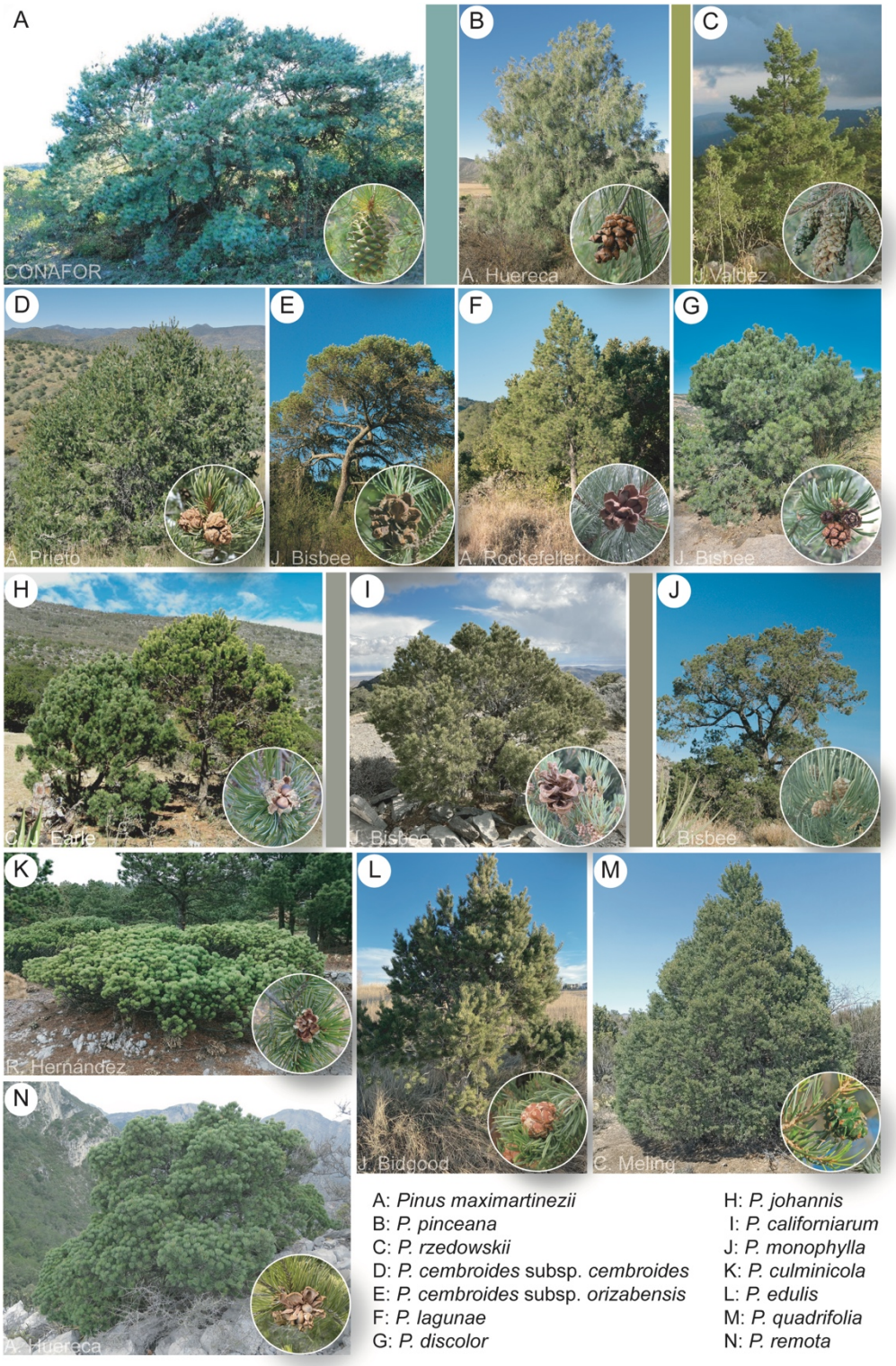


Figura 1. Pinos piñoneros de Norte América. Las imágenes muestran al árbol o arbusto y sus conos ovulados.

El primer pino piñonero reconocido para la ciencia fue *Pinus cembroides* Zucc., el cual fue descrito en 1832 a partir de una población ubicada en el estado de Hidalgo en México. Trece años más tarde, *Pinus monophylla* fue el primer pino piñonero reconocido en Estados Unidos. Siguieron *Pinus edulis* Engelm (1839), *P. pinceana* Gordon (1858), *P. quadrifolia* Parl. ex Sudw. (1897), *P. culminicola* Andresen & Beaman (1960), *P. maximartinezii* (1964), *P. remota* (Little) D.K. Bailey & Hawksw. (1966), *P. rzedowskii* (1969), *Pinus johannis* M.-F. Robert (1978), *P. discolor* D.K. Bailey & Hawksw. (1979) y finalmente fueron *P. californiarum* (1987) y *P. lagunae* (Robert-Passini) Passini (1987) los últimos en incorporarse al catálogo de los piñoneros de Norte América. Esencialmente, los piñoneros se encuentran restringidos a ambientes áridos o semiáridos desde el suroeste de Estados Unidos, hasta la región centro-sur de México (Critchfield y Little, 1966). En México, *Pinus cembroides* es la especie de piñonero más ampliamente distribuida seguido por *P. pinceana*, mientras que *P. edulis* es el piñonero más ampliamente distribuido en Estados Unidos seguido por *P. monophylla*.

En *Pinus* subsección *Cembroides* encontramos dos especies de gran importancia tanto económica como biológica, *Pinus cembroides* subsp*. cembroides* y *P. edulis* son importantes económicamente debido al alto consumo de sus semillas en México y Estados Unidos (Lanner, 1981; Farjon y Styles, 1997). El elevado valor nutritivo de las semillas de los pinos piñoneros juega un papel importante en el consumo y dispersión de semillas por aves (córvidos) y roedores. Particularmente, la reproducción de algunas especies de córvidos depende de la disponibilidad de semillas de los pinos piñoneros (Ligon, 1978). No obstante, muchas de las poblaciones de pinos piñoneros se encuentran en declive debido a la pérdida de hábitat, fragmentación, sequía extrema e incendios (Ledig et al., 2001; Suzán-Azpiri et al., 2002; Gitlin et al., 2006). Actualmente, más de la mitad de los taxa se encuentran en alguna categoría de riesgo de acuerdo con la IUCN (2021) y la Norma Oficial Mexicana NOM-059 (SEMARNAT, 2020; Tabla 1).

Tabla 1. Taxa de *Pinus* subsección *Cembroides* en alguna categoría de riesgo de acuerdo con la NOM-059 y la IUCN. Existe una falta de actualización taxonómica por parte de las dos fuentes de información a pesar de la evidencia reportada en los últimos años.

Taxón	SEMARNAT, 2020	IUCN, 2021
Pinus cembroides subsp. *cembroides* var. *bicolor* Little	Protección especial	
Pinus cembroides subsp. *lagunae* (Robert-Passini) D. K.Bailey	Protección especial	Vunerable
Pinus cembroides subsp. *orizabensis* D.K.Bailey		Amenazada
Pinus culminicola Andresen & Beaman	Peligro de extinción	Amenazada
Pinus maximartinezii Rzed.	Peligro de extinción	Amenazada
Pinus monophylla Torr. & Frém.	Protección especial	
Pinus pinceana Gordon	Peligro de extinción	
Pinus quadrifolia Parl. ex Sudw.	Protección especial	
Pinus remota (Little) D.K.Bailey & Hawksw.	Protección especial	
Pinus rzedowskii Madrigal & M. Caball.	Peligro de extinción	Vunerable

1.2. Estudios filogenéticos de *Pinus* subsección *Cembroides*

Los pinos piñoneros han sido sujetos a varias clasificaciones. El número de especies reconocidas difiere entre publicaciones (Tabla 2) y los rangos taxonómicos utilizados para pinos piñoneros varían ampliamente entre investigadores (Shaw, 1909; Malusa, 1992; Farjon y Styles, 1997; Gernandt et al., 2005; Eckenwalder, 2009; Farjon y Filer, 2013). Las discrepancias taxonómicas entre publicaciones son debidas al empleo de distintos caracteres tanto morfológicos, ecológicos y moleculares, así como diferencias en el muestreo (Malusa, 1992; Farjon y Styles, 1997; Gernandt et al., 2001, 2003, 2005; Syring et al., 2005; Parks et al., 2012). Basado en un análisis cladístico con caracteres morfológicos y ecológicos, Malusa (1992) dividió la subsección en dos linajes, un grupo con ocho especies de conos pequeños (<10 cm) y un segundo grupo de cuatro taxa con conos grandes (>10 cm). Similarmente, Farjon y Styles (1997) recuperaron la misma dicotomía de un grupo monofilético de los pinos de conos-pequeños y un grupo

parafilético de las especies de conos grandes como *P*. *maximartinezii*, *P*. *rzedowskii*, *P*. *pinceana*, incluyendo a *P*. *nelsonii.* Estas dos últimas especies se recuperaron como especies hermanas y ambas fueron incluidas en la subsección *Nelsoniae*, mientras que *P*. *rzedowskii* y *P*. *maximartinezii* fueron excluidos de la subsección por falta de información taxonómica (Farjon y Styles, 1997). No fue hasta un año después que Price et al. (1998) dividieron a los pinos piñoneros de Norte América en *Pinus* subsección *Cembroides* y en *Pinus* subsección *Rzedowskianae* Carvajal, siendo esta última una subsección monotípica.

En los últimos años la filogenia de *Pinus* subsección *Cembroides* ha sido inferida utilizando secuencias moleculares de ADN de plasto, nuclear-ribosomal y nuclear (Gernandt et al., 2001, 2003, 2005; Parks et al., 2012; Ortíz-Medrano et al., 2016). Los estudios filogenéticos de *Pinus* subsección *Cembroides* también han mostrado diferentes resultados en sus relaciones evolutivas y una falta de resolución filogenética debido al empleo de distintos caracteres moleculares y diferencias en el muestreo (Gernandt et al., 2001, 2003, 2005; Syring et al., 2005; Parks et al., 2012). Por citar algunos, se encuentra el trabajo de Gernandt et al. (2001) quienes analizaron la región del espaciador interno transcrito (ITS) del ADN ribosomal del núcleo (ADNr). Los autores encontraron que la región del ITS en especies de pinos es demasiado heterogénea, ya que muchas veces copias divergentes en individuos de la misma especie no agrupan juntas. En dos estudios posteriores con el empleo de secuencias de genes de plasto, Gernandt et al. (2003, 2005) lograron inferir las relaciones filogenéticas entre *P. culminicola*, *P. johannis* y *P. cembroides* subsp. *cembroides*, sugiriendo que *P. johannis* no es una subespecie de *P. cembroides* como se había sugerido en tratamientos taxonómicos previos.

Tabla 2. Taxones reconocidos de *Pinus* subsección *Cembroides* en diferentes clasificaciones.

Perry, 1991	Price et al., 1998	Gernandt et al., 2005	Eckenwalder, 2009	Farjon and Filer, 2013
Pinus catarinae				
P. cembroides	*P. cembroides* subsp. *cembroides*	*P. cembroides*	*P. cembroides*	*P. cembroides* subsp. *cembroides*
P. lagunae	*P. cembroides* subsp. *lagunae*		*P. lagunae*	*P. cembroides* subsp. *lagunae*
P. cembroides subsp. *orizabensis*	*P. cembroides* subsp. *orizabensis*		*P. cembroides* var. *orizabensis*	*P. cembroides* var. *orizabensis*
P. culminicola	*P. culminicola*	*P. culminicola*	*P. culminicola*	*P. culminicola*
P. discolor	*P. discolor*	*P. discolor*	*P. culminicola* var. *bicolor*	*P. cembroides* subsp. *cembroides* var. *bicolor*
P. edulis	*P. edulis*	*P. edulis*	*P. edulis*	*P. edulis*
P. johannis	*P. johannis*	*P. johannis*	*P. culminicola* var. *johannis*	
P. juarezensis	*P. juarezensis*			
	P. maximartinezii	*P. maximartinezii*	*P. maximartinezii*	*P. maximartinezii*
P. monophylla	*P. monophylla*	*P. monophylla*	*P. monophylla*	*P. monophylla*
	P. monophylla subsp. *californiarum*		*P. monophylla* subsp. *californiarum*	
	P. monophylla subsp. *fallax*		*P. monophylla* subsp. *fallax*	
	P. nelsonii			
	P. pinceana	*P. pinceana*	*P. pinceana*	*P. pinceana*
P. quadrifolia		*P. quadrifolia*	*P. quadrifolia*	*P. quadrifolia*
P. remota	*P. remota*	*P. remota*	*P. culminicola* var. *remota*	*P. remota*
		P. rzedowskii	*P. rzedowskii*	*P. rzedowskii*

Posteriormente, Parks et al. (2012) reportaron con el uso de plastomas casi completos que *P. johannis* no es un taxón infraespecífico de *P*. *cembroides* y fue recuperado como el grupo hermano de *P. culminicola*. Asimismo, Gernandt et al. (2003) reportaron con secuencias de los genes *rbcL* (ribulosa 1,5-bifosfato carboxilasa/oxigenasa) y *rpl16* (proteína ribosomal L16) que *P*. *juarezensis* es sinónimo de *P*. *quadrifolia* debido a que *P*. *quadrifolia* (incluyendo *P*. *juarezensis*) siempre se ubica como el grupo hermano de *P*. *monophylla*.

Otros taxones de *Pinus* subsección *Cembroides* que han sido clasificados en diferentes rangos taxonómicos son los pinos piñoneros con acículas solitarias que incluyen a *P*. *monophylla*, *P*. *californiarum* y *P*. *fallax.* En el pasado, *Pinus californiarum*

ha sido tratado como sinónimo o subespecie de *P*. *monophylla* (Farjon y Styles, 1997; Price et al., 1998), mientras que *P*. *fallax* ha sido tratado como variedad o subespecie de *P. californiarum*, *P*. *monophylla* o *P*. *edulis* (Silba, 1990). Gernandt et al. (2007) estudiaron a los pinos piñoneros de una sola acícula con ADN del plasto. Los autores lograron inferir la posición filogenética de *P*. *californiarum* como la especie hermana de *P*. *edulis* (dos acículas por fascículo). No obstante, las poblaciones de *P*. *fallax* no fueron muestreadas, por lo que considerarlas permitiría entender mejor la evolución de los pinos piñoneros de Norte América.

Los estudios sobre la clasificación y delimitación de especies en *Pinus* subsección *Cembroides* reportan algunas de las estrategias que se han empleado para estudiarlos. Los resultados demuestran lo difícil que ha sido establecer los límites entre algunos taxa como *Pinus discolor*, *P*. *johannis*, *P. lagunae, P*. *californiarum*, *P*. *fallax* y *P. monophylla*. Esto evidencia el conflicto existente dentro de *Pinus* subsección *Cembroides* donde en la actualidad aún carecemos de una hipótesis filogenética robusta y una clara delimitación de especies.

En los últimos años, los genes nucleares de bajo número de copia, que se presumen como marcadores independientes, con capacidad de poder aumentar la resolución filogenética (Syring et al., 2005), se han utilizado también para explorar mejor los problemas sobre incongruencia entre árboles de genes y árboles de especie (ver abajo). Un nuevo método de secuenciación que permite obtener miles de genes nucleares y secuencias casi completas de plasto o mitocondria es Hyb-Seq (Weitemier et al., 2014), el cual utiliza sondas para enriquecimiento y secuenciación masiva (Gnirke et al., 2009). En este estudio utilizamos y evaluamos el potencial que ofrecen los genes nucleares de baja copia y ADN de plasto para delimitar especies, inferir las relaciones evolutivas de las especies y re-estimar los tiempos de divergencia y calcular las tasas de diversificación en *Pinus* subsección *Cembroides*

1.3. Incongruencia filogenética

Reconstruir la historia evolutiva entre cualquier grupo de especies, genes o proteínas es uno de los principales objetivos de la biología evolutiva. Esta aproximación nos ha permitido estudiar las relaciones evolutivas, los patrones de especiación y divergencia, la delimitación de especies, la clasificación y organización de las especies son los árboles filogenéticos (O'Hara, 1992; Baum y Smith, 2012; Yang y Rannala, 2012). Los análisis filogenéticos han sido ampliamente usados para generar y probar hipótesis evolutivas empleando distintos tipos de caracteres tanto moleculares como morfológicos (Scotland et al., 2003). Particularmente, las filogenias con datos moleculares han sido importantes para entender la organización y evolución de genes y genomas (Bowers et al., 2003). En los últimos años la tecnología de secuenciación de nueva generación ha facilitado la generación de grandes conjuntos de datos genéticos que incluyen secuencias de los tres diferentes genomas de plantas: nuclear, mitocondrial y de plasto (Bräutigam y Gowik, 2010). A pesar del progreso en la generación de datos genómicos, la inferencia filogenética involucra aún desafíos que crean incertidumbre con respecto a las relaciones evolutivas de cualquier grupo de organismos o genes analizados (Huelsenbeck et al., 2000). Una causa de dicha incertidumbre es el análisis de distintos conjuntos de datos que producen filogenias que difieren unas de otras en uno o más detalles (Wendel y Doyle, 1998; Rokas et al., 2003) y este problema has sido apreciado desde hace tiempo (Doyle, 1992; Maddison, 1997).

En los últimos años el estudio de la incongruencia entre árboles individuales de genes y árboles de especies (presencia de conflicto topológico) ha recibido mayor atención en la biología evolutiva (Gernandt et al., 2018a; Stubbs et al., 2020; Wang et al., 2021; Rose et al., 2021; Cai et al., 2021; Dong et al., 2021), debido a que estos conflictos entre árboles de genes y árboles de especies impiden el objetivo de recuperar la historia evolutiva entre cualquier grupo de especies o genes. El conflicto entre árboles de genes y especies podría ser causado por uno o más procesos biológicos que afectan de distinta

manera a los diferentes conjuntos de datos (Fig. 2). Entre estos fenómenos biológicos se encuentra el sorteo incompleto de linajes (ILS por sus siglas en inglés) en el que dos o más linajes no logran unirse en su población ancestral más reciente (mirando hacia atrás en el tiempo). La duplicación y pérdida de genes (GDL, por sus siglas en inglés) que claramente implica la extinción o presencia extra de una copia de un gen en uno o varios linajes. Asimismo, se encuentra la transferencia horizontal de genes (HGT, por sus siglas en inglés) que involucra el intercambio interespecífico de material genético entre linajes (Maddison, 1997; Knowles et al., 2012).

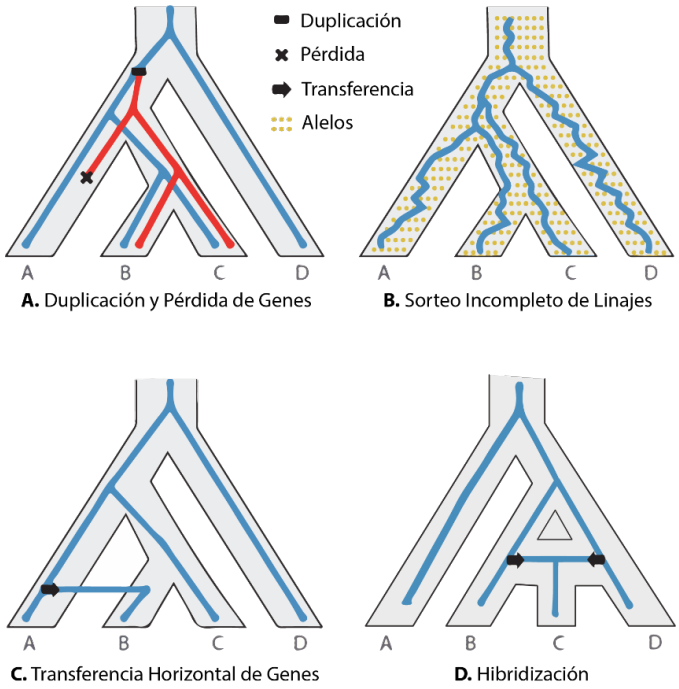


Figura 2. Procesos biológicos que causan incongruencia filogenética. Imagen tomada de Mirarab et al., 2021.

Otros procesos que causan incongruencia filogenética son la recombinación genética, la rápida diversificación y los fenómenos de evolución reticulada como introgresión e hibridización (Wendel y Doyle, 1998; Degnan y Rosenberg, 2009). La introgresión ocurre entre dos linajes cuando un evento de hibridación inicial es seguido por un retrocruzamiento en uno o ambos linajes parentales (Rieseberg y Wendel, 1993), mientras que la hibridización en sí misma es la creación de un individuo híbrido a partir de dos linajes parentales (Folk et al., 2018). También se ha reportado que la insuficiencia

de datos, el muestreo de genes y de taxones (Rokas y Caroll, 2005), así como errores en la secuenciación son causas técnicas que promueven el conflicto entre árboles de genes y de especies (Wendel y Doyle, 1998).

Particularmente, el ILS y la evolución reticulada han recibido mayor atención en los últimos años (Syring et al., 2007; Willyard et al., 2009; Gernandt et al., 2018a; Willyard et al, 2021). Por un lado, ILS es importante en organismos con grandes tamaños efectivos de población, tiempos de generación prolongados y especies que han sufrido radiaciones rápidas (Pamilo y Nei, 1988; Degnan y Rosenberg, 2009). Mientras que la reticulación a través de la hibridación o la captura de plasto desempeña un papel importante en aquellas especies que son polinizadas por viento (Rieseberg y Soltis, 1991).

De manera general, los árboles se caracterizan comúnmente por tener grandes tamaños efectivos de población, tasas de mutación y especiación más lentas, tiempos generacionales más prolongados y flujo de genes interespecífico (Petit y Hampe, 2006). Particularmente, los pinos son árboles longevos con grandes tamaños efectivos poblacionales (Syring et al., 2007). Su polinización es realizada por viento, con barreras reproductivas débiles, lo que permite que la hibridización aumente la proporción de alelos compartidos por especies reproductivamente compatibles (Mirov, 1967; Hong et al., 1993).

1.4. Métodos para estudiar incongruencia filogenética

En la última década se ha incrementado enormemente el desarrollo de varios métodos, algoritmos y aproximaciones para documentar la evidencia de la incongruencia filogenética (Kubatko et al., 2009; Liu et al., 2009; Than y Nakhleh, 2009; Mirarab et al., 2014; Roch y Steel, 2015; Zhang et al., 2018). Particularmente, ha habido un desarrollo importante en la teoría que aborda la estimación de árboles y redes filogenéticas cuando la fuente del conflicto entre árboles de especies y árboles de genes es ILS o reticulación. Esta teoría conocida como coalescencia está basada en un modelo matemático y

probabilístico que subyace a la historia evolutiva de los alelos. La esencia de la teoría es rastrear la historia de los alelos hacia atrás en el tiempo para identificar eventos que ocurrieron en el pasado desde el ancestro común más reciente de una muestra de alelos (Kingman, 1982). La teoría coalescente ahora es un tema central en genética de poblaciones (Hudson, 1990), que se ha implementado en sistemática (Pamilo y Nei, 1988). A partir del reconocimiento de la teoría de coalescencia, dos principales grupos de métodos se han desarrollado para explorar fuentes de incongruencia filogenética. Los métodos coalescentes multiespecie (MSC por sus siglas en inglés) (Hudson, 1983; Tajima, 1983; Pamilo y Nei, 1988; Maddison, 1997; Rannala et al., 2020) y la redes coalescentes multiespecie (Yu et al., 2014; Meng y Kubatko, 2009; Wen et al., 2016). La mayoría de los nuevos enfoques consideran explícitamente que los genes evolucionan dentro de un árbol de especies o una red filogenética de especies y las secuencias evolucionan a lo largo de los árboles de genes (Maddison, 1997; Degnan y Rosenberg, 2009).

Los métodos para estudiar ILS han sido sustancialmente más desarrollados. Este fenómeno biológico ha sido considerado el mayor desafío para la estimación de árboles de especies (Edwards, 2009). Los métodos para estudiar ILS se encuentran divididos en tres principales grupos: los métodos concatenados, de co-estimación y finalmente los métodos de resumen (Liu et al., 2008; Liu et al., 2009; Larget et al., 2010; Mossel y Roch, 2010; Sukumaran y Holder, 2010; Dasarathy et al., 2015; Vachaspati y Warnow, 2015).

Existe un método único en su categoría conocido como SVDquartets o Descomposición de Valores Singulares a partir de cuartetos de especies (Chifman y Kubatko, 2014). SVDquartets está implementado en PAUP* (Swofford, 2002) y solo es utilizado para modelar ILS. SVDquartets combina análisis concatenados bajo un marco coalescente para inferir las relaciones entre cuartetos de taxones que resulten en un árbol de especies, utilizando estadísticas algebraicas.

Los métodos Bayesianos implementados en BEAST y *BEAST co-estiman los árboles de genes y especies directamente de un conjunto de secuencias de genes alineados (Drummond y Rambaut, 2007; Liu, 2008). Estos métodos Bayesianos estiman árboles de especies incorporando la incertidumbre asociada con la estimación del modelo de sustitución de nucleótidos, la estimación del árbol de genes y el proceso de coalescencia (Liu, 2008). BEAST y *BEAST emplean las cadenas Markovianas para explorar la distribución de la probabilidad posterior del árbol de especies, calculando el mejor árbol en función de su probabilidad. También pueden estimar los tiempos de divergencia, así como los tamaños de las poblaciones ancestrales de un conjunto de árboles de genes (Liu y Pearl, 2007; Liu et al., 2008). Estos dos métodos son capaces de inferir árboles de especies, incluso cuando el árbol de especies se encuentra en "la zona de anomalía" (Liu y Edwards, 2009), que ocurre cuando existe una mayor frecuencia de discordancia entre los árboles de genes y el árbol de especies, que concordancia. A pesar de esta limitante, BEAST y *BEAST son precisos y estadísticamente consistentes (Leaché y Rannala, 2011). Sin embargo, computacionalmente son mucho más demandantes y hasta ahora no se han podido usar con cientos de genes (Knowles et al., 2012). Por lo tanto, no se pueden utilizar para análisis a escala filogenómica.

Los métodos de resumen sin lugar a duda son los más populares entre los métodos coalescentes multiespecie. Son más ampliamente usados para estudiar ILS y existen al menos una decena de ellos en la última década. De manera general, todos los métodos de resumen utilizan como archivo de entrada un conjunto de árboles de genes, de ahí su categoría. Estos métodos de resumen son estadísticamente consistentes y precisos a diferencia de los métodos concatenados y mucho menos demandantes computacionalmente que los métodos de co-estimación (Mirarab et al., 2016). Por tanto, se pueden utilizar para análisis a escala filogenómica. Particularmente, la precisión entre los métodos de resumen puede variar dependiendo el número de árboles de genes y el nivel de ILS (Mirarab et al., 2016). Otro aspecto que puede variar entre los métodos de resumen es la manera en como ellos resumen los tiempos de coalescencia. Algunos

métodos utilizan los tiempos de coalescencia promedio entre loci, otros calculan los rangos promedios de coalescencia entre loci o los tiempos de coalescencia mínimos (Liu et al., 2009; Mossel y Roch, 2010). También pueden variar en como ellos calculan el nivel de ILS; mientras unos utilizan cuartetos de concordancia otros utilizan factores de concordancia (CFs por sus siglas en inglés) (Larget et al., 2010; Mirarab et al., 2014). Otras características de los métodos de resumen hasta ahora descritos son que asumen que no hay recombinación dentro de los loci y que no existe el flujo de genes después de la especiación. También asumen que los árboles de genes están correctamente construidos. Entre los métodos más populares se encuentra la Estimación de Árboles de Especies bajo Máxima verosimilitud (STEM por sus siglas en inglés) (Kubatko et al., 2009), el método de Estimación de Árboles de Especies usando Rangos Promedio de coalescencia (STAR por sus siglas en inglés), el modelo de Estimación de Árboles de Especies usando Tiempos de Coalescencia Promedio (STEAC por sus siglas en inglés) (Liu et al., 2009), la Prueba de Separación Global o GLASS por sus siglas en inglés (Mossel y Roch, 2010), el método de Estimación de Árboles de Especies bajo Máxima Pseudo-verosimilitud (MP-EST por sus siglas en inglés) (Liu et al., 2010), BUCKy (Larget et al., 2010), el Algoritmo para estimar Árboles de Especies Precisos o ASTRAL por sus siglas en inglés (Mirarab et al., 2014; Mirarab y Warnow, 2015; Zhang et al., 2018; Yin et al., 2019) y el método de Árboles de Especies Precisos a partir de Distancias entre Nodos o ASTRID por sus siglas en inglés (Vachaspati y Warnow, 2015). De los métodos mencionados anteriormente, ASTRAL se encuentra entre los más populares debido a que es más preciso que MP-EST, BUCKy y CA-ML en muchas condiciones, excepto cuando la cantidad de ILS es baja (Mirarab et al., 2014). ASTRAL puede construir árboles de especies incluso con datos faltantes, politomías en árboles de entrada y la presencia de genes parálogos (Mirarab y Warnow, 2015; Zhang et al., 2018; Yin et al., 2019).

A diferencia de los métodos para explorar ILS, los métodos para estudiar redes filogenéticas son menos abundantes. Entre ellos destacan Phylonet (Than et al., 2008;

Than y Nakhleh, 2009), SNaQ (Redes de Especies usando Cuartetos) implementado en el paquete de Phylonetworks (Solís-Lemus y Ané, 2016), SpeciesNetwork (Zhan et al., 2018) implementado en BEAST 2 y MSci (Modelo Coalescente Multiespecie con introgresión) implementado en BPP (Flouri et al., 2020). Es preciso aclarar que algunos programas de redes filogenéticas utilizan como archivos de entrada árboles de genes previamente estimados por máxima verosimilitud o inferencia Bayesiana y/o los alineamientos individuales por cada gen. Aquellos métodos que estiman redes filogenéticas bajo inferencia Bayesiana utilizan las cadenas de Markov Monte Carlo o MCMC para explorar las mejores hipótesis de hibridización en función de su probabilidad posterior (Zhan et al., 2018). Todos los programas de redes son capaces de manejar miles de loci para múltiples individuos. Sin embargo, estos métodos de redes filogenéticas requieren mucha demanda computacional a una escala multilocus o genómica, principalmente cuando se trata de estimar redes filogenéticas con múltiples eventos de hibridización bajo inferencia Bayesiana o verosimilitud. Por último, Phylonet, SNaQ y SpeciesNetwork pueden estimar tanto eventos de evolución reticulada como ILS (Than y Nakhleh, 2009; Solís-Lemus y Ané, 2016; Zhan et al., 2018).

Existen otros tipos de métodos para estudiar alguno de los fenómenos de evolución reticulada. Por un lado, se encuentran los métodos llamados estadísticos *D* de Patterson que modelan procesos como hibridización o flujo de genes pero que no utilizan redes filogenéticas como ABBA-BABA (Green et al., 2010) y HyDe (Blischak et al., 2018; Kubatko y Chifman, 2019). Por el otro lado, se encuentran los modelos probabilísticos como Structure (Pritchard et al., 2000) y admixture (Alexander et al., 2009) que generan información sobre hibridización y flujo de genes a partir de métodos de agrupamiento que permiten asignar a un individuo a una población a partir de la frecuencia de genes ancestrales compartidos.

Es claro que con el paso del tiempo y con el continuo avance en la generación de datos genómicos, el desarrollo y perfección de nuevos métodos que permitan analizar los fenómenos de incongruencia filogenética de manera precisa continuará. Hasta ahora, los

métodos de árboles o redes filogenéticas bajo el modelo coalescente de especies han permitido explorar fenómenos de incongruencia filogenética que al inicio de este apartado describimos. Seguir estudiando dichos fenómenos permitirá generar más información y alcanzar el objetivo de reconstruir una hipótesis evolutiva más robusta entre cualquier grupo de especies, genes o proteínas del árbol de la vida.

1.5. Delimitación de especies en *Pinus* subsección *Cembroides*

1.5.1. Marcadores no moleculares

Tradicionalmente los límites entre especies del género *Pinus* se han inferido a partir de un pequeño número de caracteres morfológicos para discernir o incluir a ciertas entidades taxonómicas. Con el paso del tiempo diferentes fuentes de evidencia se han ido incorporando para delimitar especies de pinos como caracteres bioquímicos, anatómicos, sexuales y más recientemente caracteres moleculares (Price et al., 1998; Gernandt et al., 2001, 2003, 2005; Syring et al., 2005; Parks et al., 2012; Hernandez-León., 2013). El empleo de estas distintas fuentes de información ha provocado ajustes, modificaciones y discrepancias en cuanto al número de especies reconocidas en el género (Mirov, 1967; Perry, 1991; Farjon y Styles, 1997; Price et al., 1998; Gernandt et al., 2003, 2005). Estas diferencias se deben a una serie de factores de distinta índole que a continuación citaremos. Particularmente, la morfología puede no discriminar taxones debido a que la falta de variación, y puede enmascarar la presencia de especies crípticas en los complejos de especies de pinos (ver Silba, 1990; Vidakovic, 1991; Farjon y Style, 1997; Price et al., 1998; Businsky et al., 2014). Por un lado, se ha demostrado que la plasticidad de algunos caracteres morfológicos como lo es el número de hojas, así como la longitud de las hojas, conos y semillas dentro de una misma especie puede estar asociada a condiciones ambientales y geográficas tales como la precipitación y altitud (Poulos y Berlyn, 2007; Cole et al., 2008). Por el otro lado, los caracteres que muestran una mayor variación intraespecífica suelen presentar un elevado grado de homoplasia (Wiens, 1995). De hecho, uno de los caracteres más utilizados para delimitar especies

de pinos es el número de acículas por fascículo (Farjon y Styles, 1997), el cual ha sido reportado como homoplásico en el género (Gernandt et al., 2005).

En cuanto a los marcadores bioquímicos, los terpenos (metabolitos secundarios que dan las características organolépticas a los pinos) son otra fuente de información que se ha utilizado para separar especies de pinos (Zavarin y Snajberk, 1986; Sarac et al., 2013; Mitic et al., 2017). Uno de los primeros inconvenientes es la variabilidad en la composición de los terpenos la cual puede ser afectada por diversos factores tanto técnicos, de desarrollo y ecológicos (Hanover, 1992). Entre los factores de desarrollo se encuentra la composición de los terpenos en una fase de crecimiento del árbol y el tipo de órgano o tejido muestreado, ya que se ha demostrado que existen cambios significativos en la composición de los terpenos dentro del mismo tejido pero en diferente posición y estadio de un pino (Roberts, 1970). Desde el punto de vista técnico, el procesamiento del material vegetal, así como protocolo de aislamiento del terpeno son causas que impactan en los resultados de su composición debido a que existen miles de derivados de terpenos dentro de una misma especie. Finalmente, la concentración de los terpenos en los mismos sistemas de tejidos también puede cambiar durante las distintas estaciones del año (Hanover, 1992).

1.5.2. Marcadores moleculares

Las aloenzimas han sido otros marcadores que se han utilizado para estimar diferencias genéticas entre especies de pinos (Millar et al., 1988). La variación genética de este tipo de proteínas es conocida como polimorfismo “aloenzimático” y ha servido para examinar procesos genéticos en plantas (Weeden y Wendel, 1989). No obstante, la frecuencia de las aloenzimas dentro de una misma especie puede diferir entre estructuras morfológicas como se ha reportado en corteza y conos ovulíferos de pinos (Plessas y Strauss, 1986) y el grado de polimorfismo y diferenciación es muy bajo, tal que diferenciar poblaciones o especies es complicado (Wu et al., 1999).

Similarmente, el ADN Polimórfico por Amplificación Aleatoria (RAPD, por sus siglas en inglés) y los Polimorfismos en la Longitud de Fragmentos Amplificados (AFLPs)

tienen la capacidad de detectar diferencias a nivel de especie y entre poblaciones (Dvorak et al., 2000). En pinos se ha encontrado que el porcentaje de loci polimórficos de RAPD y AFLPs es mayor respecto a las aloenzimas (Szmidt et al., 1996). De acuerdo con Duminil y Di Michele (2009) estos marcadores son muy informativos y se han empleado para estudiar la incongruencia filogenética entre árboles de especies y árboles de genes, pues tienen la capacidad de revelar eventos de hibridización, lo cual es muy importante en las pináceas donde las barreras al flujo génico son débiles (Mirov, 1967; Hong et al., 1993). Ahora bien, delimitar complejos de especies o especies cercanamente relacionadas con rangos de distribución compartidos es difícil con RAPDs y AFLPs. Por ejemplo, Castro-Félix et al. (2008) no lograron diferenciar a las dos especies del complejo *Pinus strobiformis* Engelm. - *Pinus ayacahuite* C. Ehrenb. ex Schltdl. con 139 fragmentos polimórficos usando un método de distancia genética. Es sabido que las limitantes de estos marcadores están relacionadas con sus propiedades de dominancia y homoplasia (Després et al., 2003). Por si fuera poco, los RAPDs y AFLPs presentan una pobre reproducibilidad especialmente entre laboratorios (Bussel et al., 2005).

Uno de los marcadores genéticos más populares para estudiar y diferenciar especies de plantas a un nivel genético es la región ITS del ADN ribosomal (Álvarez y Wendel, 2003). Los ITS tienen la ventaja de presentar una alta proporción de sitios filogenéticamente informativos comparado con secuencia de ADN de plasto (Mort et al., 2007). De acuerdo con Pang et al. (2012), parte de la región del ITS (ITS2), se ha mostrado como el mejor locus de código de barras (regiones cortas de ADN universales) debido a su poder de discriminación de especies. De tal forma que la región del ITS fue propuesta como un novedoso código de barras en plantas (Li et al., 2011), no solo por su poder discriminatorio sino por su alta tasa de éxito de amplificación y obtención de secuencias en diferentes grupos de plantas en un laboratorio (Pang et al., 2012). Las secuencias de ITS presentan un fenómeno biológico conocido como evolución concertada (Baldwin et al., 1995). La evolución concertada homogeniza las diferentes copias de genes parálogos de ITS (Mort et al., 2007), sin embargo, cuando existe un gran

número de copias, la evolución concertada no opera lo suficientemente rápido. La presencia de genes parálogos puede acumularse, lo que pone limitaciones prácticas al uso y la interpretación de la variación de ITS (Syring et al., 2005). También se ha demostrado que el grado de homoplasia es más alto en ITS que en otros marcadores de ADN, muy probablemente debido a la combinación ortología/paralogía, cambios de bases compensatorias o alguna combinación de estos fenómenos (Álvares y Wendel, 2003). Finalmente, su uso para la discriminación de especies depende del tiempo de divergencia entre las especies y su tasa de evolución molecular en el linaje estudiado. Si la divergencia es demasiado reciente, estos marcadores podrían no ser informativos (Duminil y Di Michele, 2009). Por tanto, inferir los límites de especies con ITS en miembros de *Pinus* subsección *Cembroides* representa complicaciones debido a que la diversificación en este grupo de especies de pinos ha sido reportada como reciente (Saladin et al., 2017).

En pinos se han utilizado diferentes combinaciones de secuencias codificantes y no codificantes del ADN del plasto para discriminar especies (Zhou et al; 2010; Hernández-León, 2013). Dentro de los siete loci de plasto evaluados por parte del Consorcio de Código de Barras de la Vida (CBOL), marcadores como *matK* (maturasa involucrada en el procesamiento de intrones grupo II) y *rbcL* han sido considerados y recomendados como códigos de barras estándares debido a su universalidad (Hollingsworth, 2011). Estos marcadores universales tienen ciertos niveles de discriminación entre especies, por tanto, tienen una amplia aplicación (Lahaye et al., 2008). Son marcadores de fácil amplificación por PCR, longitud apropiada y buenos niveles de tasas de sustitución (Chase et al., 1993).

Además, los marcadores centrales sugeridos por el Consorcio de Códigos de Barra de la Vida (*rbcL*, *matK*, ITS, entre otros) tienen una baja tasa de discriminación de especies y deben complementarse con otros marcadores (Jeason et al., 2011; Hollingsworth, 2011). Una alternativa para incrementar el número de sitios informativos es el empleo del enfoque denominado “super código de barras” que consiste en generar

genomas completos de plasto (plastomas) (Erickson et al., 2008; Li et al., 2015). La secuenciación del genoma del plasto o plastoma es ahora relativamente sencilla, gracias a la disminución de los costos de secuenciación. Su empleo dejaría de lado algunas de las complejidades asociadas con bases de datos del Código de Barras de la Vida, las cuales se encuentran parcialmente superpuestas (Hollingsworth, 2011). Los plastomas en pináceas se distinguen de la mayoría de otros taxa por su reordenación significativa y su abundante contenido de ADN como genoma citoplásmico (Rieseberg y Soltis, 1991). Los plastomas pueden utilizarse para aclarar cuestiones taxonómicas (Whittall et al., 2010; Asaf et al., 2018) y su variabilidad mutacional en pinos es lo suficientemente heterogénea para poder estimar la divergencia entre especies (Whittall et al., 2010). Desde luego, existen inconvenientes prácticos, metodológicos y biológicos al momento de trabajar con plastomas. El procesamiento y ensamble de las lecturas de los plastomas representan algunos desafíos, como el confiar en un plastoma de referencia o realizar un ensamble *de novo* (Aguirre-Dugua y Gernandt, 2017). En este sentido, la calidad del ensamble dependería completamente del plastoma de referencia o del ensamble *de novo*. Por tanto, la delimitación de especies llevaría la incertidumbre del ensamble de los plastomas. Además de presentar herencia uniparental, existe evidencia que indica que el ADN de plasto en pinos experimenta tasas de sustitución relativamente bajas (Willyard et al., 2007). No obstante, se ha demostrado que con plastomas completos o casi completos existe suficiente variación total entre las especies, aunque la tasa de sustitución sea baja, logrando una buena discriminación entre taxones (Parks et al., 2012; Gernandt et al., 2018a).

Lo cierto es que el enfoque denominado “super código de barras” combinado con métodos formales para delimitar especies bajo el modelo coalescente no se ha empleado en ningún grupo de pinos y esta combinación podría ayudar a estudiar mejor los límites de especies en *Pinus* subsección *Cembroides*.

Una alternativa de marcadores que presentan herencia uniparental es el empleo de genes nucleares de baja copia (Sang, 2002). Los genes nucleares son marcadores

moleculares independientes con capacidad de aumentar la resolución filogenética, así como explorar mejor los problemas sobre incongruencia entre árboles de genes y árboles de especies, incorporando datos de ADN del plasto. Particularmente en pinos, se ha reportado que los genes nucleares de baja copia sí aumentan la resolución filogenética en la mayoría de los clados de todas las subsecciones del género *Pinus*, obteniendo clados más robustos y congruentes con clasificaciones previas basadas en morfología (Syring et al., 2005). Asimismo, se ha reportado que las tasas de sustitución en genes nucleares de baja copia en *Pinus* son aproximadamente diez veces más rápidas que los loci del ADN de plasto (4.1 vs. 0.04 sitios/año) (Willyard et al., 2007; De La Torre et al., 2017). Los marcadores independientes que muestran altas tasas de sustitución pueden proporcionar una resolución mayor que el ADN del plasto y ADN ribosomal para resolver relaciones entre especies estrechamente relacionadas, con radiaciones rápidas y eventos de hibridización histórica (Willyard et al., 2007). De esta manera, los genes nucleares de baja copia podrían ayudar a resolver no solo las relaciones entre especies de *Pinus* subsección *Cembroides*, donde la diversificación de linajes ha sido reciente (Gernandt et al., 2008), sino también estudiar el posible flujo de genes entre dichas especies. Al igual que el resto de los marcadores mencionados, los genes nucleares presentan inconvenientes de diferente índole. Como se ha mencionado anteriormente, los métodos de análisis y procesamiento de datos genómicos se encuentran desfasados de los métodos modernos de NSG. Por un lado, la profundidad de secuenciación y la calidad de las lecturas están asociadas al tipo de plataforma de secuenciación utilizada (Cronn et al., 2008). Mientras que el pre-procesamiento y ensamble de las lecturas de los genomas con o sin referencia requieren una alta demanda computacional y protocolos informáticos específicos. De igual forma, los errores e incertidumbres del ensamble y la detección de variantes tienen un impacto sobre cualquier clase de análisis que se realice con los datos ensamblados. Por el otro lado, la presencia de genes parálogos podrían obscurecer la delimitación de especies y el análisis evolutivo de cualquier grupo de especies (Palmé et al., 2009). A pesar de estos obstáculos, el empleo de múltiples genes

nucleares de baja copia combinado con métodos para delimitar especies bajo el modelo coalescente, tampoco se ha empleado en ningún grupo de pinos y esta combinación también podría ayudar a estudiar mejor los límites de especies en *Pinus* subsección *Cembroides*.

El reciente enfoque que combina sondas de enriquecimiento (Gnirke et al., 2009) y secuenciación masiva para capturar genes nucleares de baja copia, así como ADN de alto número de copias de plasto, mitocondria y ADN ribosomal del núcleo (Hyb-Seq; Weitemier et al., 2014), puede generar de manera eficiente una gran cantidad de datos a un costo razonable (Zimmer y Wen, 2015). Además, Hyb-Seq puede capturar facilmente secuencias de genes nucleares, ADN de plasto y ADN mitocondrial a partir del ADN total de muestras de herbario. Este enfoque de secuenciación de siguiente generación comienza a ser aceptado como una técnica prometedora en estudios moleculares en plantas y árboles (Gernandt et al., 2018a; Morales-Briones et al., 2018).

La creciente disponibilidad de datos moleculares y nuevos enfoques y métodos (Schlick-Steiner et al., 2010; Fujita et al., 2012; Zhang et al., 2013) ha generado el interés en delimitar especies de manera integradora. En la actualidad existe una gran cantidad de métodos disponibles para inferir los límites de las especies, pero su uso es poco común para delimitar especies de pinos o árboles en general. Entre los métodos más comunes se encuentra Structurama que puede inferir la estructura de una población a partir de Polimorfismos de un Solo Nucleótido o SNPs, por sus siglas en inglés (Huelsenbeck et al., 2011). También existen los métodos para discriminar especies mediante el análisis no multi locus como el modelo Mixto Generalizado de Yule y Coalescencia (GMYC; Pons et al., 2006), el método de Descubrimiento Automático con Código de Barras (ABDG; Puillandre et al., 2011), el modelo de Identificación y Evolución de Especies en R (SPIDER; Brown et al., 2012), y el método de Árboles con Procesos de Poisson (PTP; Zhang et al., 2013; Kapli et al., 2017). Similarmente se encuentran los métodos de delimitación de especies a nivel multilocus como Brownie (O 'Meara et al., 2006; O' Meara, 2010), spedeSTEM (delimitación de especies utilizando la Estimación

de Árboles de Especies bajo Máxima verosimilitud (Ence y Carstens, 2011), BPP (Filogenética y Filogeografía Bayesiana) (Yang, 2015), DISSECT (División de Individuos en Especies usando Secuencias y Árboles Colapsados de Epsilon) (Jones et al., 2015) y el método de distribución Trinomial o Tr2 (Fujisawa et al., 2016). Los métodos coalescentes brindan una alternativa atractiva para estudiar los procesos evolutivos que contribuyen a la especiación, inferir las relaciones entre especies y delimitar linajes evolutivos independientes de manera objetiva en presencia de conflicto de árboles genéticos (Fujita et al., 2012; Smith et al., 2015; Luo et al., 2018; Smith et al., 2020).

En pinos existen pocos estudios que utilicen métodos formales para delimitar especies debido a las complejas historias evolutivas causadas por el sorteo incompleto de linajes y la reticulación (Rosenberg, 2003; Hernández-León et al., 2013; Zhang et al., 2014). Hasta ahora el uso de caracteres morfológicos, bioquímicos, anatómicos, de biología reproductiva, las distancias genéticas, los análisis filogenéticos y recientemente caracteres moleculares han contribuído en gran medida a la identificación y delimitación de especies del género *Pinus* (Price et al., 1998; Gernandt et al., 2001, 2003, 2005; Syring et al., 2005; Parks et al., 2012). No obstante, cada fuente de evidencia utilizada sufre procesos que pueden introducir errores como describimos anteriormente.

Entre los estudios para delimitar especies se encuentra el de Hernández-León et al. (2014), quienes utilizaron el modelo GMYC para delimitar especies de la sección *Trifoliae* a partir de genes de plasto. Los autores no utilizaron otro método de delimitación de especies para corroborar sus hipótesis de delimitación. Los resultados obtenidos mostraron que el método sólo logró estimar 23 especies de las 49 especies reconocidas en la sección. Este es el único trabajo en pinos que emplea métodos coalescentes para delimitar especies. Existen otros trabajos donde infieren los límites de especies, pero sin utilizar métodos coalescentes. Por citar algunos, se encuentra el de Moreno-Letelier et al. (2013), quienes utilizaron genes mitocondriales, de plasto y caracteres ambientales para delimitar un pequeño complejo de tres especies de pinos del subgénero *Strobus* que están cercanamente emparentadas y cuyos límites no eran claros. Los autores lograron

diferenciar genéticamente las tres especies, *P*. *ayacahuite*, *P*. *strobiformis* y *P*. *flexilis*. En 2015, López-Reyes et al. publicaron un estudio donde combinaron caracteres moleculares y morfo-anatómicos para inferir los límites entre *P*. *maximinoi* H. E. Moore y *P*. *douglasiana* Martínez empleando análisis multivariados. Los autores no lograron diferenciar ambos linajes y concluyeron que ambas especies pueden ser consideradas como incipientes debido a una incompleta divergencia morfológica, molecular y ecológica. De manera general existen estudios en especies de árboles que demuestran algunas de las estrategias que se han empleado para delimitar especies (*Abies*: Shao y Xiang, 2015; *Quercus*: Morales-Saldaña et al. 2021), pero el campo de la delimitación de especies en árboles sigue siendo poco explorado y aún más escasos son los estudios donde incluyen métodos estrictamente desarrollados para delimitar especies.

1.6. Edades absolutas de *Pinus* subsección *Cembroides*

En términos generales, los fósiles de *Pinus* han sido fundamentales para estimar las edades de los principales linajes taxonómicos del género (Willyard et al., 2007; Gernandt et al., 2008). El empleo del registro fósil combinado con filogenias moleculares (calibración) permite fechar los tiempos de divergencia no solo en *Pinus* sino en coníferas en general (Gernandt et al., 2016). En la actualidad, el uso de fósiles y secuencias de ADN para estimar los tiempos de divergencia en pinos es cada vez más popular (Wang et al., 2000; Willyard et al., 2007; Saladin et al., 2017; Gernandt et al., 2018b; Jin et al., 2021). Particularmente, el origen (edad troncal) y diversificación (edad corona) en pinos, utilizando fósiles para calibrar las filogenias, se han estimado o re-estimado a distintos rangos taxonómicos como a nivel de sección (Hernández-León et al., 2013) y a nivel de todo el género (Willyard et al., 2007; Saladin et al., 2017; Jin et al., 2021). Estos estudios previos han mostrado diferentes estrategias de calibración, utilizando diferentes conjuntos de datos y distintos métodos para estimar las edades en *Pinus*.

En la actualidad, existen varios métodos disponibles para estimar los tiempos de divergencia, entre los que se encuentran los “relojes moleculares” (Yang, 2007;

Drummond y Rambaut, 2007; Heath et al., 2014; Bouckaert et al., 2014). El término "reloj molecular" ahora se usa de manera más amplia para referirse a un conjunto de métodos y modelos que evalúan cómo varían las tasas de evolución genética a lo largo del árbol de la vida, y usan esta información para establecer una escala de tiempo absoluta en el árbol (Zuckerkandl y Pauling, 1965; Lee y Ho, 2016). Por tanto, los relojes moleculares son fundamentales para inferir escalas de tiempo evolutivas. Entre los enfoques más comunes se encuentran los relojes moleculares estrictos y los relojes moleculares relajados. Los relojes moleculares relajados pueden estimar los tiempos de divergencia, permitiendo diferentes tasas de sustitución genética (heterogeneidad) entre las ramas del árbol a partir de secuencias de ADN e información filogenética (Drummond et al., 2006), mientras que los relojes moleculares estrictos asumen tasas homogéneas entre las ramas de la filogenia. El modelo de reloj estricto tiene un solo parámetro que es la tasa de evolución (Zuckerkandl y Pauling, 1965; Ho y Duchêne, 2014). El reloj estricto también se utiliza como modelo nulo para probar la presencia de heterogeneidad de tasas evolutivas (Peterson, 2006). A pesar del avance en el desarrollo de los métodos de reloj molecular, el componente más importante en todos los análisis de datación molecular sigue siendo la elección de las calibraciones (Ho y Duchêne, 2014; Heath et al., 2008).

Pinus subsección *Cembroides* se clasifica en la sección *Parrya* junto con otras dos subsecciones de Norte América, *Balfourianae* y *Nelsoniae* (Gernandt et al., 2005). La sección *Parrya* se clasifica junto con la sección *Quinquefoliae* Duhamel en el subgénero *Strobus* (Gernandt et al., 2005). Dentro del subgénero *Strobus*, *Pinus* subsección *Cembroides* es la segunda subsección con más especies reconocidas (ver Price et al., 1998). A pesar del número de especies descritas*, Pinus* subsección *Cembroides* no cuenta con un registro fósil abundante en comparación con el subgénero *Pinus* (ver Millar, 1993). Esta limitante aumenta el riesgo de cometer errores cuando se utilizan fósiles para calibrar árboles filogenéticos (Heath et al., 2008). A pesar de esta restricción, los fósiles que se han utilizado para estimar la edad troncal de *Pinus* subsección *Cembroides* son *Pinus lindgrenii* Knowlton fechado para el Mioceno tardío y *P. sanjuanensis* Axelrod del

Oligoceno tardío (Saladin et al., 2017; Jin et al., 2021). La ubicación filogenética para *P. lindgrenii* y *P. sanjuanensis* se ha basado en sinapomorfías morfológicas entre las especies existentes de *Pinus* subsección *Cembroides*, pero las afinidades cercanas con las especies existentes son poco conocidas debido al registro fósil incompleto. Aún así, *Pinus sanjuanensis* ha sido reportado como una especie semejante a *P. edulis* (Axelrod, 1986), mientras que *P. lindgrenii* se ha identificado como un taxón con afinidades cercanas a *P. cembroides* y *P. edulis* por Axelrod (1986) y como afín de las especies actuales *P. monophylla* y *P. edulis* debido al rango de tamaño del cono y tamaño de la semilla (Miller, 1992). Una estrategia para identificar la posición de *P. sanjuanensis* y *P. lindgrenii* dentro de *Pinus* subsección *Cembroides* es a través de la combinación de una evaluación crítica con morfología y datos genéticos (evidencia total) tal como sugiere Willyard et al. (2007).

De acuerdo con el registro fósil, la edad troncal y la edad corona de *Pinus* subsección *Cembroides* son aún controvertidos. Axelrod (1986) sugiere que su origen se remonta al Eoceno medio según hace 47 millones de años (MA). De manera similar, se ha sugerido que la presencia de los pinos piñoneros de Norte América se remonta al Oligoceno hace 27.2 MA (Wolfe y Schorn, 1990). Esta diferencia de estimados sobre el origen de *Pinus* subsección *Cembroides* tiene que ver con la falta de aceptación por parte de Wolfe y Schorn sobre las edades de los fósiles más antiguos atribuidos a las subsecciones de *Pinus* por Axelrod (1986). Contrario al registro fósil, las filogenias moleculares calibradas con fósiles sugieren que la edad de origen ocurrió en el Oligoceno temprano (Gernandt et al., 2008), el Eoceno medio (Saladin et al., 2017) o durante el Eoceno temprano (Jin et al., 2021). Desde luego, las edades estimadas dependen de la aceptación o interpretación de como los fósiles pueden estar posicionados en una filogenia.

Se ha reportado que *Pinus* subsección *Cembroides* diversificó en América del Norte durante el Oligoceno (Jin et al., 2021) o el Mioceno (Willyard et al., 2007; Gernandt et al., 2008; Saladin et al., 2017). Aunque han radiado recientemente (Gernandt et al.,

2008) en esta subsección se encuentran especies como *P*. *maximartinezii*, *P*. *pinceana* y *P*. *rzedowskii* con una morfología divergente en conos, semillas, acículas y anatomía de la madera (Malusa, 1992; Farjon y Styles, 1997) y una variabilidad genética sustancialmente mayor (Delgado et al., 1999; Ledig et al., 2001), pero para el clado de pinos piñoneros de conos pequeños aún no se comprende bien la dinámica de diversificación porque las relaciones filogenéticas no se han resuelto o siguen siendo controvertidas (Saladin et al., 2017). Otra limitante es que las filogenias para relojes moleculares no han incluido a todas las especies existentes debido a que algunos taxones siguen siendo tratados como sinónimos, variedades, o subespecies (ver sección 1.2 de esta tesis). Los estudios de relojes moleculares demuestran algunas de las estrategias de calibración (asignación de fósiles en diferentes clados) para estimar los tiempos de divergencia en el género *Pinus* a distintas escalas taxonómicas (Willyard et al., 2007; Gernandt et al., 2008; Hernández-León et al., 2013; Saladin et al., 2017; Jin et al., 2021). Sin embargo, hasta el momento, las inferencias se han realizado por arriba del nivel de subsección, pero a nivel de subsección no existe un solo estudio en pinos que además incluya a todos los taxa reconocidos para un grupo. Se sabe que la diversidad no reconocida y las especies no muestreadas tienen un impacto negativo en las estimaciones del tiempo de divergencia (Cusimano y Renner, 2010), y la precisión de los tiempos de divergencia disminuye debido al muestreo incompleto (Heath et al., 2008). Por tal motivo, es necesario realizar un análisis con un muestreo taxonómico que incluya todas las especies reconocidas dentro de *Cembroides*, que permita aumentar la precisión en la inferencia de los tiempos de divergencia, con el empleo de múltiples genes de baja copia como una estrategia para aumentar el número de sitios informativos tal como sugieren Magallón y Sanderson (2005).

2. OBJETIVOS

2.1. Objetivo general

- Inferir las relaciones evolutivas, así como estudiar eventos de incongruencia filogenética e inferir los límites entre especies en *Pinus* subsección *Cembroides*, a partir de secuencias de genes nucleares de baja copia.

2.1.1. Objetivos específicos

- Inferir las relaciones evolutivas de las especies de *Pinus* subsección *Cembroides* mediante análisis de coalescencia.

- Explorar la importancia relativa del sorteo incompleto de linajes y evolución reticulada como causas de discordancia filogenética en *Pinus* subsección *Cembroides.*

- Estudiar los límites entre especies de *Pinus* subsección *Cembroides* a partir de métodos coalescentes multiespecie.

- Re-estimar los tiempos de divergencia dentro de *Pinus* subsección *Cembroides* utilizando genes nucleares de copia baja y todas las especies existentes.

- Estimar las tasas de diversificación en *Pinus* subsección *Cembroides*.

3. ARTÍCULOS DE INVESTIGACIÓN

3.1.

Phylogenetics of *Pinus* subsection *Cembroides* Engelm. (Pinaceae) inferred from low-copy nuclear gene sequences

Montes, J.-R., P. Peláez, A. Willyard, A. Moreno-Letelier, D. Piñero, and D. S. Gernandt. (2019)

Systematic Botany 44(3):501–518.
URL: https://doi.org/10.1600/036364419X15620113920563

Systematic Botany (2019), 44(3): pp. 501–518

DOI 10.1600/036364419X15620113920563
Date of publication August 6, 2019

Phylogenetics of *Pinus* Subsection *Cembroides* Engelm. (Pinaceae) Inferred from Low-Copy Nuclear Gene Sequences

José Rubén Montes,[1] Pablo Peláez,[1] Ann Willyard,[2] Alejandra Moreno-Letelier,[3] Daniel Piñero,[4] and David S. Gernandt[1,5]

[1]Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Apartado Postal 70-233, Ciudad de México, 04510, México; ruben.montes@st.ib.unam.mx; pablo.pelaez@cinvestav.mx
[2]Biology Department, Hendrix College, 1600 Washington Ave, Conway, Arkansas 72032, USA; willyard@hendrix.edu
[3]Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México, 04510, México; amletelier@ib.unam.mx
[4]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70-275, Ciudad de México, 04510, México; pinero@unam.mx
[5]Author for correspondence (dgernandt@ib.unam.mx)

Communicating Editor: Sasa Stefanovic

Abstract—*Pinus* subsection *Cembroides* comprises approximately 15 taxa distributed from the southwestern United States to south central Mexico. Despite previous phylogenetic studies based on morphology, nuclear ribosomal DNA, and plastid DNA, we still lack a robust phylogenetic hypothesis and clear delimitation for the closely-related species within the group. We studied the evolutionary relationships within subsection *Cembroides* and explored incomplete lineage sorting and reticulation using low-copy number nuclear genes. Concatenation and multispecies coalescent phylogenies were inferred from samples representing all taxa from subsection *Cembroides* and outgroups corresponding to the closely-related subsections *Balfourianae*, *Nelsoniae*, *Gerardianae*, and *Krempfianae*. The concatenation and coalescence-based trees mainly agreed with one another in recovering *Pinus* subsection *Cembroides* as monophyletic and in recovering similar relationships among species as in previous plastid DNA-based studies. Phylogenetic position and admixture analysis suggest that *P. californiarum* should be treated as a separate species from *P. monophylla*. Furthermore, our results support recognizing *P. fallax* as a species rather than as an infraspecific taxon of *P. monophylla* or *P. edulis*. The ASTRAL-III tree was consistent with the presence of very high levels of ILS in the group of pinyon pines with small cones. Analyses that account for both incomplete lineage sorting and reticulation identify some unexpected hybridization scenarios that were not reported in the literature.

Keywords—Coalescence, pinyon pine, reticulation, target enrichment.

Pinus subsection *Cembroides* Engelm. is a clade of North American pinyon pines occurring in arid or semi-arid environments from the southwestern United States to south central Mexico (Critchfield and Little 1966). The species of this subsection are small to medium-sized trees or shrubs characterized by secondary leaves with deciduous fascicle sheaths. Except for *P. rzedowskii* Madrigal, Caball.M., all taxa have enlarged seeds with a thickened sclerotesta (Madrigal and Caballero 1969) and are functionally wingless. The morphological characters that have been used in species identification in this group include the number of needles per fascicle, needle length and width, the distribution of stomata on the adaxial and abaxial leaf surfaces, and cone morphology (Malusa 1992; Farjon and Styles 1997). *Pinus* subsection *Cembroides* is classified in section *Parrya* Mayr together with two other North American subsections, *Balfourianae* Engelm. and *Nelsoniae* Burgh (Gernandt et al. 2005). *Pinus* subsection *Nelsoniae* is monotypic, with the pinyon pine, *P. nelsonii* Shaw distinguished by persistent fascicle sheaths and connate needles (Little and Critchfield 1969; Gernandt et al. 2001). Section *Parrya* is classified together with section *Quinquefoliae* Duhamel in subgenus *Strobus* Lemmon (Gernandt et al. 2005).

Seeds of *P. cembroides* subsp. *cembroides* Zucc. and *P. edulis* Engelm. are important sources of food in Mexico and the United States (Lanner 1981; Farjon and Styles 1997). The high nutritive value in seeds encourages interactions among pinyon pines, rodents, and corvid birds. The needles are sometimes used for medical treatments (Lanner 1981). The International Union for Conservation of Nature (IUCN 2017) lists five pinyon pine taxa as vulnerable or endangered and the Mexican government lists nine as protected (SEMARNAT 2010; Table 1).

Pinus subsection *Cembroides* has been the focus of several phylogenetic studies (Malusa 1992; Farjon and Styles 1997; Gernandt et al. 2001, 2003; Parks et al. 2012; Flores-Rentería et al. 2013) and the number of recognized species and infraspecific taxa differs in recent works (Malusa 1992; Eckenwalder 2009; Farjon and Filer 2013; Table 2). Phylogenetic results in the subsection have varied due to the use of different morphological, ecological, and molecular characters and differences in sampling (Malusa 1992; Farjon and Styles 1997; Gernandt et al. 2001, 2003, 2005; Syring et al. 2005). Based on a cladistic analysis of morphological and ecological characters, Malusa (1992) divided subsection *Cembroides* into a group of eight species with small seed cones and a second group of four species with large cones. In a restriction site study of noncoding plastid DNA, Pérez de la Rosa et al. (1995) recovered *P. nelsonii* as separate from subsection *Cembroides*. A cladistic analysis of morphological characters in Neotropical species of *Pinus* subgenus *Strobus* by Farjon and Styles (1997) recovered the small-coned pinyons as monophyletic, and the four large-cones species as paraphyletic to subsection *Strobus*. Two of the large-cone species, *P. nelsonii* and *P. pinceana* Gordon, formed a clade, leading the authors to classify them together in subsection *Nelsoniae*. In contrast, Price et al. (1998) classified *P. nelsonii* and *P. pinceana* in subsection *Cembroides* and *P. rzedowskii* in the monotypic subsection *Rzedowskianae* Carvajal.

More recently, phylogenetic relationships of pinyon pines have been inferred from sequences of plastid and nuclear ribosomal DNA (Gernandt et al. 2001, 2003, 2005; Parks et al. 2012; Ortiz-Medrano et al. 2016). Gernandt et al. (2001) reported phylogenetic analyses of the ITS region of nrDNA. The authors found that divergent copies of the ITS region in individuals of the same species do not group together. Nevertheless, results from the ITS region study and two subsequent phylogenetic studies using plastid DNA sequences corroborated the separation of *P. nelsonii* from the

TABLE 1. Taxa classified in *Pinus* subsection *Cembroides* with geographic distribution and conservation risk category. * Subject to special protection, and ** Listed as endangered by the Mexican government (SEMARNAT 2010). † Vulnerable and †† Endangered by the IUCN (2017).

Taxon	Distribution
Pinus californiarum D.K.Bailey	California (CA), Baja California (BC)
Pinus cembroides subsp. *cembroides* Zucc.	Arizona (AZ), New Mexico (NM), Texas (TX), Chihuahua (CH), Coahuila (CL), Durango (DG), Hidalgo (HG), Jalisco (JC), Nuevo León (NL), Querétaro (QO), San Luis Potosí (SP), Sonora (SR),
Pinus cembroides subsp. *orizabensis* D.K.Bailey††	Puebla (PL), Tlaxcala (TL), Veracruz (VZ)
Pinus culminicola Andresen & Beaman** ††	Coahuila (CL), Nuevo León (NL)
Pinus discolor D.K.Bailey & Hawksw	Arizona (AZ), New Mexico (NM), Durango (DG), San Luis Potosí (SP), Sonora (SR)
Pinus edulis Engelm.	Arizona (AZ), Colorado (CO), Nevada (NV), New Mexico (NM), Utah (UT), Texas (TX), Wyoming (WY)
Pinus fallax (Little) Businský	Arizona (AZ), Utah (UT), New Mexico (NM)
Pinus johannis M.F.Robert*	Coahuila (CO), San Luis Potosí (SP), Zacatecas (ZS)
Pinus lagunae (Passini) D.K.Bailey* †	Baja California Sur (BS)
Pinus maximartinezii Rzed.* ††	Durango (DG), Zacatecas (ZS)
Pinus monophylla Torr. & Frém.*	Arizona (AZ), California (CA), Idaho (ID), Nevada (NV), Oregon (OR), Utah (UT)
Pinus pinceana Gordon**	Coahuila (CL), Hidalgo (HG), Querétaro (QO), San Luis Potosí (SP), Zacatecas (ZS)
Pinus quadrifolia Parl. ex Sudw. *	California (CA), Baja California (BC)
Pinus remota (Little) D.K.Bailey & Hawksw. *	Texas (TX), Chihuahua (CH), Coahuila (CL), Nuevo León (NL)
*Pinus rzedowskii*** †	Michoacán (MICH)

other pinyon pines. Plastid DNA studies also have suggested that *P. johannis* M.F.Robert and *P. discolor* D.K.Bailey & Hawksw., are not infraspecific taxa of *P. cembroides* but instead close relatives of *P. culminicola* Andresen & Beaman (Gernandt et al. 2003, 2005; Parks et al. 2012; Ortiz-Medrano et al. 2016).

Some, but not all the results from phylogenetic studies have been followed in subsequent taxonomic treatments (e.g. Eckenwalder 2009; Debreczy and Rácz 2011; Farjon and Filer 2013). Farjon and Filer (2013) recognized that *P. nelsonii* and *P. pinceana* are not closely related, but that *P. nelsonii* belongs to a more distant group, and they recognized that *P. pinceana* is closely related to *P. maximartinezii* Rzed. Farjon and Filer (2013) treated *P. johannis* as a subspecies of *P. cembroides* although this has been contradicted by phylogenetic analysis of plastid DNA. Another plastid DNA study left in question whether the single-leaf pinyon pines are monophyletic, recovering *P. californiarum* D.K.Bailey as sister to *P. edulis* rather than to *P. monophylla* Torr. & Frém. (Gernandt et al. 2007). *Pinus californiarum* has been treated as a synonym of *P. monophylla* for sharing a single needle, or separated as *P. californiarum* or *P. monophylla* var. *californiarum* (D.K.Bailey) Silba based principally on differences in the number of resin canals, number of stomatal lines, and diameter of the needle (Silba 1990; Farjon and Styles 1997; Price et al. 1998).

TABLE 2. Two classifications of the species recognized in *Pinus* subsection *Cembroides*.

Farjon and Filer (2013)	Gernandt et al. (2005)
Pinus cembroides subsp. *cembroides*	
Pinus cembroides subsp. *cembroides* var. *bicolor*	*Pinus cembroides*
Pinus cembroides subsp. *lagunae*	
Pinus cembroides var. *orizabensis*	*Pinus culminicola*
Pinus culminicola	*Pinus discolor*
Pinus edulis	*Pinus edulis*
Pinus maximartinezii	
Pinus monophylla	*Pinus johannis*
Pinus pinceana	
Pinus quadrifolia	*Pinus maximartinezii*
Pinus remota	*Pinus monophylla*
Pinus rzedowskii	*Pinus pinceana*
	Pinus quadrifolia
	Pinus remota
	Pinus rzedowskii

Introgressive hybridization and gene flow have been reported in species of *Pinus* subsection *Cembroides* (e.g. Mirov 1967; Lanner 1974a, 1974b; Lanner and Phillips 1992; Malusa 1992). Some populations of *P. edulis* (predominantly two needles per fascicle) and *P. monophylla* (predominantly single-needled) occur in sympatry (Fig. 1) in the eastern Great Basin where trees of *P. edulis* with both single needles and two needles per fascicle have been observed (Lanner 1974a). In a study of natural hybridization in pinyon pines in northwestern Arizona, Lanner and Phillips (1992) analyzed variation in morphological characters over different years. Based on differences in the frequency of needle and resin canal numbers at 22 sites, they concluded that bidirectional introgression was occurring between *P. edulis* and *P. monophylla*, and proposed that overlapping phenology, wind-dispersed pollen, and weak premating barriers were responsible.

Lanner (1974b) proposed that *P. quadrifolia* Parl. ex Sudw. originated from hybridization between *P. monophylla* (treated here as *P. californiarium*; predominantly single-needled) and a previously undescribed species with five needles per fascicle, *P. juarezensis* Lanner (treated here as a synonym of *P. quadrifolia*). According to Lanner (1974b), this would explain extreme needle number variation in *P. quadrifolia*, a characteristic that is frequently observed in pine artificial hybrids when parents differ in this character (Keng and Little 1961). The geographical distributions of the taxa overlap broadly (Fig. 1), with sympatric populations common in Baja California (e.g. in the Sierra Juárez), suggesting that putative hybrids of several types co-exist in a hybrid swarm (Lanner 1974b). Farjon and Styles (1997) reported that pollen dispersal occurs in April and May in *P. monophylla* and in March and April in *P. quadrifolia*, which would allow for interspecific gene flow. Nonetheless, the proposal to recognize *P. juarezensis* as one of the parental species has not been widely accepted. *Pinus juarezensis* was considered a synonym of *P. quadrifolia* by subsequent authors (Farjon and Styles 1997; Gernandt et al. 2003; Eckenwalder 2009). Plastid DNA from three samples of *P. quadrifolia* (two of which were from the type locality of *P. juarezensis*) grouped together and formed a sister group to *P. monophylla* from California, indicating that *P. quadrifolia* has not captured pollen of *P. californiarum* (Gernandt et al. 2003).

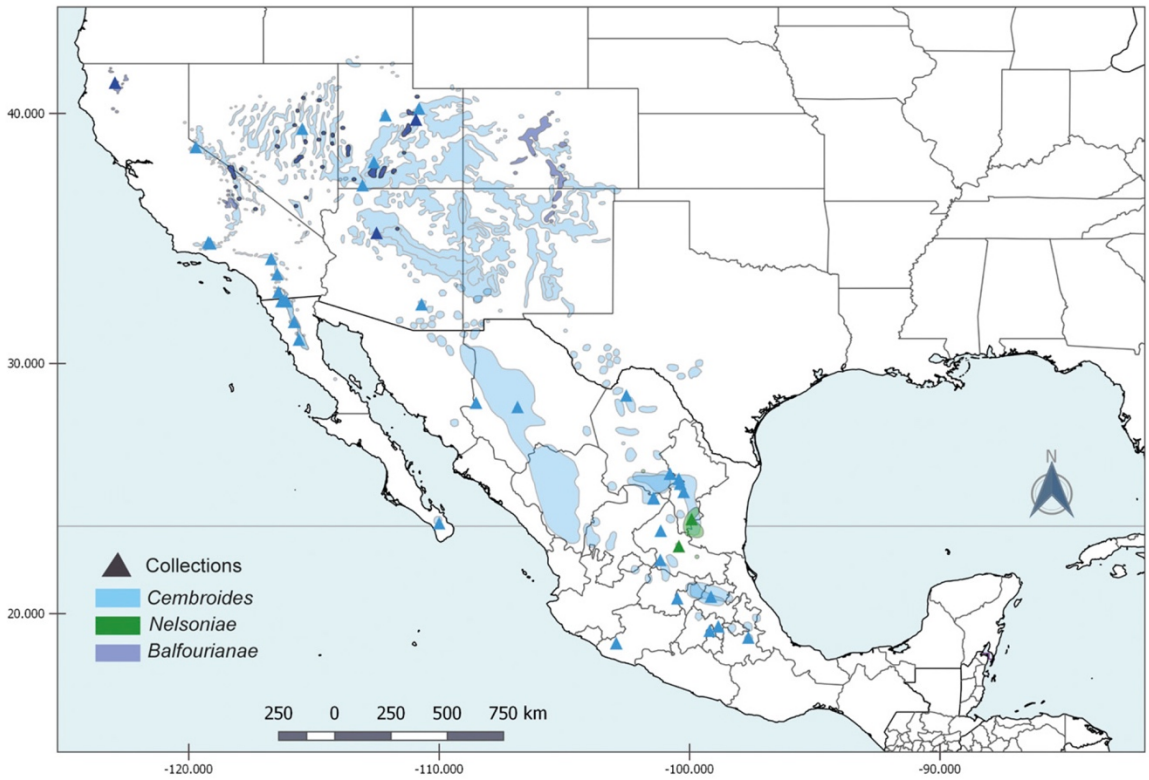


FIG. 1. Map showing the distribution of *Pinus* section *Parrya* based on Critchfield and Little (1966). Species records are based on herbarium collections and data from field studies. The individual sample sites are indicated by triangles (Appendix 1). Colors represent North American subsections that belong to section *Parrya*.

In his description of *P. remota*, Little (1968) suggested the possibility that in the past its populations had been in contact with those of *P. edulis*, permitting introgressive hybridization. The geographic distribution of *P. remota* and *P. edulis* may have been more extensive in the past (Late Quaternary), bringing the species into contact, maybe in the Chihuahuan Desert (Van Devender 1990). Additionally, there is a possible overlap in phenology between *P. remota*, which disperses its pollen between March and April, and *P. edulis*, which disperses pollen in a short period in the spring (Lanner 1970; Farjon and Styles 1997). In the cladistic analysis of morphological characters by Malusa (1992), *P. remota*, which has three needles per fascicle, formed a clade with the single-needle pinyon pines (*P. californiarum*, *P. fallax* (Businský) Little, and *P. monophylla*). The morphological character *P. remota* shares with the single-leaf pinyons is an elevated number of resin canals. An increase in the number of resin canals may have arisen in the species not by independent evolution but by hybridization, likely between *P. remota* and *P. fallax* (Malusa 1992). Plastid DNA of *P. remota* is unique, ruling out the possibility that it was acquired recently through hybridization (Gernandt et al. 2003).

Plastid DNA can provide important insights into phylogenetic relationships, but exclusive dependence on it as a phylogenetic marker can be misleading because plastid capture through introgression has been documented in diverse plant lineages (e.g. Delgado et al. 2007; Gernandt et al. 2018; Morales-Briones et al. 2018). Reticulation through hybridization and incomplete lineage sorting (ILS) also play important roles in gene tree discordance (Rieseberg and Soltis 1991; Maddison 1997) by producing very similar patterns of shared genetic diversity and obscuring phylogenetic relationships, which greatly limits our understanding of the processes of diversification in many lineages (Maddison and Knowles 2006). On one hand, weak premating barriers to gene flow can permit the introduction of alleles from other species (Mirov 1967). On the other, ILS can reduce the differentiation between species. Therefore, conifer species with long generation times (Petit and Hampe 2006) often share genetic variation (e.g. Wang et al. 2011; DeGiorgio et al. 2014). New DNA sequencing technologies and the development of coalescent-based frameworks allow us to better study hybridization and ILS. A promising approach is the use of low-copy nuclear genes to increase phylogenetic resolution and explore the problems of discordance between gene trees and species trees. For example, Syring et al. (2005) found that phylogenetic inference based on low-copy nuclear genes in *Pinus* subgenus *Strobus* increased resolution and robustness in most of the subsectional clades recovered in their study. Gernandt et al. (2018) demonstrated the utility of low-copy nuclear genes to explore ILS and reticulation at the species level in *Pinus* subsection *Australes*. For this reason, our aims were to infer the phylogenetic relationships for species of *Pinus* subsection *Cembroides* and explore the relative importance of incomplete lineage sorting and reticulation as causes of phylogenetic

discordance. We used targeted sequence capture (Gnirke et al. 2009), also known as Hyb-Seq (Weitemier et al. 2014), to characterize nuclear DNA sequences from multiple individuals per species and perform concatenated and multispecies coalescent analyses. This study represents the most complete taxonomic sampling to date for *Pinus* subsection *Cembroides* and its close relatives.

Materials and Methods

Taxon Sampling—We sampled 60 individuals, with most taxa in subsections *Cembroides* and *Nelsoniae* represented by multiple populations (Appendix 1). Vouchers were deposited in the Herbario Nacional de México (MEXU), Instituto de Biología, Universidad Nacional Autónoma of México and the Oregon State University Herbarium (OSC). The individuals represent all taxa of *Pinus* subsection *Cembroides* recognized by Gernandt et al. (2005). For outgroups we included three individuals representing two of three species of subsection *Balfourianae* and three of *P. nelsonii*, the only member of subsection *Nelsoniae* (Fig. 1). These two subsections were recovered as the sister group of subsection *Cembroides* in previous phylogenetic analyses and together with subsection *Cembroides* are classified in section *Parrya* (Gernandt et al. 2005). We also included representative species from section *Quinquefoliae*: subsections *Gerardianae* Loudon (2), *Krempfianae* Little and Critchfield (1), and *Strobus* as more distant outgroups (Appendix 1).

DNA Extraction and Quantification—For extraction of genomic DNA, we followed the modified CTAB method of Doyle and Doyle (1987) for diploid leaf tissue and used a Wizard genomic DNA purification kit (Promega, Madison, Wisconsin) for haploid seed megagametophyte tissue. We used a Nanodrop spectrophotometer 2000/2000c (Thermo Scientific, Waltham, Massachusetts) to measure the absorbance maxima ratio. Samples with 800 ng or more of DNA and an A260/A280 between 2.0 and 2.2 were selected for sequencing. We used a Qubit fluorometer v. 3.0 and dsDNA HS assay kit (Life Technologies, Carlsbad, California) to measure DNA concentration.

Probe Design—Details on probe design, library preparation, sequencing, and gene assembly were described by Gernandt et al. (2018). Briefly, a total of 1045 putative single copy nuclear genes were screened for probe design from *Pinus* species (*P. taeda* L., *P. pinaster* Aiton, and *P. sylvestris* L.; Neves et al. 2013; Willyard et al. 2007). The exon sequences were submitted to Arbor Biosciences (Ann Arbor, Michigan, USA), and 120 bp RNA bait sequences were used to perform a BLAST search on the *P. taeda* draft genome v. 1.0 (Neale et al. 2014; Wegrzyn et al. 2014).

Illumina Library Preparation and Target Enrichment—Genomic libraries were prepared with between 100 and 500 ng of DNA per sample. The DNA was fragmented into ca. 250 bp with a bioruptor and barcode adapters were ligated for sequencing on the Illumina using a TruSeq library prep kit (Illumina, San Diego, California). Libraries were pooled into 24 samples in equimolar ratios and enrichment was carried out with MYbaits biotinylated RNA baits following the manufacturer's protocol v. 2.3.1 (Arbor Biosciences, Ann Arbor, Michigan). The samples were combined in equal concentrations (48×) and sequenced using an Illumina Hi-Seq 2500 with the 100 bp module with paired reads.

Data Selection—We used two different data sets as input for phylogenetic analyses. The principal data set consisted of genes assembled with HybPiper v. 1.2 (Johnson et al. 2016). Data are available in the Dryad Digital Repository (Montes et al. 2019). A second data set consisted of single nucleotide polymorphisms (SNPs) identified with SAMtools (see below). For both, Illumina reads were filtered in Trimmomatic v. 0.36 (Bolger et al. 2014). For nuclear genes, a total of 60 paired R1 and R2 files in *fastq* format were filtered in Trimmomatic, removing bases at read ends with qualities $<$ Q20 using a 4 bp sliding window, and removing reads with a length $<$ 30 bp following trimming (Weitemier et al. 2014). Only reads with both pairs surviving were assembled into individual alignments with HybPiper (Johnson et al. 2016). HybPiper used BWA v. 0.7.1 (Li and Durbin 2009) to align reads to the reference nuclear gene sequences. SAMtools v. 0.1.19 (Li et al. 2009) was used to sort the reads into separate directories for each gene. The pipeline subsequently used SPAdes v. 3.10.1 (Bankevich et al. 2012) for de novo assembly of each gene individually using the retrieved reads. A total of 969 genes from 996 targets were assembled successfully for at least one sample.

The resulting gene assemblies were imported into Geneious v. R9 (Kearse et al. 2012). Individual nuclear gene alignments were performed with MAFFT v. 7.0 (Katoh et al. 2002). We used the following criteria proposed by Gernandt et al. (2018) to filter the multiple sequence alignments: 1) missing one or more samples, 2) fewer than 50% of sites, 3) pairwise similarity less than 93%, and 4) putative paralogs. The first three criteria were applied in Geneious, whereas the paralogs were detected when assembling the only two haploid references, *P. cembroides* USA and *P. bungeana* CA, with HybPiper (Johnson et al. 2016). Paralogs script identified contigs with lengths $\geq$ 85% of the reference sequence, indicating multiple long-length matches. After filters, we excluded 665 of the 969 genes assembled. The remaining 304 genes were carried forward for phylogenetic analysis.

Phylogenetic Analysis Using Low-Copy Number Nuclear Genes—We analyzed a concatenated alignment of 304 nuclear genes for 60 individuals with maximum likelihood in RAxML-HPC v. 8.2.10 (Stamatakis 2014). We performed 1000 heuristic searches for the best tree applying a general time reversible model with the gamma parameter (GTR + G) separately to each gene. Bootstrapping was performed with 1000 replicates as part of the heuristic search with RAxML. The best tree was imported into FigTree v. 1.4.0 for further editing (Rambaut 2012). Samples with unstable topological positions ("rogues") were identified using the maximum dropset size in RogueNaRok (Aberer et al. 2013), available as an online webserver (http://rnr.h-its.org/). These individuals were excluded from the concatenated alignment and multispecies coalescent analyses.

Phylogenetic Analysis Using Coalescent-Based Methods—Coalescent analyses that accommodate ILS and hybridization were performed on the low-copy nuclear genes. Because estimating reticulation is computationally demanding, we used a nuclear gene tree as a guide for choosing a subsample of individuals representative of *Pinus* subsection *Cembroides*. The species tree was estimated with minimizing deep coalescences (MDC), a parsimony method (Maddison 1997; Than and Nakhleh 2009) and networks were estimated by maximum parsimony from gene trees. To study reticulation scenarios, we used PhyloNet v. 3.6.1 (Than et al. 2008) to analyze 22 sequences representing 22 taxa, of which 15 belong to subsection *Cembroides*. The remaining seven taxa were from the outgroup (one individual per species). We used the Perl script BeforePhylo.pl v. 0.9.0 (https://github.com/qiyunzhu/BeforePhylo.git) to produce individual gene alignment files with the reduced representation of samples. The maximum likelihood tree for each of the 304 genes was inferred with RAxML using the GTR + G model and bootstrapping with the autoMRE option. We sequentially permitted up to three reticulation events under the MDC criterion (InferNetwork_MP). For each reticulation setting, we performed 10 independent searches of 5 to 20 (-x 5 to -x 20). The network with the lowest number of extra lineages was selected and displayed graphically with Dendroscope v. 3.0 (Huson and Scornavacca 2012). Based on the results from the 22-terminal analysis, we evaluated the effect of constraining *P. remota* and *P. quadrifolia* as hybrids. We included the taxa proposed to have been involved in past reticulation events with these species. For *P. remota*, these include *P. cembroides*, *P. edulis*, and *P. fallax* (Malusa 1992; Little 1966). For *P. quadrifolia* we included *P. californiarum* (Lanner 1974b).

Species and lineage tree inference was also performed with SVDquartets (Chifman and Kubatko 2014) in PAUP* v. 4.0a150 (Swofford 2002). SVDquartets is a robust method for multilocus data that is designed to build unrooted quartets that accommodate ILS but not reticulation and evaluates the correspondence of the nodes (see Chifman and Kubatko 2014). For this method we used a NEXUS input file that included a data set block specifying the 304 gene partitions and a taxon set assigning each individual to its corresponding species. The file included 58 sequences representing the 15 subsection *Cembroides* taxa and ten individuals from the outgroup, the same as in the concatenated analysis. A maximum of 1,000,000 randomly chosen quartets were analyzed and branch support was estimated with 1000 bootstrap replicates (bs). The trees were displayed graphically in FigTree v. 1.4.0 (Rambaut 2012).

In addition, we concatenated these 304 low-copy number nuclear genes with SNPs (see below) for 51 samples to infer a species tree with SVDquartets. The species trees resulting from the nuclear gene alignment and the concatenated SNPs analyses were displayed graphically as a tanglegram with Dendroscope (Huson and Scornavacca 2012).

The species tree was also estimated with ASTRAL-III v. 5.6.3 (Mirarab and Warnow 2015; Zhang et al. 2017). We used 58 individuals representing 22 taxa, removing two samples, one because it had an unstable position in the RAxML analysis of the concatenated alignment and the other because of the amount of shared ancestry between two species observed with SNPs. Individual trees from the 304 nuclear genes were estimated with RAxML using the GTR+G model and bootstrapping with the autoMRE option. The best tree from each analysis was concatenated and used as the input file for ASTRAL-III. Branch lengths in coalescent units and local posterior

probabilities were estimated for the species tree (Sayyari and Mirarab 2016). We performed character state reconstruction by mapping of the number of leaves per fascicle as an ordered multistate character on the species tree inferred with ASTRAL-III to evaluate the origin of single-needle pinyons based on the likelihood ancestral state approach in Mesquite v. 3.6 (Maddison and Maddison 2018).

Sequence Read Alignment and SNP Calling—Single nucleotide polymorphisms constitute a valuable source of genetic variation to study evolutionary relationships in non-model organisms (Leaché and Oaks 2017). Although only 304 assembled genes were selected to carry out phylogenetic analyses, we observed that for other target genes some regions were assembled and could be used to identify genetic markers in spite of our gene filtering criteria. For this reason, we performed reference-guided SNP calling in an effort to increase the number of sites for phylogenetic analysis. This method provided an alternative way of interpreting the Illumina data compared to assembling with HybPiper and applying our ad hoc filters. It has the potential to identify more informative sites and reduce bias introduced by HybPiper, which only returns one allele per gene. Also, it might be more reliable for removing paralogs or improving homology across samples. Demultiplexed sequence reads from each of the 52 samples were evaluated using FastQC and MultiQC and filtered with Trimmomatic v. 0.36 to discard low quality and adapter sequences (Bolger et al. 2014; Ewels et al. 2016). The minimum length of reads to be kept was set to 50 bp. Paired cleaned sequences were mapped against the *Pinus taeda* genome v. 2.0 (Neale et al. 2014; Wegrzyn et al. 2014) using BWA-MEM with default parameters. The SAM files were converted to BAM files with SAMtools v. 0.1.19 (Li et al. 2009). Uniquely mapped and sorted reads were obtained for each alignment file using the view and sort routines from SAMtools. Potential PCR duplicates were discarded using the Picard tool MarkDuplicates (http://broadinstitute.github.io/picard). SAMtools mpileup and BCFtools (with the options for biallelic variants SNPs/indels, no-BAQ, minimum mapping quality of 20, and minimum base quality of 25) were used for variant calling. SNPs were further filtered with VCFtools v. 0.1.13 (Danecek et al. 2011) for phylogenetic and admixture analyses.

SNP-Based Phylogenetic and Admixture Analyses—The SNPs were filtered if genotypes were not called across all samples (100%), minimum mean depth was below 5, minimum quality score was below 30, or minor allele frequency was lower than 0.05. The variant calling format (VCF) file was converted into a tab-delimited text file with VCFtools and subsequently SNPs were concatenated into a FASTA file including heterozygous sites. A multiple sequence alignment was generated with MAFFT v. 7.0 (Katoh et al. 2002) and used as input to infer a maximum likelihood tree in RAxML-HPC v. 8.0.26 (Stamatakis 2014). The maximum likelihood and bootstrap searches were carried out with the GTR+G model and a total of 1000 bootstrap replicates were computed.

For the admixture analysis, SNPs were filtered if genotypes called were below 80% across all samples, minimum mean depth was below 5, and minimum quality score was below 30. The VCF file was converted to an ordinary PLINK file (.ped) using PLINK v. 1.9 (Purcell et al. 2007). Maximum likelihood estimation of individual ancestries (population structure analysis) was carried out using ADMIXTURE v. 1.3.0 (Alexander et al. 2009). ADMIXTURE's cross-validation procedure (5-fold) was run for ancestral populations values (K) from 2–5. The Q-matrices generated for the different K values (1–4) were then clustered and plotted using CLUMPAK v. 1.1 (Kopelman et al. 2015).

Results

Filtering and Editing Sequences—Statistics for the 60 samples assembled in HybPiper are summarized in Appendix 2. Fifty samples corresponded to subsection *Cembroides* and 10 to the outgroups. Mean coverage of the 969 genes was 226.42×. We excluded 68.6% of genes as a result of the data filtering steps (Table 3), resulting in a final tally of 304 included gene alignments.

Phylogenetic Analyses Using Nuclear Genes—The alignment of 304 concatenated nuclear genes was 222,129 bp in length and included 16,503 parsimony informative sites and 13,639 variable but parsimony uninformative sites. The RAxML analysis recovered *Pinus* subsection *Cembroides* as monophyletic (100% bs). Rooting with section *Quinquefoliae* resulted in recovering subsections *Nelsoniae* and *Balfourianae* sister to one another and in turn sister to subsection *Cembroides* (Fig. 2). This section *Parrya* clade of exclusively North American taxa was recovered with high bootstrap support (100% bs). In the outgroup, the subsections *Krempfianae* and *Gerardianae* were united (100% bs). In the subsection *Cembroides* clade, eight of 13 taxa represented by multiple individuals were recovered as exclusive lineages (*P. culminicola*, *P. johannis*, *P. fallax*, *P. maximartinezii*, *P. monophylla*, *P. pinceana*, *P. remota*, and *P. rzedowskii*; Fig. 2). The bootstrap values for most branches in the phylogeny were high. The three large-cone pinyon pines with a restricted geographical distribution in Mexico formed a well-supported clade, in which *P. maximartinezii* and *P. pinceana* formed a monophyletic group (100% bs) sister to *P. rzedowskii* (100% bs). The large-cone clade was sister to a clade of the remaining (small-cone) species of subsection *Cembroides*. In the small-cone clade, *P. johannis* and *P. discolor* were sister to *P. culminicola* (94% bs). Five individuals of *P. remota* were sister to the clade of *P. johannis*, *P. discolor*, and *P. culminicola* (92% bs). *Pinus cembroides* subsp. *cembroides*, *P. cembroides* subsp. *orizabensis*, and *P. lagunae* were paraphyletic to the clade of *P. discolor*, *P. johannis*, *P. culminicola*, and *P. remota*, with individuals of *P. cembroides* subsp. *cembroides* from Chihuahua and San Luis Potosí forming a clade with two individuals of *P. lagunae* (Baja California Sur), and *P. cembroides* subsp. *orizabensis* (Puebla) sister to an individual of *P. cembroides* subsp. *cembroides* from Hidalgo. The two individuals of *P. fallax* formed a group with one sample of *P. edulis* (93% bs), all from Utah (USA). *Pinus monophylla*, *P. californiarum*, and *P. quadrifolia* were recovered as a clade with high bootstrap support (98% bs). *Pinus monophylla* was recovered as monophyletic and sister to *P. quadrifolia*, and both as sister to a single sample of *P. californiarum* from Baja California.

Only one sample, *Pinus monophylla* UT1 (DSG478), was identified as having an unstable position with RogueNaRok (dropset size 3.0). We did not include it in the subsequent coalescence analyses with MDC, SVDquartets, and ASTRAL-III, and it was removed from the concatenated analysis in RAxML. Also, one sample of *P. californiarum* (CA2; DSG403) was removed from the coalescence and concatenated nuclear genes analyses to avoid the probability of bias in the evolutiionary relationships due to the proportions of shared ancestry observed for this sample with SNPs (see Fig. 3A).

SNP-Based Phylogenetic and Admixture Analyses—Mapping the Illumina reads to the genome of *P. taeda* identified 26,499

Table 3. Results of the data after filtering. [a]One or more samples were not assembled to the reference sequence (996 genes). [b]Fewer than 50% of sites identical. [c]Pairwise identity less than 93%. [d]The HybPiper script identified contigs with lengths ≥ 85% of the reference sequence, indicating multiple long-length matches (see Johnson et al. 2016).

	Missing data[a]	Percent of alignment columns identical[b]	Pairwise identities[c]	Paralogs[d]	Eliminated by visual inspection	Total
Genes excluded	160	299	30	103	73	665
Percent	16.5%	30.9%	3.1%	10.6%	7.5%	68.6%

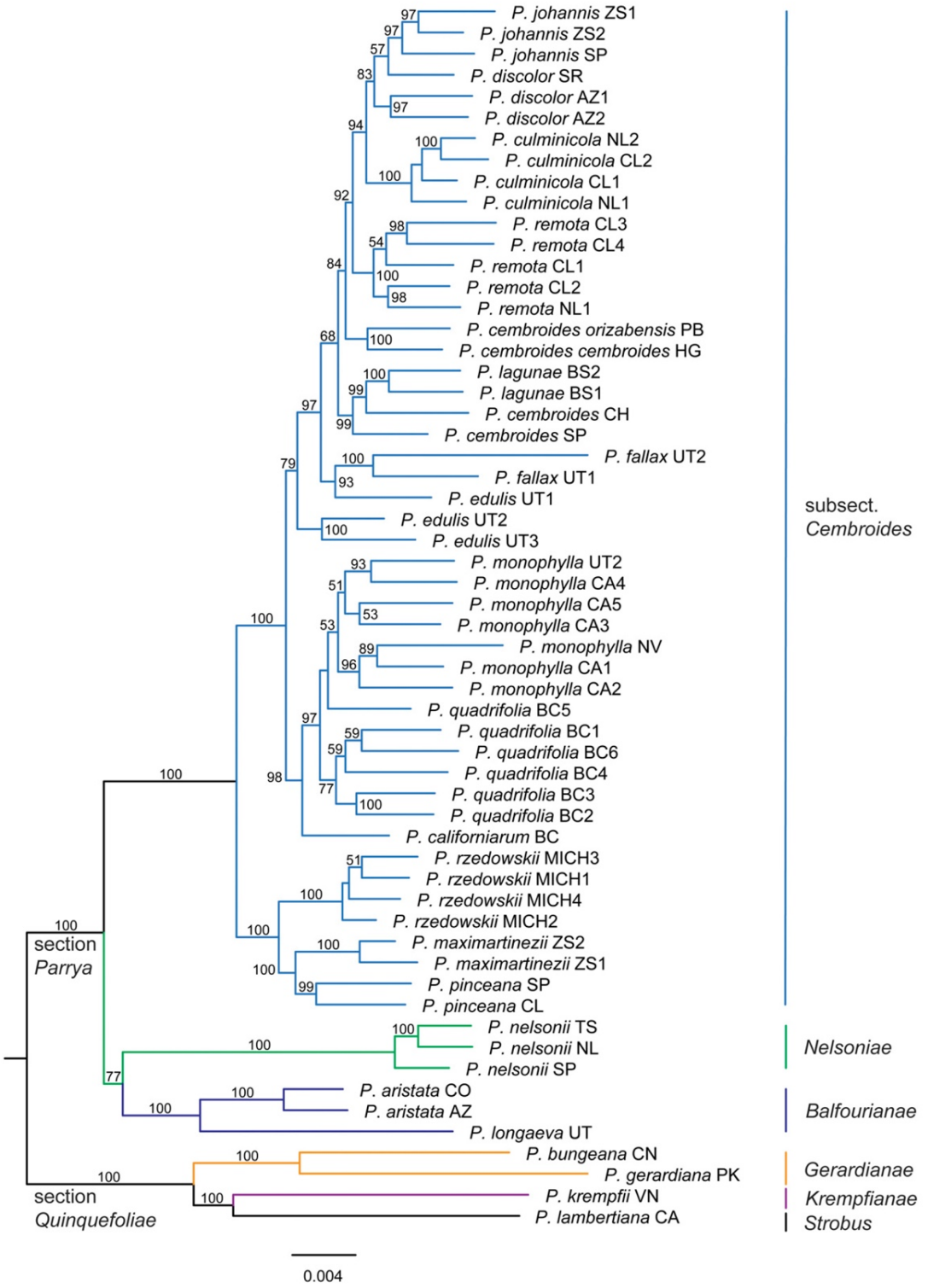


FIG. 2. Phylogeny of *Pinus* subsection *Cembroides*. Maximum likelihood tree inferred from the concatenated alignment (222,129 bp; 58 terminals and 304 genes). Bootstrap values > 50% are shown above the branches. The taxonomic subsections are represented by colors in some trees in this study (blue = *Cembroides*; green = *Nelsoniae*; navy blue = *Balfourianae*; orange = *Gerardianae*; violet = *Krempfianae*; black = *Strobus*). The sample names indicate the taxon and the state of collection. Locality codes are provided in Table 1.

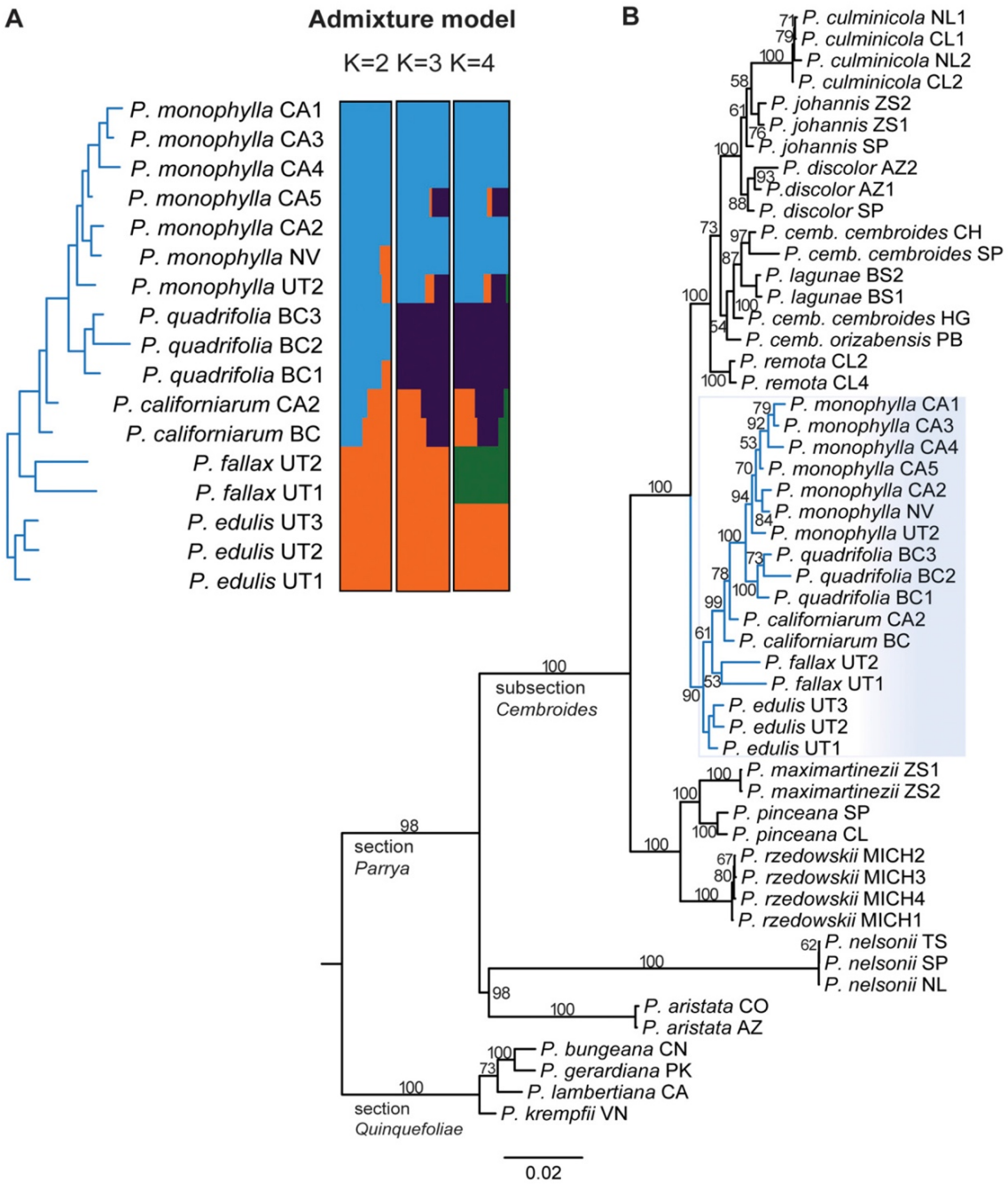


FIG. 3. Admixture analysis of a pinyon pine clade. A. Analyses were performed for *K* values ranging from 2 to 4 with a matrix containing 26,499 single nucleotide polymorphisms. Different colors represent different clusters. The combination of different colors in a bar indicates the degree of admixture. Samples in the admixture model are in the same topological order as in the maximum likelihood phylogeny which is shown at right for comparison. B. Maximum likelihood tree inferred from single nucleotide polymorphisms. Bootstrap values > 50% are shown above the branches.

SNPs genotyped in all 52 pines, including 7262 parsimony informative sites and 735 variable but parsimony uninformative sites. The resulting maximum likelihood tree inferred with RAxML (Stamatakis 2014; Fig. 3B) had a similar backbone topology as that inferred from the concatenated nuclear genes (Fig. 2), with *Pinus* subsection *Cembroides* monophyletic (Fig. 3B) and divided into two main clades comprising the large-cone pinyon pines (100% bs) and the small-cone pinyon pines (100%

bs). In the small-cone clade, *P. californiarum*, *P. monophylla*, *P. quadrifolia*, *P. fallax*, and *P. edulis* formed a well-supported monophyletic group (90% bs). *Pinus culminicola*, *P. johannis*, *P. discolor*, *P. lagunae*, *P. cembroides* subsp. *cembroides*, *P. cembroides* subsp. *orizabensis*, and *P. remota* formed another well-supported clade (100% bs). In contrast to the analysis of the HybPiper exon assembly (Fig. 2), the maximum likelihood tree inferred from SNPs grouped the two subspecies of *P. cembroides* and *P. lagunae* together in an exclusive lineage (Fig. 3B). *Pinus californiarum* individuals were sister to *P. monophylla* + *P. quadrifolia*, but the two samples of *P. californiarum* were paraphyletic in the SNP-based phylogeny, suggesting significant genetic variation within this species (Fig. 3B)

The interesting phylogenetic relationships for the clade comprising *P. californiarum*, *P. monophylla*, *P. quadrifolia*, *P. fallax*, and *P. edulis*, together with previous evidence of hybridization and introgression between some of these species, led us to explore the genetic structure of the individuals from this clade using ADMIXTURE. In total, 67,938 SNPs corresponding to only these five species were used. The cross-validation errors for ancestral structured populations values increased as the number of populations increased. Therefore, although the lowest cross-validation error was for K = 2, we used a cluster range of 2–4 to explore the dynamics of classifications in the different number of populations. Interestingly, for the lowest K all the samples were mainly defined by one subpopulation with the exception of the sample corresponding to *P. californiarum* CA2, which showed the strongest signs of admixture (Fig. 3A). For K = 3, samples from *P. quadrifolia* defined a third substructure and admixture patterns for sample CA2 was maintained. *Pinus edulis* and *P. fallax* were grouped differentially only at the highest K value. *Pinus californiarum* CA2 and *P. californiarum* BC1 consistently showed evidence of admixture across all three values of K. *Pinus monophylla* individuals were mainly assigned to the same population regardless of which K value was used (Fig. 3A). Also, *P. edulis* and *P. fallax* showed consistency regarding their classification in the same population for the first two K values.

Phylogenetic Results for Coalescent-Based Methods—HybPiper exon data were used to infer a species tree under the MDC criterion with Phylonet on a subset of subsection *Cembroides* individuals (22). *Pinus* subsection *Cembroides* was recovered as monophyletic (8084 lineages; Fig. 4). The analyses sequentially permitting up to three reticulation events (Fig. 4B–D) always identified gene flow within subsection *Cembroides*; none was identified among the outgroups. Allowing for a single reticulation resulted in a reduction from 8084 to 7619 lineages and involved introgression from *P. monophylla* into *P. edulis* (Fig. 4B). Allowing for two reticulations resulted in a reduction to 7453 lineages (Fig. 4C). The first reticulation involved introgression from *P. fallax* into *P. cembroides* subsp. *cembroides* (Fig. 4C: H1) and the second involved introgression from *P. edulis* into *P. lagunae* (Fig. 4C: H2). Allowing for three reticulations resulted in a reduction to 7211 lineages (Fig. 4D). The first reticulation (Fig. 4D: H1) involved introgression from *P. culminicola* into *P. quadrifolia*, the second involved introgression from a possible extinct taxon into *P. lagunae* (Fig. 4D: H2), and the third involved introgression from a possible extinct taxon into *P. cembroides* subsp. *cembroides* (Fig. 4D: H3).

Specifying *P. remota* as a hybrid under MDC resulted in a reduction from 8084 lineages inferred in the species tree (Fig. 4A) to 7683 in the network (Fig. 5A). *Pinus remota* was sister to *P. johannis* and gene flow was inferred from *P. fallax* (28% inheritance probability). Specifying *P. quadrifolia* as a hybrid resulted in a reduction to 7627 lineages (Fig. 5B). In this network *P. quadrifolia* was sister to *P. monophylla* (54% inheritance probability) and gene flow was inferred from *P. lagunae* (46% inheritance probability) (Fig. 5B).

Pinus subsection *Cembroides* was monophyletic and relationships among the outgroups were well supported in the SVDquartets and ASTRAL-III trees (Figs. 6B–8). The Mexican pinyon pine *P. nelsonii* (subsection *Nelsoniae*) and *P. aristata* and *P. longaeva* were sister to subsection *Cembroides* in the SVDquartets lineage and species trees (Figs. 6–7) but was sister to subsections *Cembroides* + *Balfourianae* in the ASTRAL-III tree (Fig. 8). *Pinus maximartinezii* and *P. pinceana* were sister and in turn sister to *P. rzedowskii* in all trees (Figs. 2–8). In the lineage tree individuals of *P. monophylla* formed a group with *P. quadrifolia* and *P. californiarum* but individuals of the same species were not recovered as exclusive lineages. Individuals of *P. edulis* also were not recovered as exclusive lineages (Fig. 6B). *Pinus lagunae* and the taxonomic varieties of *P. cembroides* were not recovered as exclusive lineages (Fig. 6), whereas *P. remota*, *P. culminicola*, and *P. johannis* were each recovered as exclusive lineages (> 95% bs).

In both the SVDquartets and ASTRAL-III analyses, *P. remota* was sister to the *P. culminicola*, *P. johannis*, and *P. discolor* clade (Figs. 7–8). The relationships among *P. discolor*, *P. culminicola*, and *P. johannis* differed between ASTRAL-III and SVDquartets (Figs. 6–8). Whereas *P. culminicola* was sister to *P. johannis* and *P. discolor* in the SVDquartets species tree, in the ASTRAL-III tree *P. discolor* was sister to *P. culminicola* and *P. johannis* (Fig. 8). The position of *P. fallax* and *P. edulis* also differed in the two coalescence analyses. Moreover, the position of *P. edulis* as sister to both varieties of *P. cembroides*, *P. culminicola*, *P. discolor*, *P. johannis*, *P. lagunae*, and *P. remota* was not well supported in the SVDquartets species tree (65% bs) indicating a weak node (Fig. 7). Branch lengths in the tree inferred with ASTRAL-III were short within subsection *Cembroides* except for the branch subtending the clade with *P. maximartinezii*, *P. pinceana*, and *P. rzedowskii*.

The likelihood results for character state reconstruction on the ASTRAL-III species tree supported a common origin of reduction to single needles in *P. californiarum, P. fallax*, and *P. monophylla* followed by a single increase in the number of needles per fascicle in *P. quadrifolia* (-log 33.50).

Comparison Between SVDquartets Analyses from SNPs and Nuclear Genes—The concatenated nuclear gene alignment was 222,129 bp in length (16,503 parsimony-informative sites) compared to 26,499 characters for the SNP data set (7262 parsimony-informative sites). The analyses performed in SVDquartets recovered species trees that were topologically similar. Trees based on both SNPs and concatenated nuclear genes (Fig. 9) recovered subsection *Cembroides* as monophyletic with high branch support and the large- and small-cone clades were sister groups in both analyses. The relationships within subsection *Cembroides* differed for several small-cone species. The bootstrap values for the analyses based on concatenated genes and SNPs were similar, but in some positions the support was higher using concatenated genes (Fig. 9B). In the concatenated gene analysis, *P. remota* was sister to the *P. culminicola* + *P. discolor* + *P. johannis* clade (96% bs), whereas with SNPs, *P. cembroides* subsp. *orizabensis* was sister to the *P. culminicola* + *P. discolor* + *P. johannis* clade

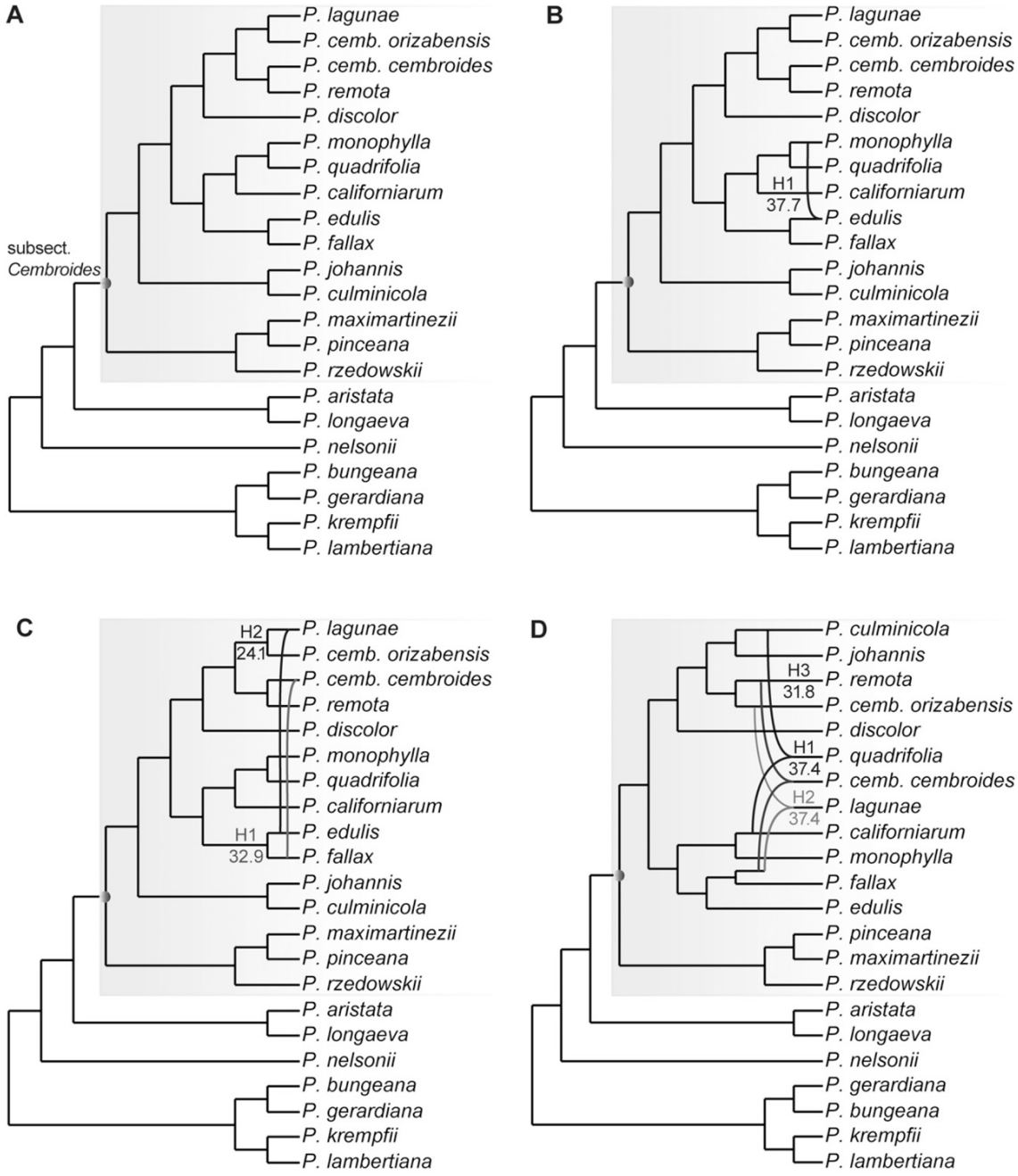


FIG. 4. Results of minimizing deep coalescences (MDC) analyses with and without reticulation. Inheritance probabilities for the minor edge are indicated for the networks (B, C, and D), with the reticulations represented by lines. A. Best MDC tree with no reticulation. B. Best MDC network permitting one reticulation event. C. Best MDC network permitting two reticulation events. D. Best MDC network permitting three reticulation events.

(61% bs) with SNPs. The clade of *P. cembroides* subsp. *cembroides*, *P. cembroides* subsp. *orizabensis*, and *P. lagunae* was recovered as monophyletic with the concatenated gene alignment (72% bs), but with SNPs it was paraphyletic due to the placement of *P. cembroides* subsp. *orizabensis* as sister to the clade of *P. discolor*, *P. johannis*, *P. culminicola*, and *P. remota* (61% bs). In the SNPs tree, *P. fallax* was sister to the single-needle pinyons together with *P. quadrifolia* (63% bs) whereas in the concatenated gene tree it was sister to all other small-cone species (100% bs). In both analyses of SVDquartets,

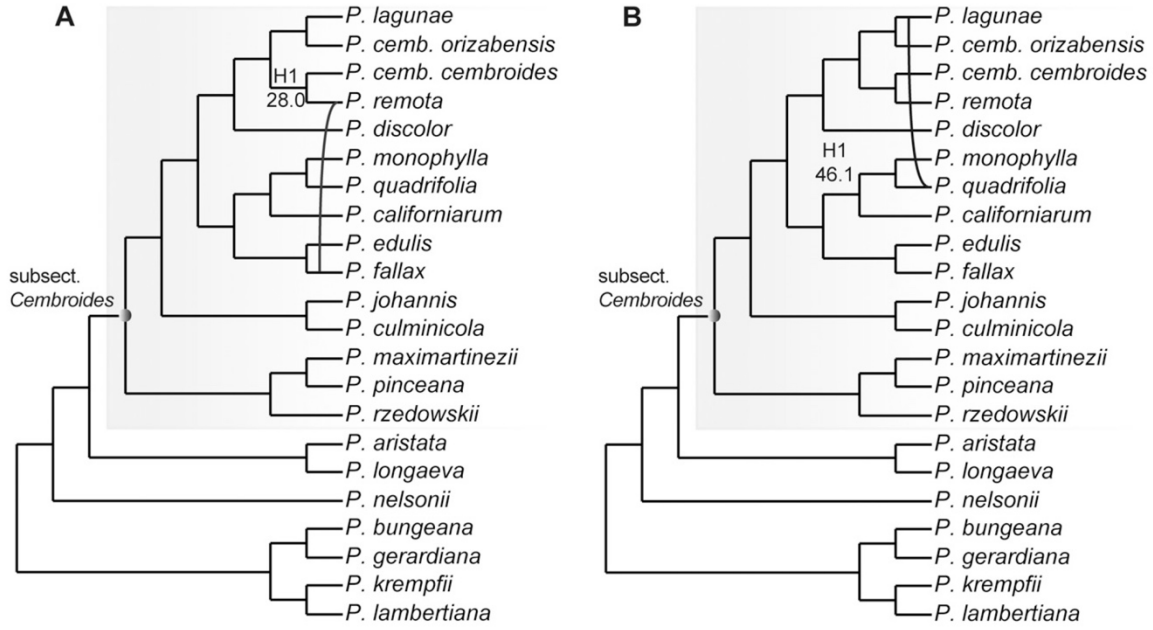


FIG. 5. Tests constraining *P. remota* and *P. quadrifolia* to be hybrids under the minimizing deep coalescences criterion (MDC). A. Best MDC network constraining *P. remota* as a hybrid. B. Best MDC network constraining *P. quadrifolia* as a hybrid. Reticulation events are represented by lines and inheritance probabilities are indicated at reticulations.

P. monophylla was sister to *P. quadrifolia* rather than to *P. californiarum*. With SNPs, *P. nelsonii* + *P. aristata* were sister to subsection *Cembroides* (50% bs), whereas with the nuclear exon alignment only *P. nelsonii* was recovered as sister to subsection *Cembroides* (61% bs).

DISCUSSION

Phylogeny of Pinus Subsection Cembroides—The phylogenies inferred with low-copy nuclear genes for *Pinus* subsection *Cembroides* recovered two main lineages (Figs. 2, 8). Nonetheless, some interrelationships vary among analysis and with previous studies. In studies with the nrDNA ITS region (Gernandt et al. 2001) and cpDNA (Gernandt et al. 2005; Parks et al. 2012; Ortiz-Medrano et al. 2016), *P. rzedowskii* is sister to the remaining species of *Pinus* subsection *Cembroides*. Here, *P. rzedowskii* is sister to *P. pinceana* and *P. maximartinezii* (100% bs). The same relationship of these large-cone pinyon pines was recovered with plastid DNA sequences by Gernandt et al. (2007) but did not receive high support (72% bs). In both the concatenated and coalescence-based analyses, the *P. rzedowskii* + *P. pinceana* + *P. maximartinezii* clade is always sister to the rest of the pinyon pines with strong support (100% bs). If *P. rzedowskii* forms a clade with *P. maximartinezii* + *P. pinceana*, instead of being the sister group of all the other species in subsection *Cembroides,* this implies either multiple gains of enlarged, functionally wingless seeds in subsection *Cembroides*, or a reversion to small winged seeds in *P. rzedowskii* (Gernandt et al. 2007). Maximum likelihood analyses of plastid DNA support the sister relationships between *P. discolor* + *P. johannis* and *P. culminicola* (Gernandt et al. 2007; Parks et al. 2012). The results with low-copy nuclear genes support that relationship with the concatenated alignment in RAxML and the SVDquartets species tree (Figs. 2, 7). In contrast, in the ASTRAL-III and SNPs trees (Figs. 3A, 8), *P. discolor* is recovered as sister to *P. culminicola* + *P. johannis*; this relationship was also supported with plastomes (Parks et al. 2012). Although Farjon and Styles (1997) treated *P. discolor* and *P. johannis* as a single variety of *P. cembroides* (as *P. cembroides* var. *bicolor* Little), the stomata of *P. discolor*, *P. johannis*, and *P. culminicola* are limited to the adaxial surfaces and the seed megagametophyte is white rather than pink. Our results with low-copy nuclear genes from the concatenated alignment, SVDquartets, and ASTRAL-III, agree with analyses of plastid DNA (Gernandt et al. 2007) in recovering *P. johannis* + *P. discolor* + *P. culminicola* as the sister group of *P. remota*, rather than grouping these species with *P. cembroides*. This relationship was not recovered with cladistic analyses of morphological characters (Malusa 1992). The coalescence analyses at the species level strongly support the clade of pinyon pines that are distributed in the Sierra Madre Oriental (SMO), *P. discolor* (mainly distributed in the Sierra Madre Occidental and the Sky Islands of the United States, but also present in the southern part of the Sierra Madre Oriental), *P. johannis*, and *P. culminicola* (100% bs in SVDquartets and ASTRAL-III; Figs. 7–8). The monophyly of the three taxa has been attributed to the evolution from an ancestor that was resistant to the calcareous soils that predominate in the SMO (Malusa 1992). Therefore, limestone soils tolerance may be a synapomorphy for *P. discolor*, *P. johannis*, and *P. culminicola*.

Infraspecific taxonomies of *P. cembroides* and their relationship to *P. lagunae* have disagreed (Zavarin and Snajberk 1985; Passini 1987; Farjon and Styles 1997; Farjon and Filer 2013). Phylogenetic results based on three plastid DNA regions recovered *P. cembroides* subsp. *cembroides*, *P. cembroides*

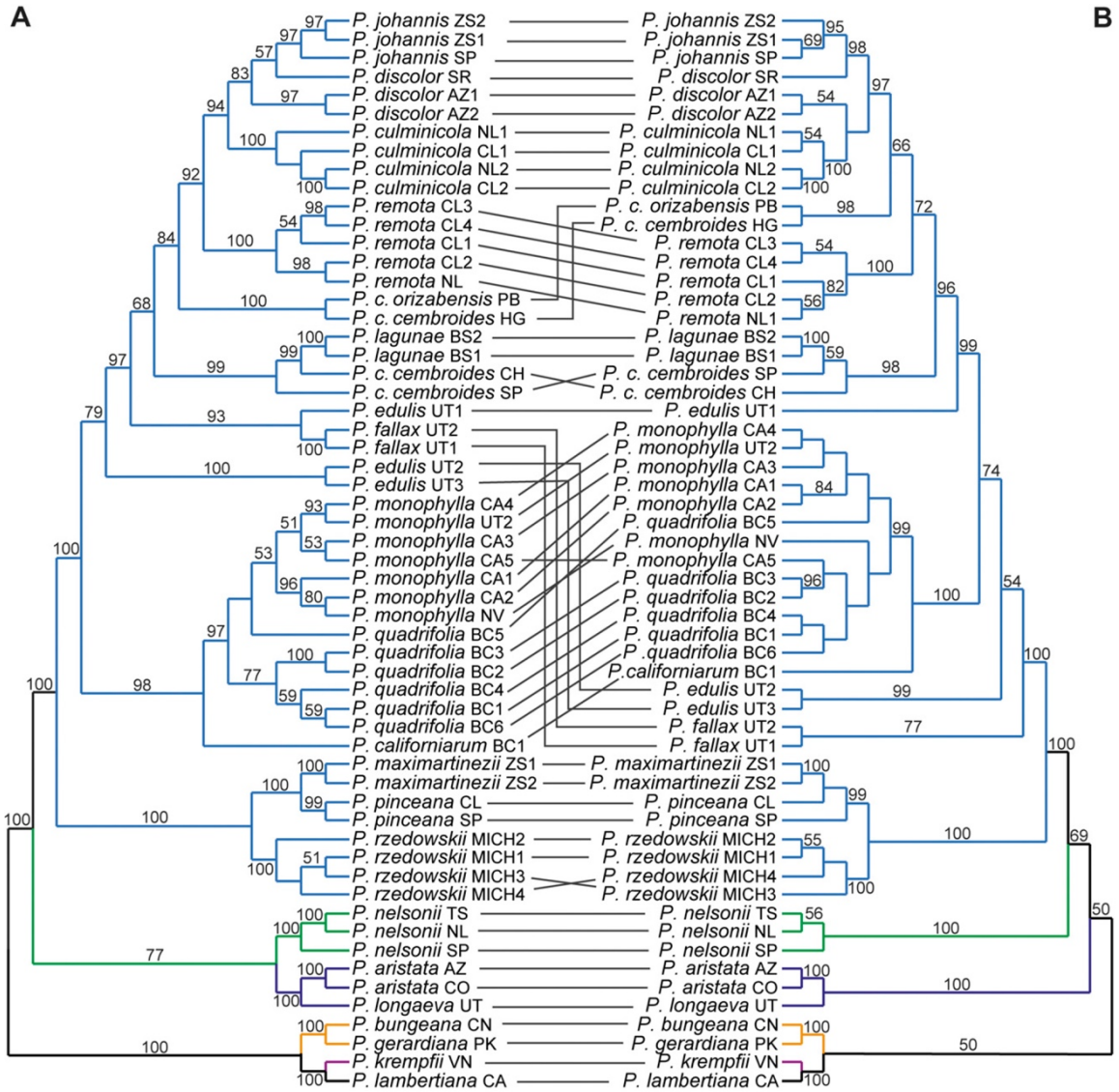


FIG. 6. Tanglegram of RAxML tree and SVDquartets lineages tree. A. Maximum likelihood tree based on concatenated alignment with 304 nuclear genes. B. Coalescent-based tree based on 1,000,000 quartets. Trees were estimated with a matrix of 222,129 bp and 58 terminals. Bootstrap values > 50% are shown above the branches. The colors of the branches follow Fig. 2.

subsp. *orizabensis*, and *P. lagunae* as a single exclusive lineage (Gernandt et al. 2003). Our results from nuclear gene assemblies (but not from SNPs; Fig. 9A) coincide with Gernandt et al. (2003) in recovering both varieties of *P. cembroides* and *P. lagunae* as an exclusive lineage (Figs. 7–9). Previously, Whang et al. (2001) reported differences of the leaf internal cuticle among *P. cembroides* subsp. *cembroides*, *P. cembroides* subsp. *orizabensis*, and *P. lagunae*. For instance, *P. cembroides* subsp. *orizabensis* differs from *P. cembroides* subsp. *cembroides* and *P. lagunae* by the width of the epidermal cell apex (thick in *P. cembroides* subsp. *orizabensis*), continuity of cell walls, stomatal apparatus shape, and cuticular flange-guard cell. *Pinus cembroides* subsp. *cembroides* and *P. lagunae* share more characters of the internal cuticle with each other than with *P. cembroides* subsp. *orizabensis*.

Zavarin and Snajberk (1985) found that the populations of *P. cembroides* subsp. *orizabensis* from southern Puebla and northeastern Veracruz differ from *Pinus cembroides* subsp. *cembroides* and *P. lagunae* in their chemical composition of monoterpenes but are very similar morphologically. They suggested that the divergence of southern populations of *P. cembroides* (*P. cembroides* subsp. *orizabensis*) as an isolated taxon most likely resulted from climatic and geographic isolation (middle Miocene). They also suggested that the isolation of *P. lagunae* from Baja California was related to movement of the coastal region from California during the

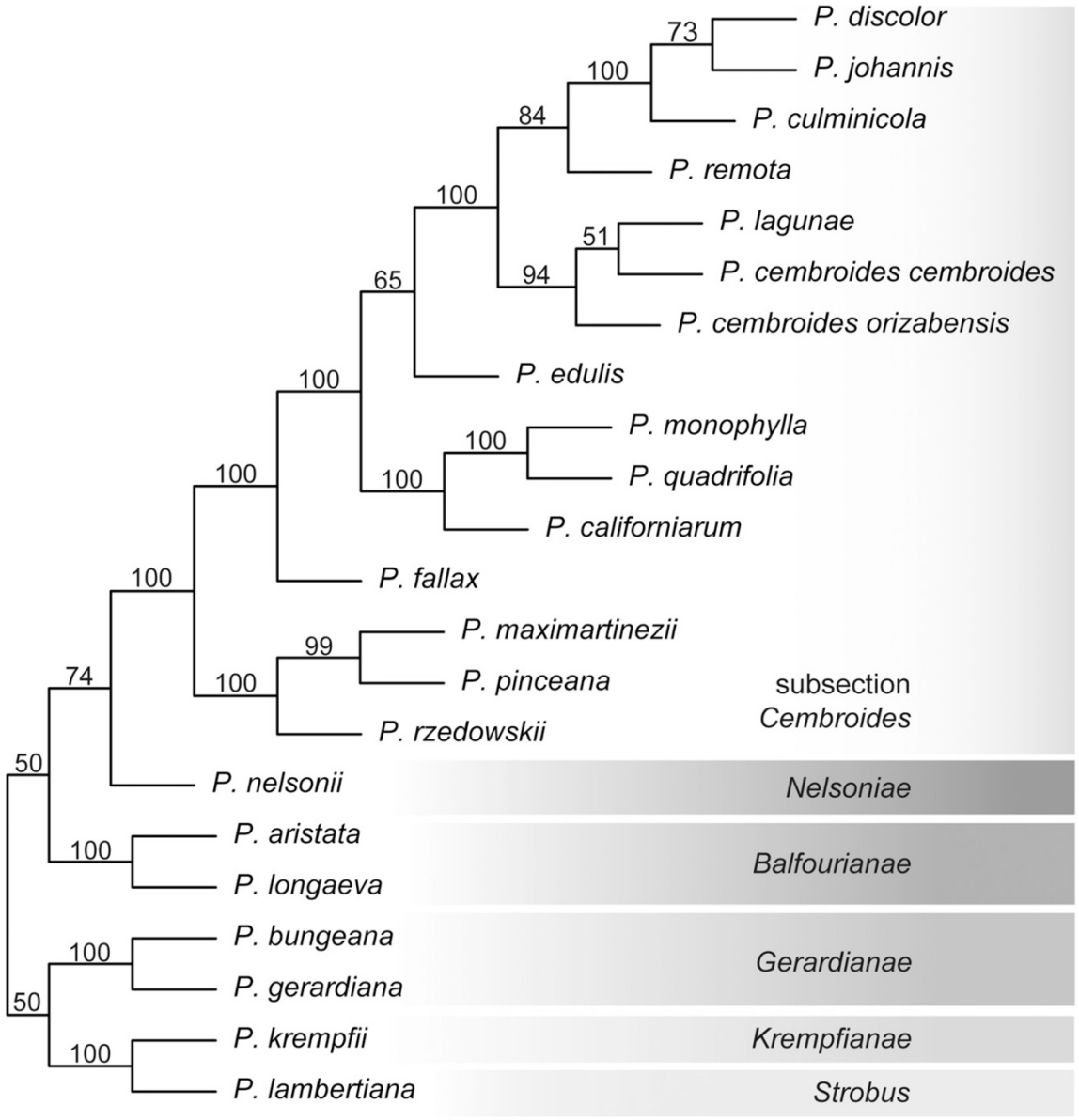


FIG. 7. Species tree inferred with SVDquartets. Coalescent-based tree estimated from 1,000,000 quartets (222,129 bp, 58 taxa and 304 genes). Total weight of incompatible quartets = 12.89%, and total weight of compatible quartets = 87.11%. Bootstrap values > 50% are shown above the branches.

Miocene and the formation of the Sierra Madre Occidental, resulting in the separation of this population from the rest of the continental populations by the breach of the Gulf of California (Zavarin and Snajberk 1985). However, this vicariance event is probably too old to explain divergence between *P. lagunae* and *P. cembroides*. The entire small-cone pinyon pine clade was estimated to have diversified during the Miocene, ~11 MYA (Gernandt et al. 2008; Saladin et al. 2017). More studies are needed to determine whether these three taxa represent independent evolutionary entities or the same species.

The coalescence-based analyses of concatenated genes and SNPs support the sister relationship between *P. monophylla* and *P. quadrifolia* and place both taxa as sister to *P. californiarum*. This relationship was not recovered with morphology, where *P. quadrifolia* is sister to *P. johannis* + *P. discolor* + *P. culminicola*, united by sharing resinous cones (Malusa 1992). In this study we observed two different topologies with respect to the origin of a single needle. In both the SVDquartets and ML analysis of the concatenated alignment, single-needle taxa occurred in independent lineages (Figs. 2, 7). In contrast, ASTRAL-III and analyses of SNPs

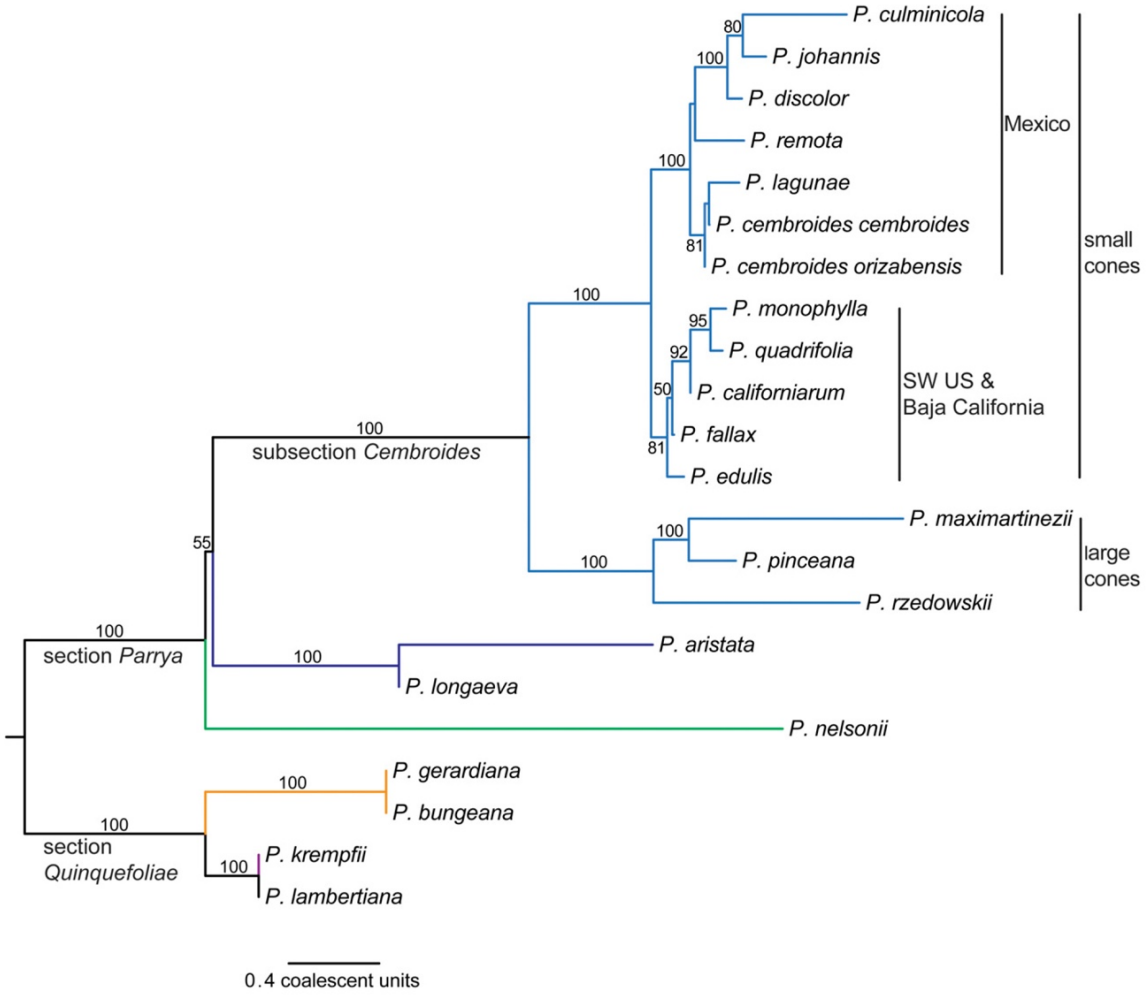


FIG. 8. Species tree inferred with ASTRAL-III. The tree was estimated under a species coalescent model using 304 nuclear gene trees with 58 individuals as input. Final quartet score = 73,455,073 and final normalized quartet score = 0.58 (very high incomplete lineage sorting). The bootstrap support values are provided above branches. The branch length represents coalescent units. The colors of the branches follow Fig. 2.

grouped *P. californiarum*, *P. fallax*, and *P. monophylla* together but paraphyletic to *P. quadrifolia* (Figs. 8–9). We performed a character state reconstruction with the ASTRAL-III species tree and the results supported a single reduction to single needles and common origin but with one independent loss in *P. quadrifolia*. Cole et al. (2013) compared variation in needle number to environmental variation and found that the proportions of the number of needles in *P. edulis* and *P. fallax* depend on annual fluctuations in precipitation. In addition, it was shown that *P. edulis* and *P. fallax* occur in an area with monsoon precipitation extremes, whereas *P. monophylla* and *P. californiarum* occur in areas with high levels of winter precipitation (Cole et al. 2008).

In morphology-based views of phylogeny, *P. monophylla*, *P. californiarum*, and *P. fallax* are recovered together by sharing resinous cones, single needles (predominantly), and thinner seed coats (Malusa 1992). With nuclear genes, *Pinus monophylla*, *P. californiarum*, and *P. fallax* are not recovered as monophyletic by the ASTRAL-III and SVDquartets analyses (Figs. 7–8). In fact, in our results *P. fallax* and *P. californiarum* are separate from *Pinus monophylla*, suggesting that *P. fallax* and *P. californiarum* are not taxonomic varieties of *P. monophylla* as proposed by Silba (1990). Taxonomic uncertainty between *P. monophylla* and *P. edulis* can be attributed to the existence of trees with both one and two leaves per fascicle (Tausch and West 1987). However, *P. edulis* is not sister to *P. monophylla*. Furthermore, environmental studies indicate that *P. edulis* is more similar to *P. fallax*, and *P. monophylla* is more similar to *P. californiarum*. Besides, *P. fallax* occurs in an area with moderate summer rains, similar to *P. edulis* (Malusa 1992). Our analyses support the separation of *P. californiarum* from *P. monophylla*. Bailey (1987) segregated *P. californiarum* from *P. monophylla* based principally on the length and amount of curl-back of fascicle sheaths, the number of leaf resin canals, and the number of rows of foliar stomata. The ASTRAL-III analysis recovered a clade with the species

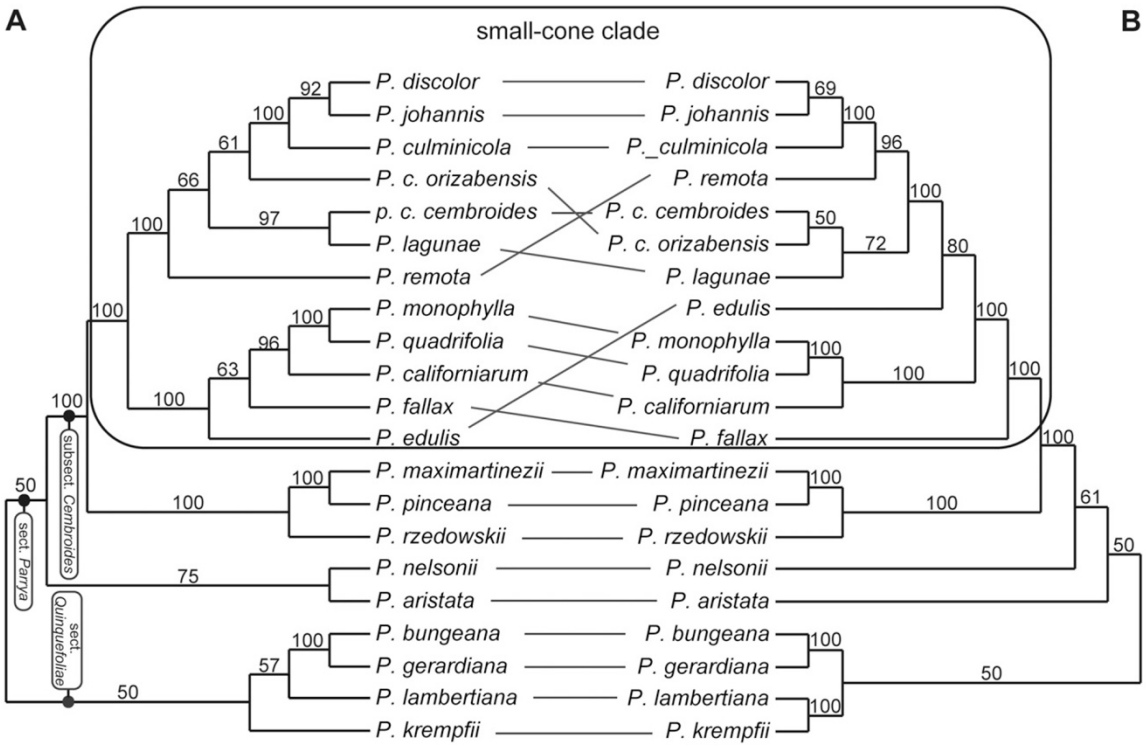


FIG. 9. Comparison between nuclear genes-based and single nucleotide polymorphisms-based trees inferred in SVDquartets. A. Single nucleotide polymorphisms tree based on 26,499 characters with a total weight of incompatible quartets = 5.35%, and total weight of compatible quartets = 94.65%. B. Coalescent-based tree based on 304 nuclear gene trees with a total weight of incompatible quartets = 14.05%, and total weight of compatible quartets = 85.95%. The trees were estimated from 1,000,000 quartets in 51 terminals. Bootstrap values > 50% are shown above the branches.

from the southwestern United States, *P. monophylla*, *P. californiarum*, *P. quadrifolia*, *P. edulis*, and *P. fallax*. The pinyon pines in this region are mainly allopatric or parapatric in distribution (Malusa 1992), but *P. californiarum* and *P. quadrifolia* co-occur in California and Baja California. *Pinus monophylla* occurs in California, extending east (and north) into the Great Basin in the states of Utah, Colorado, Arizona, and Idaho (Critchfield and Little 1966; Farjon 2005; Cole et al. 2008), and populations of *P. fallax* and *P. edulis* co-occur in Arizona and New Mexico (*P. edulis* reaches eastern Nevada and southeastern California). If *P. edulis* and *P. fallax* (adapted to early summer or periodic drought) are paraphyletic to *P. monophylla* (adapted to summer-autumn drought), *P. californiarum*, and *P. quadrifolia* (Figs. 2, 7), this relationship may be explained by vicariance or may have evolved in response to summer drought (Cole et al. 2008). In fact, these taxa are distributed widely in geographical regions with distinct precipitation regimes (Cole et al. 2008). Particularly, *P. monophylla* occurs in regions with different precipitation from *P. fallax* and *P. edulis*, which are characterized by high monsoon precipitation (Cole et al. 2008). The ecological similarities of *P. fallax* to *P. edulis* (Bailey 1987) rather than to *P. monophylla*, and its genetic distinctness from *P. monophylla* (Fig. 8), support our decision to treat *P. fallax* as a separate species from *P. monophylla*.

Hybridization and Introgression—Both natural and artificial hybridization have been well-documented in conifers (Saylor and Smith 1966; Lanner 1974a, 1974b; Delgado et al. 2007; Wachowiak et al. 2011; Zhou et al. 2017). Overlapping phenology and weak reproductive barriers can influence the direction of pollen-mediated gene flow in natural populations (Hamilton et al. 2013). In conifers, plastid DNA is paternally inherited, and higher gene flow for plastid DNA has been attributed to the high migration capacity of wind-dispersed pollen (Rieseberg and Soltis 1991; Petit et al. 2005).

In pines, plastid introgression has been observed at a low or moderate frequency in sympatric or parapatric populations (Dounavi et al. 2001; Delgado et al. 2007; Zhou et al. 2017). In closely-related *Pinus* species, shared genetic variation can be the result of introgression following secondary contact, with genetic differentiation in parapatric populations lower than in allopatric populations (Zhou et al. 2017). Using the minimizing deep coalescence criterion, we explored reticulation in subsection *Cembroides*. No gene flow was detected between subsection *Cembroides* and other closely related lineages (Figs. 4–5). This result supports the observation by Mirov (1967) that species of subsection *Cembroides* do not form hybrids with species from other pine subsections. In fact, *Pinus* subsection *Cembroides* may have diverged from other subsections (particularly *Balfourianae*) relatively early (Axelrod 1986).

The majority of taxa involved in the reticulation events detected here have geographic distributions that are somewhat close to one another (*P. californiarum*, *P. edulis*, *P. fallax*, *P. monophylla*, *P. quadrifolia*, and *P. remota*; only *P. lagunae* and *P. cembroides* subsp. *orizabensis* are geographically isolated). This coincides with other studies where gene flow has been

reported (Edwards-Burke et al. 1997; Delgado et al. 2007; Zhou et al. 2017). Our results coincide with some hypotheses proposed by Lanner (1974a). We detected gene flow in *P. edulis* in only one reticulation scenario (Fig. 4B). This suggests that *P. edulis* is introgressed with *P. monophylla*. Some populations of *P. edulis* and *P. monophylla* occur in sympatry in the eastern Great Basin where trees of *P. edulis* with both single needles and two needles per fascicle have been observed (Lanner 1974a). *Pinus edulis* also occurs in Arizona and New Mexico (Cole et al. 2013), and sympatric populations of *P. edulis* and *P. monophylla* have been reported in in the Mojave Desert in southeastern California (Munz and Keck 1959; Critchfield and Little 1966), western Utah, and Nevada (Farjon and Filer 2013). For this reason, it would not be unusual if *P. edulis* is introgressed with *P. monophylla* where their populations are in contact. The direction of gene flow inferred from *P. monophylla* to *P. edulis* is consistent with the prevailing winds, which are from west to east. Overlapping phenology could facilitate introgression from *P. monophylla* to *P. edulis*, since both disperse pollen in a short period in the spring (Lanner 1970; Farjon and Styles 1997). Furthermore, the distribution range of *P. edulis* may have been more extensive in the past (Cole et al. 2008), resulting in more widespread contact with *P. monophylla*.

The MDC method detected reticulation in taxa for which gene flow had not been suspected. Currently *Pinus fallax* and *P. cembroides* subsp. *cembroides* have widely separated distributions. One possible explanation for reticulation between the two species is that this inference is incorrect (Fig. 4C). As an alternative explanation, we suggest studying species distribution and demographic history to test whether these two taxa came into contact in the past. The potential distribution and demographic history for *P. fallax* suggests that it may have been more widely distributed in southwestern California, southern Nevada, throughout Arizona and extending beyond into Utah, Colorado, and New Mexico (Cole et al. 2008). Another detected reticulation event that was unexpected was between *P. lagunae* and *P. edulis* (Fig. 4C). These species also are widely separated geographically. The genetic diversity shared by these allopatric taxa seems more likely to be influenced by the retention of ancestral polymorphism. However, ancient introgression events may have been interrupted by migration of the populations of *P. edulis* northward in present-day USA during the Holocene (Cole et al. 2008).

We detected reticulation in *P. quadrifolia* and *P. culminicola* (Fig. 4D) although their populations are allopatric and gene flow had not been suspected. This inference may be incorrect, or long-distance pollen dispersal could have resulted in introgression. It would be interesting to study the past distribution and demographic history of *P. culminicola*, *P. cembroides* subsp. *cembroides*, and *P. lagunae* to test whether they were formerly in contact with other species.

The Phylonet analysis also detected reticulation in taxa such as *P. lagunae* and *P. cembroides* subsp. *cembroides* (Fig. 4D) for which gene flow had not been suspected. The origin of these reticulations implies the existence of an extinct taxon. It is also possible that this inference is incorrect or not significant. Copetti et al. (2017) explained the origin of a reticulation with the existence of an extinct or unsampled taxon in cacti, but the authors do not discuss the result.

Specifying *P. quadrifolia* as a hybrid under MDC (Fig. 5B), our results did not indicate that *P. quadrifolia* is introgressed with *P. californiarum* as reported by Lanner (1974b). *Pinus californiarum* was recovered as the sister to *P. quadrifolia* and *P. monophylla*, but no reticulation was detected between *P. californiarum* or *P. monophylla*. We did detect reticulation in *P. quadrifolia* and *P. lagunae* for which gene flow had not been suspected (Fig. 5B). Long-distance pollen dispersal could have resulted in introgression. Likewise, *Pinus lagunae* may be a relictual population left behind from a time when its ancestors had a range that extended northwest into what is today Baja California and southern California.

Another species reported as a possible hybrid is *P. remota*. Our results do not support the proposal of Little. Little (1968) suggested the possibility that in the past its populations had been in contact with those of *P. edulis*, permitting introgressive hybridization. Nonetheless, reticulation with *P. remota* was inferred from *P. fallax* and not from *P. edulis* (Fig. 5A; Little 1968). No contemporary *P. fallax* populations come into contact with *P. remota*, but past secondary contact could have resulted in introgression. Some populations of *P. fallax* (western New Mexico) occur in proximity to *P. remota* (Texas and northeast Mexico). According to the packrat middens record for the Late Quaternary, *P. remota* appear to have expanded into the south of Edwards Plateau to the west or southwest (Van Devender 1990). Likewise, *P. fallax* in the Sonoran Desert may have expanded from California chaparral to Arizona across the Pleistocene-Holocene boundary (Betancourt et al. 1990).

Although some populations of *P. californiarum* and *P. monophylla* are found in limited sympatry in California (San Bernardino Co.), we did not detect gene flow in any direction, but the admixture analysis provided additional information about the relationship between *P. californiarum* and *P. monophylla* (Fig. 3A). Admixture analyses should be interpreted with care since it is difficult to estimate the real number of clusters, especially for species with long generation times like conifers; however, with the increased use of sequencing technologies and massive generation data, our capacity to detect admixture has substantially improved (Pritchard et al. 2000). The admixture analysis we carried out on the pinyon pines subclade from SW US and Baja California provided a picture of possible interbreeding in this group of pines. Distribution of ancestry fractions indicate that *P. californiarum* is introgressed but this scenario was not recovered in Phylonet. The admixture results indicate that *P. quadrifolia* shares little genetic variation with *P. monophylla* (s. s.). In addition, it stands out that *P. quadrifolia* is clustered in a singular population with $K > 2$. This pattern of ancestral structure suggests that *P. quadrifolia* accumulates particular genetic diversity and could be a valid species and not a hybrid. For $K = 5$ (data not shown) *P. monophylla* was not clustered into a separate population. It is also important to mention that according to clustering patterns *P. fallax* seems to share more genetic variation with *P. edulis* than *P. monophylla*. Clustering of *P. fallax* into a new well-defined population at a particular value of K could reflect enough genetic differentiation from *P. edulis* to support its isolation as a species. Although admixture patterns support in most cases our phylogenetic results, increased sampling is required for a more complete perspective of the admixture events for these complex and widely geographically distributed populations. Despite the reduced sample size, we were able to provide a preliminary insight into the admixture events occurring in this group of pines.

In conclusion, using target enrichment to characterize 304 nuclear genes and 26,499 SNPs, we corroborated the monophyly of *Pinus* subsection *Cembroides* in all analyses. The inferred phylogenies also corroborate other relationships

previously recovered with plastid DNA. The results suggest that *P. fallax* and *P. californiarum* could be considered as valid species rather than as infraspecific taxa of *P. monophylla* or *P. edulis*. Also, our admixture results suggest that *P. quadrifolia* accumulates particular genetic diversity and could be a valid species and not a hybrid. The single-needle pinyons were recovered as a non-monophyletic group, with character reconstructions consistent with a single derivation and subsequent loss of the single-needle condition. The ASTRAL-III tree was consistent with the presence of ILS (very high) in the group of pinyon pines with small cones based on the short length in coalescent units of internal branches. Respecting reticulation events, we identified *P. remota* as having genes introgressed from *P. fallax*, and *P. quadrifolia* having genes introgressed from *P. lagunae*. Some hybridization scenarios were unexpected and not reported in the literature. Finally, further study is needed to determine the relative roles of ILS and introgression in explaining shared genetic diversity in *Pinus* subsection *Cembroides*.

Acknowledgments

The authors thank Alejandra Vázquez-Lobo and Patricia Ornelas for valuable suggestions. We also thank Aaron Liston for his collaboration in the study design, and Xitlali Aguirre-Dugua, Mario Montes, and Angélica Castolo for their participation in fieldwork, and José Delgadillo Rodríguez for logistical support. We also thank Patricia Rosas Escobar for laboratory assistance, and Alison Devault and Jake Enk at Arbor Biosciences for valuable technical assistance. We also thank Dra. Lidia I. Cabrera and the LANABIO of the Instituto de Biología, UNAM. Funding was provided by PAPIIT-DGAPA, UNAM Grant IN209816 and CONACyT Grant 221694. This work is part of the Ph.D. thesis of J.R. Montes in the Posgrado de Ciencias Biológicas, UNAM.

Author Contributions

JRM performed field- and labwork, assembled the DNA sequences, performed the phylogenetic analyses, and was the primary author for the manuscript. JRM, AML, AW, DP, and DSG designed the study. PP provided SNP data and performed admixture analyses. AML and DSG participated in fieldwork and analyses. All authors reviewed and edited the manuscript.

Literature Cited

Aberer, A. J., D. Krompass, and A. Stamatakis. 2013. Pruning rogue taxa improves phylogenetic accuracy: An efficient algorithm and webservice. *Systematic Biology* 62: 162–166.

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

Axelrod, D. I. 1986. Cenozoic history of some western American pines. *Annals of the Missouri Botanical Garden* 73: 565–641.

Bailey, D. K. 1987. A study of *Pinus* subsection *Cembroides* I: The single-needle pinyons of the Californias and the Great Basin. *Notes from the Royal Botanic Garden Edinburgh* 44: 275–310.

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

Betancourt, J. L., T. R. Van Devender, and P. S. Martin. 1990. *Packrat Middens: The Last 40,000 Years of Biotic Change*. Tucson, Arizona: University of Arizona Press.

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

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

Critchfield, W. B. and E. L. Little. 1966. *Geographic Distribution of the Pines of the World*. Washington, D.C.: Department of Agriculture Miscellaneous Publication 991.

Cole, K. L., J. F. Fisher, S. T. Arundel, J. Cannella, and S. Swift. 2008. Geographical and climatic limits of needle types of one-and two-needled pinyon pines. *Journal of Biogeography* 35: 257–269.

Cole, K. L., J. F. Fisher, K. Ironside, J. I. Mead, and P. Koehler. 2013. The biogeographic histories of *Pinus edulis* and *Pinus monophylla* over the last 50,000 years. *Quaternary International* 310: 96e110.

Copetti, D., A. Búrquez, E. Bustamante, J. L. M. Charboneau, K. L. Childs, L. E. Eguiarte, S. Lee, T. L. Liu, M. M. McMahon, N. K. Whiteman, R. A. Wing, M. F. Wojciechowski, and M. J. Sanderson. 2017. Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences USA* 114: 12003–12008.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

Debreczy, Z. and I. Rácz. 2011. *Conifers Around the World*. Budapest: DendroPress Ltd.

DeGiorgio, M., J. Syring, A. J. Eckert, A. Liston, R. Cronn, D. B. Neale, and N. A. Rosenberg. 2014. An empirical evaluation of two-stage species tree inference strategies using a multilocus dataset from North American pines. *BMC Evolutionary Biology* 14: 67.

Delgado, P., R. Salas-Lizana, A. Vázquez-Lobo, A. Wegier, M. Anzidei, E. R. Alvarez-Buylla, G. G. Vendramin, and D. Piñero. 2007. Introgressive hybridization in *Pinus montezumae* Lamb. and *Pinus pseudostrobus* Lindl. (Pinaceae): Morphological and molecular (cpSSR) evidence. *International Journal of Plant Sciences* 168: 861–875.

Dounavi, K. D., N. Koutsias, and K. P. Panetsos. 2001. Natural interspecific hybridization between *Pinus brutia* (Ten.) and *Pinus halepensis* (Mill.), verified by using the logistic regression modeling on morphological characters. *Forest Genetics* 8: 151–158.

Doyle, J. J. and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Eckenwalder, J. E. 2009. *Conifers of the World: The Complete Reference*. Portland, Oregon: Timber Press.

Edwards-Burke, M. A., J. L. Hamrick, and R. A. Price. 1997. Frequency and direction of hybridization in sympatric populations of *Pinus taeda* and *P. echinata* (Pinaceae). *American Journal of Botany* 84: 879–886.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32: 3047–3048.

Farjon, A. 2005. *Pines: Drawings and Description of the Genus* Pinus. Leiden, Netherlands: Brill.

Farjon, A. and B. Filer. 2013. *An Atlas of the World's Conifers: An Analysis of their Distribution, Biogeography, Diversity and Conservation Status*. Leiden, Netherlands: Brill.

Farjon, A. and B. Styles. 1997. *Flora Neotropica Monograph*: Pinus *(Pinaceae)*. New York: The New York Botanical Garden.

Flores-Rentería, L., A. Wegier, D. Ortega Del Vecchyo, A. Ortíz-Medrano, D. Piñero, A. V. Whipple, F. Molina-Freaner, and C. A. Domínguez. 2013. Genetic, morphological, geographical and ecological approaches reveal phylogenetic relationships in complex groups, an example of recently diverged pinyon pine species (subsection *Cembroides*). *Molecular Phylogenetics and Evolution* 69: 940–949.

Gernandt, D. S., A. Liston, and D. Piñero. 2001. Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: Implications for molecular systematic studies of pine species complexes. *Molecular Phylogenetics and Evolution* 21: 449–467.

Gernandt, D. S., A. Liston, and D. Piñero. 2003. Phylogenetic of *Pinus* subsections *Cembroides* and *Nelsoniae* inferred from cpDNA sequences. *Systematic Botany* 28: 657–673.

Gernandt, D. S., G. Gaeda López, S. Ortiz García, and A. Liston. 2005. Phylogeny and classification of *Pinus*. *Taxon* 54: 29–42.

Gernandt, D. S., O. Zerón, and I. Goyenechea. 2007. Inferencia filogenética mediante secuencias de DNA: Un ejemplo con los pinos piñoneros. Pp. 55–65 In *La Sistemática Base del Conocimiento de la Biodiversidad*, vol. 2, eds. A. Contreras-Ramos, C. Cuevas Cardona, I. Goyenechea, and U. Iturbide. Hidalgo: Universidad Autónoma del Estado de Hidalgo.

Gernandt, D. S., S. Magallón, G. Geada López, O. Zerón Flores, A. Willyard, and A. Liston. 2008. Use of simultaneous analyses to guide fossil-based calibrations of Pinaceae phylogeny. *International Journal of Plant Sciences* 169: 1086–1099.

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

Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D. B. Jaffe, E. S. Lander, and C. Nusbaum. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27: 182–189.

Hamilton, J. A., C. Lexer, and S. N. Aitken. 2013. Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* x *P. glauca*). *Molecular Ecology* 22: 827–841.

Huson, D. H. and C. Scornavacca. 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic Biology* 61: 1061–1067.

IUCN. 2017. IUCN red list of threatened species version 2017–3. Gland, Switzerland and Cambridge, UK: IUCN.

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

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

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

Keng, H. and E. L. Little. 1961. Needle characteristics of hybrid pines. *Silvae Genetica* 10: 131–146.

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

Lanner, R. M. 1970. Origin of the summer shoot of pinyon pines. *Canadian Journal of Botany* 48: 1759–1765.

Lanner, R. M. 1974a. Natural hybridization between *Pinus edulis* and *Pinus monophylla* in the American Southwest. *Silvae Genetica* 23: 108–116.

Lanner, R. M. 1974b. A new pine from Baja California and the hybrid origin of *Pinus quadrifolia*. *The Southwestern Naturalist* 19: 75–95.

Lanner, R. M. 1981. *The Piñon Pine: A Natural and Cultural History*. Reno, Nevada: University of Nevada Press.

Lanner, R. M. and A. M. Phillips. 1992. Natural hybridization and introgression of pinyon pines in northwestern Arizona. *International Journal of Plant Sciences* 153: 250–257.

Leaché, A. D. and J. R. Oaks. 2017. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology Evolution and Systematics* 48: 69–84.

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

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

Little, E. L. 1966. A new pinyon variety from Texas. *Wrightia* 3: 181–187.

Little, E. L. 1968. Two new pinyon varieties from Arizona. *Phytologia* 17: 329–342.

Little, E. L. and W. B. Critchfield. 1969. *Subdivisions of the Genus* Pinus *(Pines)*. Washington D.C.: US Forest Service Miscellaneous Publication 1144.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.

Maddison, W. P. and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55: 21–30.

Maddison, W. P. and D. R. Maddison. 2018. Mesquite: A modular system for evolutionary analysis, version 3.6. Technical report. Software available from <http://mesquiteproject.org>.

Madrigal, S. X. and D. M. Caballero. 1969. Una nueva especie mexicana de *Pinus*. *Boletín Técnico del Instituto Nacional de Investigaciones Forestales* 26: 1–11.

Malusa, J. 1992. Phylogeny and biogeography of the pinyon pines (*Pinus* subsect. *Cembroides*). *Systematic Botany* 17: 42–66.

Mirarab, S. and T. Warnow. 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31: i44–i52.

Mirov, N. T. 1967. *The Genus* Pinus. New York: Ronald Press.

Montes, J. R., P. Peláez, A. Willyard, A. Moreno-Letelier, D. Piñero, and D. S. Gernandt. 2019. Data from: Phylogenetics of *Pinus* subsection *Cembroides* Engelm. (Pinaceae) inferred from low-copy nuclear gene sequences. Dryad Digital Repository. https://doi.org/10.5061/dryad.1f8p5d6.

Morales-Briones, D. F., K. Romoleroux, F. Kolár, and D. C. Tank. 2018. Phylogeny and evolution of the neotropical radiation of *Lachemilla* (Rosaceae): Uncovering a history of reticulate evolution and implications for infrageneric classification. *Systematic Botany* 43: 17–35.

Munz, P. A. and D. D. Keck. 1959. *A California Flora*. Berkeley, California: University of California Press.

Neale, D. B., J. L. Wegrzyn, K. A. Stevens, A. V. Zimin, D. Puiu, M. W. Crepeau, C. Cardeno, M. Koriabine, A. E. Holtz-Morris, J. D. Liechty, P. J. Martínez-García, H. A. Vasquez-Gross, B. Y. Lin, J. J. Zieve, W. M. Dougherty, S. Fuentes-Soriano, L. S. Wu, D. Gilbert, G. Marçais, M. Roberts, C. Holt, M. Yandell, J. M. Davis, K. E. Smith, J. F. Dean, W. W. Lorenz, R. W. Whetten, R. Sederoff, N. Wheeler, P. E. McGuire, D. Main, C. A. Loopstra, K. Mockaitis, P. J. de Jong, J. A. Yorke, S. L. Salzberg, and C. H. Langley. 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* 15: R59.

Neves, L. G., J. M. Davis, W. B. Barbazuk, and M. Kirst. 2013. Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal* 75: 146–156.

Ortiz-Medrano, A., D. P. Scantlebury, A. Vázquez-Lobo, A. Mastretta-Yanes, and D. Piñero. 2016. Morphological and niche divergence of pinyon pines. *Ecology and Evolution* 6: 2886–2896.

Parks, M., R. Cronn, and A. Liston. 2012. Separating the wheat from the chaff: Mitigating the effects of noise in plastome phylogenomic data set from *Pinus* L. (Pinaceae). *BMC Evolutionary Biology* 12: 100.

Passini, M. F. 1987. The endemic pinyon of lower California: *Pinus lagunae*. *Phytologia* 63: 337–338.

Pérez de la Rosa, J. A., S. A. Harris, and A. Farjon. 1995. Noncoding chloroplast DNA variation in Mexican pines. *Theoretical and Applied Genetics* 91: 1101–1106.

Petit, R. J. and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics* 37: 187–214.

Petit, R. J., J. Duminil, S. Fineschi, A. Hampe, D. Salvini, and G. G. Vendramin. 2005. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* 14: 689–701.

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

Price, R. A., A. Liston, and S. H. Strauss. 1998. Phylogeny and systematics of *Pinus*. Pp. 49–68 in *Ecology and Biogeography of Pinus*, ed. D. M. Richardson. Cambridge, UK: Cambridge University Press.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81: 559–575.

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

Rieseberg, L. H. and D. E. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5: 65–84.

Saladin, B., A. B. Leslie, R. O. Wüest, G. Litsios, E. Conti, N. Salamin, and N. E. Zimmermann. 2017. Fossils matter: Improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evolutionary Biology* 17: 95.

Saylor, L. C. and B. W. Smith. 1966. Meiotic irregularity in species and interspecific hybrids of *Pinus*. *American Journal of Botany* 53: 453–468.

Sayyari, E. and S. Mirarab. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* 33: 1654–1668.

SEMARNAT. 2010. Norma Oficial Mexicana NOM-059-ECOL-2001. Protección ambiental. Especies nativas de México de flora y fauna silvestres. Categorías de riesgo y especificaciones para su inclusión, exclusión o cambio. Lista de especies en riesgo. Diario Oficial de la Federación, 30 de diciembre de 2010.

Silba, J. 1990. A supplement to the international census of the coniferae, II. *Phytologia* 68: 7–78.

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

Swofford, D. L. 2002. PAUP* Phylogenetic analysis using parsimony (*and other methods), v. 4.0b10. Sunderland, Massachusetts: Sinauer Associates.

Syring, J., A. Willyard, R. Cronn, and A. Liston. 2005. Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. *American Journal of Botany* 92: 2086–2100.

Tausch, R. J. and N. E. West. 1987. Morphological variation/precipitation relationships of Great Basin single-needled pinyon. Pp. 86–91 In *Proceedings of the Pinyon-Juniper Conference*, ed. R. L. Everett. Reno, Nevada: General Technical Report INT-215. U.S.D.A., Forest Service.

Than, C. and L. Nakhleh. 2009. Species tree inference by minimizing deep coalescences. *PLoS Computational Biology* 5: e1000501.

Than, C., D. Ruths, and L. Nakhleh. 2008. PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9: 322.

Van Devender, T. R. 1990. Late Quaternary vegetation and climate of the Chihuahuan Desert, United States and Mexico. Pp. 104–133 In *Packrat Middens: The Last 40,000 Years of Biotic Change*, eds. J. L Betancourt, T. R. Van Devender, and P. S. Martin. Tucson, Arizona: The University of Arizona Press.

Wachowiak, W., A. E. Palmé, and O. Savolainen. 2011. Speciation history of three closely related pines *Pinus mugo* (T.), *P. uliginosa* (N.) and *P. sylvestris* (L.). *Molecular Ecology* 20: 1729–1743.

Wang, J., R. J. Abbott, Y. L. Peng, F. K. Du, and J.-Q. Liu. 2011. Species delimitation and biogeography of two fir species (*Abies*) in central China: Cytoplasmic DNA variation. *Heredity* 107: 362–370.

Wegrzyn, J. L., J. D. Liechty, K. A. Stevens, L.-S. Wu, C. A. Loopstra, H. A. Vasquez-Gross, W. M. Dougherty, B. Y. Lin, J. J. Zieve, P. J. Martínez-García, C. Holt, M. Yandell, A. V. Zimin, J. A. Yorke, M. W. Crepeau, D. Puiu, S. L. Salzberg, P. J. de Jong, K. Mockaitis, D. Main, C. H. Langley, and D. B. Neale. 2014. Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* 196: 891–909.

Weitemier, K., S. C. K. Straub, R. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. HybSeq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 9: 1400042.

Whang, S. S., J.-H. Pak, R. S. Hill, and K. Kim. 2001. Cuticle micromorphology of the leaves of *Pinus* (Pinaceae) from Mexico and Central America. *Botanical Journal of the Linnean Society* 135: 349–373.

Willyard, A., J. Syring, D. S. Gernandt, A. Liston, and R. Cronn. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution* 24: 90–101.

Zavarin, E. and K. Snajberk. 1985. Monoterpenoid and morphological differentiation within *Pinus cembroides*. *Biochemical Systematics and Ecology* 13: 89–104.

Zhang, C., E. Sayyari, and S. Mirarab. 2017. ASTRAL-III: Increased scalability and impacts of contracting low support branches. Pp. 53–75 in *Comparative Genomics, 15th International Workshop, RECOMB CG*, eds. J. Meidanis and L. Nakhleh. Barcelona, Spain: Springer Press. doi: 10.1007/978-3-319-67979-2_4.

Zhou, Y., L. Duvaux, G. Ren, L. Zhang, O. Savolainen, and J. Liu. 2017. Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity* 118: 211–220.

Appendix 1. Collection information for individuals included in the study. Voucher information for this study, presented in the following order: Taxon; voucher specimen: collector and number, (herbarium acronym), locality.

Ingroup: ***Pinus californiarum*** D.K.Bailey; *D.S. Gernandt 403, 1561*, (MEXU), Mexico, Baja California. ***Pinus cembroides*** **subsp.** ***cembroides*** Zucc.; *D.S. Gernandt 444, 593, 1042*, (MEXU), Mexico. ***Pinus cembroides*** **subsp.** ***orizabensis*** D.K.Bailey; *D.S. Gernandt 7399*, (MEXU), Mexico, Puebla. ***Pinus culminicola*** Andresen & Beaman; *D.S. Gernandt 1135, 1137, 01S6, D.O. Burge 1212*, (MEXU), Mexico. ***Pinus discolor*** D.K.Bailey & Hawksw.; *D.S. Gernandt 1067*, (MEXU), Mexico, Sonora, *D.S. Gernandt 785*, (MEXU), *F. Hammond 02S2*, (OSC), United States, Arizona. ***Pinus edulis*** Engelm.; *D.S. Gernandt 485, 1020, 1028*, (MEXU), United States, Utah. ***Pinus fallax*** (Little) Businský; *D.S. Gernandt 492, 494*, (MEXU), United States, Utah. ***Pinus johannis*** M.F.Robert; *D.S. Gernandt 501, 7999, 8199*, (MEXU), Mexico. ***Pinus lagunae*** (Robert-Passini) D.K.Bailey; *A.M. González 9279, 6399*, (MEXU), Mexico, Baja California Sur. ***Pinus maximartinezii*** Rzed. *D.S. Gernandt 1010, 7799*, (MEXU), Mexico, Zacatecas. ***Pinus monophylla*** Torr. & Frém.; *D.S. Gernandt 478, 480, 1214, 1509, 1512, 1513, A. Liston 1298, R. Halse 6668*, (MEXU), United States. ***Pinus pinceana*** Gordon; *D.S. Gernandt 1163, 8999*, (MEXU), Mexico. ***Pinus quadrifolia*** Parl. ex Sudw.; *D.S. Gernandt 961, 1099, 1499, 1560, 1599*, (MEXU), *D. Gernandt, A. Liston & Ann Willyard 035*, (OSC), Mexico, Baja California. ***Pinus remota*** (Little) D.K.Bailey & Hawksw.; *D.S. Gernandt 801, 1301 19498, 22498, 23298*, (MEXU), Mexico. ***Pinus rzedowskii*** Madrigal & M. Caball. *D.S. Gernandt 635, 636, 637*, (MEXU), *R. Businský 47131*, (OSC), Mexico, Michoacán.

Outgroup: ***Pinus aristata*** Engelm.; *K. Ferrell 30, 37*, (OSC), United States. ***Pinus bungeana*** Zucc. ex Endl.; *J.E.R 03S3A*, (OSC), China. ***Pinus gerardiana*** Wall. ex D.Don; *R. Businský 41105*, (OSC), Pakistan, Gilgit-Baltistan. ***Pinus krempfii*** Lecomte; *P. Thomas 242*, (E), Vietnam, Lam Dong. ***Pinus lambertiana*** Douglas; *D.S. Gernandt 1195*, (MEXU), United States, California. ***Pinus longaeva*** D.K.Bailey; *D.S. Gernandt 1027*, (MEXU), United States, Utah. ***Pinus nelsonii*** Shaw. *D.S. Gernandt 1096, 10198, 31798*, (MEXU), Mexico.

Appendix 2. Sequence statistics for the 60 *Pinus* samples assembled in HybPiper.For each sample, species name is followed by sample ID; sequencing run; yield (Mb); reads; reads mapped; genes mapped; and percent recovered gene.

Subsection ***Cembroides***. ***Pinus californiarum***: DSG1509; 2; 894; 7,601,929; 3,800,860; 969; 97.3. ***P. californiarum***: DSG1561; 1; 932; 8,243,891; 4,121,790; 969; 97.3. ***P. californiarum:*** DSG1512; 1; 999; 8,764,857; 4,382,336; 969; 97.3. ***P. californiarum***: DSG1513; 2; 1108; 9,405,258; 4,702,662; 969; 97.3. ***P. californiarum***: DSG403; 2; 1150; 9,862,288; 4,931,170; 968; 97.2. ***P. californiarum***: AL1298; 2; 1283; 11,713,774; 5,856,883; 969; 97.3. ***P. cembroides*** **subsp.** ***cembroides***: DSG593; 2; 1003; 8,658,650; 4,329,349; 969; 97.3. ***P. cembroides*** **subsp.** ***cembroides***: DSG444; 2; 1203; 10,785,141; 5,392,515; 969; 97.3. ***P. cembroides*** **subsp.** ***cembroides***: DSG1042; 2; 1275; 11,310,686; 5,655,408; 969; 97.3. ***P. cembroides*** **subsp.** ***orizabensis***: DSG7399; 2; 923; 8,474,182; 4,237,030; 968; 97.2. ***P. culminicola***: DOB1212; 1; 733; 6,396,584; 3,198,382; 968; 97.2. ***P. culminicola***: DSG1135; 2; 1068; 9,108,291; 4,553,996; 969; 97.3. ***P. culminicola***: 01S6; 2; 1120; 9,549,507; 4,774,712; 968; 97.2. ***P. culminicola***: DSG1137; 2; 1223; 10,962,417; 5,481,235; 969; 97.3. ***P. discolor***: 02s2; 2; 1036; 9,070,577; 4,535,221; 968; 97.2. ***P. discolor***: DSG1067; 2; 1191; 10,417,878; 5,208,859; 968; 97.2. ***P. discolor***: DSG785; 2; 1180; 10,617,625; 5,308,819; 969; 97.3. ***P. edulis***: DSG485; 2; 989; 9,054,447; 4,527,033; 969; 97.3. ***P. edulis***: DSG1028; 1; 994; 9,061,862; 4,530,904; 969; 97.3. ***P. edulis***: DSG1020; 2; 1126; 9,546,477; 4,773,241; 969; 97.3. ***P. fallax***: DSG494; 2; 484; 4,261,767; 2,130,955; 968; 97.2. ***P. fallax***: DSG492; 2; 904; 7,384,002; 3,692,088; 969; 97.3. ***P. johannis***: DSG501; 2; 907; 7,844,117; 3,922,046; 969; 97.3. ***P. johannis***: DSG08199; 2; 1313; 11,297,858; 5,648,973; 969; 97.3. ***P. lagunae***: AGM9263; 2; 1069; 9,553,367; 4,776,613; 969; 97.3. ***P. lagunae***: AGM9279; 2; 1233; 10,814,782; 5,407,465; 969; 97.3. ***P. maximartinezii***: DSG07799; 2; 937; 8,321,810; 4,160,907; 969; 97.3. ***P. maximartinezii***: DSG1010; 2; 704; 10,502,750; 5,251,251; 969; 97.3. ***P. maximartinezii***: DSG6499; 2; 1174; 10,502,750; 5,251,251; 969; 97.3. ***P. monophylla***: RH6668; 1; 618; 5,605,098; 2,802,565; 969; 97.3. ***P. monophylla***: DSG1214; 1; 942; 8,427,317; 4,213,596; 969; 97.3. ***P. monophylla***: DSG478; 2; 999; 8,718,219; 4,359,102; 969; 97.3. ***P. monophylla***: DSG480; 2; 1241; 11,108,138; 5,554,149; 969; 97.3. ***P. pinceana***: DSG1163; 2; 909; 8,001,565; 4,000,671; 969; 97.3. ***P. pinceana***: DSG8999; 2; 1320; 11,789,410; 5,894,652; 968; 97.2. ***P. pinceana***: DSG7999; 2; 1576; 14,115,122; 7,057,552; 969; 97.3. ***P. quadrifolia:*** DSG1599; 2; 862; 7,306,710; 3,653,325; 968; 97.2. ***P. quadrifolia***: DSG1560; 1; 890; 8,193,597; 4,096,686; 968; 97.2. ***P. quadrifolia***: DSG01499; 2; 1096; 9,892,499; 4,946,163; 969; 97.3. ***P. quadrifolia***: DSG961; 2; 1195; 10,111,721; 5,055,873; 969; 97.3. ***P. quadrifolia***: DSG01099; 2; 1152; 10,445,399; 5,222,697; 968; 97.2. ***P. quadrifolia***: quad035; 2; 1336; 10,879,895; 5,440,165; 968; 97.2. ***P. remota***: DSG1301; 2; 981; 8,112,584; 4,056,345; 969; 97.3. ***P. remota***: DSG19498; 2; 972; 8,686,551; 4,343,208; 969; 97.3. ***P. remota***: DSG22498; 2; 1010; 8,887,062; 4,443,557; 969; 97.3. ***P. remota***: DSG801; 2; 1123; 9,854,338; 4,926,998; 969; 97.3. ***P. remota***: DSG23298; 2; 1117; 10,125,019; 5,062,440; 969; 97.3. ***P. rzedowskii***: DSG637; 2; 1199; 10,663,690; 5,331,879; 968; 97.2. ***P. rzedowskii***: RB47131; 2; 1219; 10,740,126; 5,370,129; 968; 97.2. ***P. rzedowskii***: DSG635; 2; 1310; 11,448,532; 5,724,214; 968; 97.2. ***P. rzedowskii***: DSG636; 2; 1287; 11,475,267; 5,737,519; 969; 97.3. **Subsection** ***Balfourianae***. ***P. aristata***: KF37; 2; 680; 6,039,708; 3,020,016; 969; 97.3. ***P. aristata***: KF30; 2; 813; 7,027,768; 3,513,851; 969; 97.3. ***P. longaeva***: DSG1027; 1; 681; 6,029,708; 3,010,016; 969; 97.3. **Subsection** ***Gerardianae***. ***P. bungeana***: 03s3A; 2; 842; 7,036,428; 3,518,309; 969; 97.3. ***P. gerardiana***: RB41105; 2; 464; 4,111,029; 2,055,501; 968; 97.2. **Subsection** ***Krempfianae***. ***P. krempfii***: PT242; 2; 864; 7,248,613; 3,624,217; 969; 97.3. **Subsection** ***Nelsoniae***. ***P. nelsonii***: DSG1096; 2; 954; 8,461,342; 4,230,564; 969; 97.3. ***P. nelsonii***: DSG10198; 2; 1141; 10,225,576; 5,112,715; 968; 97.2. ***P. nelsonii***: DSG31798; 2; 1261; 11,150,773; 5,575,312; 968; 97.2. **Subsection** ***Strobus***. ***P. lambertiana***: DSG1195; 1; 1122; 10,161,689; 5,080,591; 969; 97.3.

3.2. Coalescent-based species delimitation in North American pinyon pines using low-copy nuclear genes and plastomes

Montes, J.-R., A. Moreno-Letelier, P. Peláez, and D. S. Gernandt. (2022)

American Journal of Botany 109(5):1–21.
URL: DOI: 10.1002/ajb2.1847

Received: 27 July 2021 | Accepted: 4 March 2022

DOI: 10.1002/ajb2.1847

RESEARCH ARTICLE

American Journal of Botany

Coalescent-based species delimitation in North American pinyon pines using low-copy nuclear genes and plastomes

José-Rubén Montes[1] | **Pablo Peláez**[2] | **Alejandra Moreno-Letelier**[3] | **David S. Gernandt**[4]

[1]Posgrado en Ciencias Biológicas, Instituto de Biología, Universidad Nacional Autónoma de México, 04510, Ciudad de México, Mexico

[2]Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, 62210, Cuernavaca, Morelos, Mexico

[3]Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, 04510, Ciudad de México, Mexico

[4]Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, 04510, Ciudad de México, Mexico

Correspondence

David S. Gernandt, Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, 04510, Ciudad de México, Mexico.
Email: dgernandt@ib.unam.mx

Abstract

Premise: Accurate species delimitation is essential for evolutionary biology, conservation, and biodiversity management. We studied species delimitation in North American pinyon pines, *Pinus* subsection *Cembroides,* a natural group with high levels of incomplete lineage sorting.

Methods: We used coalescent-based methods and multivariate analyses of low-copy number nuclear genes and nearly complete high-copy number plastomes generated with the Hyb-Seq method. The three coalescent-based species delimitation methods evaluated were the Generalized Mixed Yule Coalescent (GMYC), Poisson Tree Process (PTP), and Trinomial Distribution of Triplets (Tr2). We also measured admixture in populations with possible introgression.

Results: Our results show inconsistencies among GMYC, PTP, and Tr2. The single-locus based GMYC analysis of plastid DNA recovered a higher number of species (up to 24 entities, including singleton lineages and clusters) than PTP and the multi-locus coalescent approach. The PTP analysis identified 10 species whereas Tr2 recovered 13, which agreed closely with taxonomic treatments.

Conclusions: We found that PTP and GMYC identified species with low levels of ILS and high morphological divergence (*P. maximartinezii*, *P. pinceana*, and *P. rzedowskii*). However, GMYC method oversplit species by identification of more divergent samples as singletons. Moreover, both PTP and GMYC were incapable of identifying some species that are readily identified morphologically. We suggest that the divergence times between lineages within North American pinyon pines are so disparate that GMYC results are unreliable. Results of the Tr2 method coincided well with previous delimitations based on morphology, DNA, geography, and secondary chemistry.

KEYWORDS

coalescent theory, conifers, GMYC, Hyb-Seq, incomplete lineage sorting, Pinaceae, pines, PTP, Tr2

Species are fundamental units of study in several areas of the biological sciences. The accurate delimitation of species boundaries is essential for evolutionary biology, conservation biology, and biodiversity management (Sites and Crandall, 1997; Sites and Marshall, 2003; Yang and Rannala, 2010). Species delimitation is related to species concepts, which have been long debated by biologists because there are discrepancies among species definitions (de Queiroz, 2007). Definitions are independent of methodological aspects of lineage delimitation, but criteria for inferring species boundaries have been used both for conceptualization and species delimitation. Nonetheless, species concepts exhibit an underlying unity that provides the basis for a unified concept of species as separately evolving lineages, which allows us to address the problem of delimitation more directly (de Queiroz, 2007).

Interest in integrative species delimitation has grown thanks to the increasing availability of molecular data, new approaches, and methods (Schlick-Steiner et al., 2010; Fujita et al., 2012; Zhang et al., 2013) that have been developed to test species boundaries. Many coalescent-based methods have been developed, including single-locus methods such as the Generalized Mixed Yule Coalescent (GMYC; Pons et al., 2006) and Poisson Tree Processes (PTP; Zhang et al., 2013; Kapli et al., 2017). Multi-locus species delimitation methods include Brownian processes (Brownie; O'Meara et al., 2006; O'Meara, 2010) and Species Tree Estimation under Maximum Likelihood (spedeSTEM; Ence and Carstens, 2011). Methods have also been developed for biallelic genetic markers such as Single Nucleotide Polymorphisms and Amplified Fragment Length Polymorphisms Phylogenies (SNAPP; Bryant et al., 2012), Bayesian Phylogenetics and Phylogeography (BP&P; Yang, 2015), Division of Individuals into Species using Sequences and Epsilon-Collapsed Trees (DISSECT; Jones et al., 2015), and the Trinomial Distribution method (Tr2; Fujisawa et al., 2016). Coalescent-based methods provide an attractive alternative for studying the evolutionary processes that contribute to speciation, inferring the relationships among species, and delimiting independent evolutionary lineages objectively in the presence of gene-tree conflict (Fujita et al., 2012; Smith et al., 2015, 2020; Luo et al., 2018).

In pines and other conifers, there are few studies using objective methods for the delimitation of species. *Pinus* L. classification and species delimitation initially used a small number of morphological characters. Over time additional evidence was incorporated from anatomy, reproductive biology, biochemistry, and molecular markers (Price et al., 1998; Gernandt et al., 2001, 2003, 2005; Syring et al., 2005; Parks et al., 2012; Willyard et al., 2021). Nevertheless, each source of evidence used undergoes processes that can introduce error, such as plasticity of certain morphological characters in response to the environment (e.g., number of needles per fascicle and needle length and width), transfer of genetic information among genomic compartments, low interspecific variability (e.g., plastid DNA markers), and uniparental inheritance of plastid DNA, which is susceptible to "plastid capture" (Mirov, 1967; Rieseberg and Soltis, 1991; Liston et al., 1999; Gernandt et al., 2003; Kan et al., 2007; Mort et al., 2007; Poulos and Berlyn, 2007; Cole et al., 2008; Tsutsui et al., 2009; Turna and Güney, 2009; Nobis et al., 2012; Cole et al., 2013). Lineage delimitation of trees is also difficult because of complex evolutionary histories caused by incomplete lineage sorting (ILS) and reticulation resulting from hybridization and introgression (Rosenberg, 2003; Hernández-León et al., 2013; Zhang et al., 2014).

Trees are commonly characterized by large population sizes, longevity, slower mutation and speciation rates, and longer generation times (Petit and Hampe, 2006). Studies of species boundaries have been carried out in *Populus* L. (Wang et al., 2011), *Cycas* L. (Feng et al., 2016), and *Pinus* (Moreno-Letelier et al., 2013; Zhang et al., 2014; López-Reyes et al., 2015; Willyard et al., 2017). Species delimitation has been done using clustering algorithms (Pritchard et al., 2000) and grouping individuals in populations but without evaluating the evolutionary divergence of clusters. An exception was an attempt to delimit species of North American hard pines using plastid DNA and a single-locus based coalescent-based method (Hernández-León et al., 2013).

Pinus subsection *Cembroides* Engelm. is a clade of North American pinyon pines with a fossil record that extends to the Late Oligocene (Wolfe and Schorn, 1990). The pinyon pines are restricted to arid or semi-arid environments extending from the southwestern United States to south central Mexico (Critchfield and Little, 1966). They comprise approximately 15 pine taxa of exceptional ecological importance (Lanner, 1981; Farjon and Styles, 1997). The North American pinyon pines are small to medium-sized trees or shrubs with 1 to 5 secondary leaves with deciduous fascicle sheaths, ovulate cones with a short peduncle, and seeds that are functionally wingless in all species except *P. rzedowskii* Madrigal & M. Caball. (Madrigal and Caballero, 1969; Malusa, 1992). The International Union for Conservation of Nature (IUCN, 2019) lists five pinyon pine taxa as vulnerable or endangered whereas the Mexican government lists nine as protected (SEMARNAT, 2010).

The circumscription of recognized species in *Pinus* subsect. *Cembroides* differs in recent works (Gernandt et al., 2005; Farjon and Filer, 2013; Gernandt and Pérez de la Rosa, 2014; Montes et al., 2019) and the taxonomic ranges used for pinyon pines has varied widely, due in part to placing emphasis on different molecular and structural characters (Malusa, 1992; Farjon and Styles, 1997; Gernandt et al., 2005; Farjon and Filer, 2013). Disagreements in the classification of some taxa include whether or not to elevate *P. cembroides* Zucc. subsp. *orizabensis* D.K. Bailey and *P. cembroides* subsp. *lagunae* (Robert-Passini) D.K. Bailey (here treated as *P. lagunae* (Robert-Passini) Passini), to the rank of species (Gernandt et al., 2005; Farjon and Filer, 2013; Montes et al., 2019). Similarly, phylogenetic analyses clearly support separating *P. discolor* D.K. Bailey & F.G. Hawksw. and *P. johannis* M.F. Robert from *P. cembroides* (Gernandt et al., 2003, 2005; Parks et al., 2012; Montes et al., 2019) although they have been treated as infraspecific taxa of *P. cembroides* (Farjon and Filer, 2013).

Pinus californiarum D.K. Bailey, *P. fallax* (Little) Businský, and *P. monophylla* Torr. & Frém. all have solitary needles (Malusa, 1992). *Pinus californiarum* has been treated as an independent lineage, as a synonym of *P. monophylla*, or as a variety, *P. monophylla* var. *californiarum* (D.K. Bailey) Silba (Silba, 1990; Farjon and Styles, 1997; Price et al., 1998). *Pinus californiarum* was recovered as sister to *P. monophylla* and *P. quadrifolia* Parl. ex Sudw. in a previous phylogenetic study of low-copy nuclear DNA, suggesting that it could be considered as a valid species rather than as an infraspecific taxon of *P. monophylla* (Montes et al., 2019). *Pinus fallax* and *P. californiarum* may also be valid species rather than infraspecific taxa of *P. monophylla* or *P. edulis* Engelm. (Montes et al., 2019). *Pinus fallax* was originally described

as *P. edulis* var. *fallax* Little (1968) but has been treated as *P. californiarum* subsp. *fallax* (Little) D.K. Bailey, or *P. monophylla* subsp. *fallax* (Little) Silba (Farjon and Styles, 1997; Cole et al., 2008; Farjon and Filer, 2013). *Pinus fallax* occurs in environmental conditions with moderate summer rainfall (Malusa, 1992; Cole et al., 2008) unlike *P. monophylla* and *P. californiarum*, which inhabit places with dry summers (Cole et al., 2008).

Pinus subsect. *Cembroides* offers an opportunity to study the boundaries among species that have evolved few morphological differences (Price et al., 1998; Gernandt et al., 2008) and species with clear morphological divergences and presumably relatively deep divergences. In this study, our aims were to: (1) infer the species boundaries in *Pinus* subsect. *Cembroides* using coalescent-based methods; (2) compare species delimitation hypotheses of single and multi-locus coalescent methods; (3) reexamine the taxonomic validity of some taxa in the light of multi-locus data; and (4) study admixture in three subgroups: (a) *Pinus cembroides* subsp. *cembroides*, *P. cembroides* subsp. *orizabensis*, and *P. lagunae*; (b) *P. johannis* and *P. discolor*; and (c) *P. californiarum*, *P. fallax*, and *P. monophylla*.

MATERIALS AND METHODS

Sampling

We included 3–8 individuals per species from material deposited in the Herbario Nacional de México (MEXU) and from field collections (Appendix S1). Ninety-three individuals were sampled, including 80 corresponding to subsect. *Cembroides*, four to subsect. *Balfourianae* Engelm., and three to subsect. *Nelsoniae* Burgh. These three subsections comprise section *Parrya* Mayr of subgenus *Strobus* Lemmon (Gernandt et al., 2005). From sect. *Quinquefoliae* Duhamel we included two individuals of subsect. *Gerardianae* Loudon, one of subsect. *Krempfianae* Little and Critchfield, and one of subsect. *Strobus* Loudon.

We extracted DNA from haploid seed megagametophyte of 10 individuals using a Wizard genomic DNA purification kit (Promega, Madison, Wisconsin, USA) and from diploid leaf tissue of 83 individuals using the CTAB method (Doyle and Doyle, 1987).

Illumina library preparation, probe design, and Hyb-Seq sequencing

We used 500 ng of total DNA per sample to prepare the genomic libraries. DNA fragments of ca. 250 bp were size-selected with a bioruptor sonicator and the length distribution of fragments was evaluated by automated electrophoresis using a 2100 Bioanalyzer System (Agilent, Santa Clara, California, USA). Barcode adapters were ligated for Illumina sequencing using a NEBNext library prep kit for three samples (*P. lagunae* BS3-BS5) and a TruSeq library prep kit for all other samples (Illumina, San Diego, California, USA).

Pools of 24 samples were enriched for nuclear targets with MYbaits version 2.3.1 biotinylated RNA baits (Arbor Biosciences, Ann Arbor, Michigan, USA) following the manufacturer's protocol. Probes were designed for 1045 putative low-copy nuclear genes from *Pinus taeda* L. (Willyard et al., 2007; Neves et al., 2013; Gernandt et al., 2018; see Appendix S2 for more details). The samples were spread across eight sequencing runs (Appendix S3) that also included Pinaceae samples for other studies. Different mixtures of enriched and unenriched libraries were used for successive runs, according to recovery of plastid sequences in prior runs. We combined samples into three different multiplex sets (Appendix S3) to sequence on a single lane each of an Illumina Hi-Seq. 2500 or Hi-Seq. 4000 using the 100, 125, or 150 bp modules with paired reads (Appendix S3).

Processing of Hyb-Seq data

Illumina reads were demultiplexed based on their barcode and processed with Trimmomatic version 0.32 (Bolger et al., 2014) using the parameters for paired end reads suggested by the authors (Appendix S2). Trimmed reads were assembled with the HybPiper version 1.2 pipeline (Johnson et al., 2016). This pipeline first performs read sorting with the BWA method (Li and Durbin, 2009) using the nuclear gene sequences from the probe design step as references and then assembles each gene using SPAdes version 3.10.1 (Bankevich et al., 2012). Assembled gene files (without introns) were imported into Geneious version R11.0.5 (Kearse et al., 2012) and aligned with MAFFT version 7.0 (Katoh et al., 2002). The genes were filtered under five ad hoc exclusion criteria: (1) missing sequence for one or more samples (341); (2) pairwise identity less than 93% (168); (3) fewer than 50% of identical sites (67); (4) genes detected as possible paralogs (164); and (5) genes with anomalously high substitution rates based on profiling phylogenetic informativeness (22) (Appendix S2). The first three criteria were applied in Geneious and the fourth with the HybPiper script paralog_investigator.py (Johnson et al., 2016). The paralog script identified contigs with lengths ≥85% of the reference sequence, indicating multiple long-length matches.

We estimated the informativeness of characters with the PhyDesign web application (Townsend, 2007; López-Giráldez and Townsend, 2011). The input files for PhyDesign were the concatenated alignment with 229 gene partitions (those remaining after filtering) and the ultrametric tree. Maximum likelihood trees were inferred in RAxML version 8.2.10 (Stamatakis, 2014) under the general time reversible model with the gamma parameter (GTR + G) and 1000 bootstrap searches. The trees were ultrametricized with clock-based likelihood in PAUP* version 4.0a150 (Swofford, 2002) using the HKY85 substitution model and "Thorne" parameterization for

clock optimization. Genes with unusually high substitution rates, resulting in recent illusory spikes in the informativeness plot were eliminated (Appendix S4).

We used the modified protocol by Aguirre-Dugua and Gernandt (2017) to assemble plastomes. We removed duplicated reads in Geneious and trimmed low-quality bases with Trimmomatic. De novo assembly was then performed on the sequences that mapped to the reference using SPAdes. The resulting scaffolds were imported into Geneious, eliminating those that were <500 bp in length. We mapped the scaffolds to the consensus sequence that was previously generated in the mapping step and extracted a new consensus sequence. Finally, we remapped the reads to the consensus sequence to produce a final consensus. From the assembled plastomes, we chose those with the highest coverage for phylogenetic analyses (Appendix S5).

The plastome sequences were aligned with MAFFT in Geneious. Poorly aligned or divergent regions (with elevated numbers of differences, insertions, and deletions) were deleted using the Gblocks webserver version 0.91b (Castresana, 2000) with the following options: (1) smaller final blocks; (2) gap position within the final blocks; and (3) less strict flanking positions.

Phylogenetic analysis

We performed a maximum likelihood analysis of the 207 nuclear genes and 93 terminals in RAxML (Stamatakis, 2014). We performed heuristic searches with the -#autoMREoption, which automatically determines the sufficient number of bootstrap replicates (bs) and a GTR + G model to calculate the heterogeneity of rates for each of the multiple alignments. We also concatenated the 207 gene alignments and performed a maximum likelihood analysis in RAxML with 1000 heuristic searches and the GTR + G model. The plastome sequences were analyzed in RAxML with 1000 heuristic searches on the alignment partitioned into coding and noncoding regions and applying the GTR + G model to each.

We also performed Bayesian inference using MrBayes version 3.2.7 (Huelsenbeck and Ronquist, 2001) on the plastid DNA alignment with partition blocks for coding and noncoding regions. The nucleotide substitution model was chosen using the Akaike Information Criterion test (AIC) in jModelTest version 2.1.10 (Miller et al., 2010; Darriba et al., 2012). The analysis was conducted using the GTR model allowing both invariant sites and rate heterogeneity (I + G). The analysis was run using three heated chains and one cold chain, and a heating of 0.2. Two independent runs of 40,000,000 generations were performed with sampling every 1000 generations, discarding 0.25 as a burn-in fraction. We used Tracer version 1.7.1. (Rambaut et al., 2018) to corroborate chain convergence. Tree topologies were summarized in a 50% majority rule tree and the consensus tree was imported into FigTree version 1.4.0 for further editing (Rambaut, 2012).

Identification of single nucleotide polymorphisms

To provide an alternative method for assessing variable sites for the species delimitation analyses, SNP calling was performed through the alignment of the Hyb-Seq data to the *Pinus taeda* genome. SNPs were initially called from all 93 samples. Quality of demultiplexed sequence reads was assessed with FastQC version 0.11.7 and MultiQC version 1.5 (Andrews, 2010; Ewels et al., 2016). Sequence quality trimming and adapter removal were performed using Trimmomatic with default parameters (Bolger et al., 2014). Paired cleaned reads were mapped against the *P. taeda* genome version 2.01 using BWA-MEM (Li, 2013 [Preprint]; Neale et al., 2014; Wegrzyn et al., 2014). BAM files were sorted and uniquely mapped reads were extracted with the sort and view routines of SAMtools version 0.1.19 (Li et al., 2009). The Picard tool MarkDuplicates version 2.5.0 was used to discard duplicate reads (website: http://broadinstitute.github.io/picard). SAMtools mpileup was used to call variants with the following parameters: (1) biallelic variants only; (2) no-BAQ; (3) minimum mapping quality of 20; and (4) minimum base quality of 25 (Li, 2011). A first step filter was applied with VCFtools version 0.1.13 to keep variants genotyped in 50% of the samples, having a minimum quality score of 25, a minor allele count less than three, and a minimum mean depth of three reads (Danecek et al., 2011). In a second filtering step, SNPs were removed if genotypes were not present across all samples (100%), minimum mean depth was below 10, and minimum quality score was below 30. With respect to the SNPs called for the *P. cembroides* complex (17 samples), the *P. johannis-discolor* complex (11 samples), and the *P. californiarum-fallax-monophylla* complex (27 samples), samples were extracted separately after the first step filter with VCFtools and then subjected to the second filter. The variant calling format (VCF) file containing all 93 samples was converted into a tab-delimited text file, and subsequently SNPs were concatenated into a FASTA file including heterozygous sites. A multiple sequence alignment was generated with MAFFT.

Lineage tree estimation

Lineage tree inference was performed with the coalescent method ASTRAL-III version 5.7.3 (Mirarab and Warnow, 2015; Zhang et al., 2018). We used 207 nuclear gene trees with 93 terminals estimated previously with RAxML as input for ASTRAL-III. Branch lengths, coalescent units, and local posterior probabilities (lpp) were estimated with the lineage tree. We also explored gene conflict using gene-concordance factors (gCFs) in IQ-TREE-2 version 2.1.2 (Minh et al., 2020).

Tree inference based on SNPs was performed with the coalescent method SVDquartets (Chifman and Kubatko, 2014) in PAUP* (Swofford, 2002). SVDquartets performs

well even when there is variation in effective population sizes, presence of ILS, and gene flow (Long and Kubatko, 2018, 2019). It infers unrooted trees from quartets based on multi-locus and unlinked SNP data (Chifman and Kubatko, 2014). For this method we used a NEXUS input file that included 26,180 SNPs and 93 terminals. All possible quartets were analyzed and branch support was estimated with 1000 bootstrap replicates.

Coalescent-based species delimitation

Species boundaries were inferred under the coalescent framework with both low-copy nuclear genes and plastid DNA. For plastomes, we employed the Generalized Mixed Yule Coalescent (GMYC) using single and multiple thresholds models (sGMYC; Pons et al., 2006; mGMYC; Monaghan et al., 2009), and the Poisson Tree Processes method (PTP; Zhang et al., 2013; Kapli et al., 2017). The rooted triplets method (Tr2; Fujisawa et al., 2016) was used to perform species delimitation with the nuclear genes.

The GMYC was designed to test species delimitation with single-locus data. The method distinguishes between Yule (branching events inter-specific) and coalescence processes (branching events intraspecific) based on the difference in branching rates across all species in the phylogeny (Pons et al., 2006). A likelihood ratio test (LRt) is used to assess the timing of branching events from a null model (same species) and an alternative model (different species). The likelihood score from a chi-square test allows detecting significant changes between branching events inter-specific and branching events intraspecific (Pons et al., 2006). We used the ultrametric plastid tree with branch lengths estimated using maximum likelihood as input and performed the analyses in RStudio version 1.2.1335 (RStudio team, 2019; website: http://www.rstudio.com) using the "*splits*" package.

PTP uses two functions (speciation/coalescence) for modeling the transition point (node) between inter-specific and intraspecific branching events in a phylogeny (Zhang et al., 2013; Kapli et al., 2017). PTP uses an exponential distribution that represents the number of accumulated substitutions (k) for speciation events (n) (Kapli et al., 2017) and assumes that speciation and coalescence rates are different. PTP implements Markovian chains to assess delimitation support on a phylogenetic tree (Kapli et al., 2017). We used the (non-ultrametric) tree estimated from the plastid DNA alignment with MrBayes as input for PTP version 0.51. The bPTP analysis was run for 500,000 generations (as recommended by the authors), with a thinning of 100, and discarding 0.1 generations as a burn-in fraction.

Species delimitation was also performed with Tr2 in Python version 2.7 (Fujisawa et al., 2016). Tr2 uses Bayesian model comparison and reduces the likelihood calculations because phylogenies are decomposed into rooted triplet topologies. The method explores the best delimitation model from a guide tree and multi-locus data. Posterior probability is used to find the best delimitation model from a set of possible hypotheses. The best model is the one with a posterior probability close to 0 (Fujisawa et al., 2016). Tr2 estimates a null model (without a priori assignment of individuals to species) and allows alternate assignment of individuals to species to test among hypotheses. The 207 maximum likelihood gene trees obtained in RAxML were the input for Tr2. We compared the likelihood scores of five alternative hypotheses based on taxonomic classifications (Appendix S6).

Identifying genetic clustering from SNPs

We identified genetic clusters from three subsets of SNPs using a principal components analysis (PCA) and discriminant analysis of principal components (DAPC). Analyses were performed separately for subsets of taxa from three species complexes. The first subset corresponded to *P. discolor* and *P. johannis* ($K = 2$), the second to *P. cembroides* subsp. *cembroides*, *P. cembroides* subsp. *orizabensis*, and *P. lagunae* ($K = 2$–3), and the third to *P. californiarum*, *P. fallax*, *P. monophylla*, *P. edulis*, and *P. quadrifolia* ($K = 4$–5); this clade was called "one-needle pines + *P. edulis* + *P. quadrifolia*". Both PCA and DAPC analyses were carried out following the code by Grünwald et al. (2016) in RStudio with the "*ape*" version 5.6 (Paradis and Schliep, 2019) and "*poppr*" version 2.8.5 packages (Kamvar et al., 2014). The first two components were used for plotting using *ggplot2* package version 3.2.2 (Villanueva et al., 2016). The DAPC was performed to maximize the discrimination between groups with the same parameters as in the PCA (see Grünwald et al., 2016). We assigned the samples a priori to each species and evaluated the species assignments based on the results of the PCA. We illustrated the probability of population membership and assigned probability of populations membership.

Admixture analysis

To estimate the genetic admixture proportions within the three sample complexes the VCF files were converted to ordinary PLINK files using PLINK version 1.9 (Purcell et al., 2007). The optimal K value (evaluated from 2-10) and the admixture proportions (Q-values) up to $K = 6$ were obtained with ADMIXTURE version 1.3.0 (Alexander et al., 2009). The Clumpak program version 1.1 was used for the clustering and plotting of Q-matrices (Kopelman et al., 2015).

RESULTS

Pre-processing and processing of data

For nuclear DNA, paired reads from 99 individuals were assembled to 996 reference genes in HybPiper. The average

number of reads mapped to the references was 5,064,027. The number of genes recovered was 969 with a mean coverage of 249.3×. We retained 207 genes after filtering (see Materials and Methods).

For plastid DNA, we assembled 82 plastomes, 69 corresponding to subsect. *Cembroides* and 13 to close relatives. The mean number of reads per sample was 8,985,593 and the mean number of reads that mapped to the plastid reference was 86,426. The plastomes had a mean length of 117,610 bp and a mean coverage of 73.5×, ranging from 8.5 to 1925×. From the 82 assembled plastomes, we chose 59 with a coverage of 20× or higher and with a mean coverage of 102.2× (Appendix S5). The three samples with the highest coverage (*P. aristata* Engelm. AZ2; 1926.2×, *P. bungeana* Zucc. ex Engelm. CN; 1098.7×, and *P. cembroides* subsp. *orizabensis* PL; 437.2×) were sequenced with version 1 of the probes, which included two plastome regions. This caused a spike in total coverage at two places in the plastome.

Phylogenetic analyses

We assembled a concatenated nuclear alignment with 140,845 bp and 11,281 informative sites. The best ML tree with partitions by gene (Appendix S7) agreed in topology with the best tree without partitions (Appendix S8), except for minor differences in relationships among different samples of the same species and bootstrap values. *Pinus* subsection *Cembroides* was monophyletic, and samples for six species were recovered as monophyletic lineages (Appendix S7).

The plastid DNA alignment was 117,210 bp after removal of ambiguously aligned sites. It included 3378 informative characters and 1884 variable but parsimony uninformative characters. The Bayesian inference analysis of the plastid DNA alignment recovered the monophyly of subsects. *Cembroides*, *Balfourianae*, and *Gerardianae*. In subsect. *Cembroides*, *P. cembroides* subsp. *cembroides*, *P. maximartinezii*, *P. monophylla*, *P. quadrifolia*, *P. remota* (Little) D.K. Bailey & F.G. Hawksw., and *P. rzedowskii* were recovered as monophyletic lineages. Of the taxa recovered as monophyletic, only *P. remota* had an incongruent phylogenetic position between the nuclear and plastid trees (Appendices S7 and S9).

Lineage tree estimation

The SVDquartets and ASTRAL analyses recovered *Pinus* subsection *Cembroides* as monophyletic and a greater number of taxa as monophyletic lineages than were recovered with analyses of plastid DNA and concatenated nuclear genes. In both coalescent trees, *P. discolor*, *P. culminicola* Andresen & Beaman, *P. johannis*, *P. lagunae*, *P. maximartinezii*, *P. monophylla*, *P. pinceana* Gordon, *P. quadrifolia*, *P. remota*, and *P. rzedowskii* were recovered as exclusive lineages, whereas *P. cembroides* subsp. *cembroides*, *P. cembroides* subsp. *orizabensis*, *P. californiarum*, *P. edulis*, and *P. fallax* were nonmonophyletic (Figures 1 and 2). The majority of monophyletic lineages recovered in both SVDquartets and ASTRAL received high support (>80% bs and 0.8 lpp, respectively) except for *P. discolor* and *P. johannis* in the ASTRAL tree (Figure 2). The interrelationships in the trees based on SNPs and low-copy nuclear genes were different in small-cone species but identical in big-cone species (Figures 1 and 2). The level of ILS in the ASTRAL species tree was very high (0.5), indicating that only 50% of the total (331,985,270) quartets estimated from gene trees agree with the lineage tree. The local posterior probabilities for a third of branches were >0.8 and coalescent branch lengths in the ASTRAL tree were short for most relationships in the small-cone clade with an exception in *P. culminicola* (Figure 2). The percent of gene discordance showed high levels of conflict in all branches within subsect. *Cembroides*. In some branches, no gene tree agreed (Figure 2).

Coalescent-based species delimitation

From the ultrametric tree inferred with 59 plastid genomes, we performed delimitation analyses with both single and multiple threshold GMYC models. Both analyses were congruent in all independent coalescent groups identified, except in the "D", "E", and "I" groups (Figure 3). The number of species (clusters and singletons) described below considers only subsect. *Cembroides*. Single threshold GMYC identified 21 species (P = 0.05). The single threshold coalescent model exhibited significantly better fit over the null model ($\log L_{GMYC} = -529.7634$, $\log L_{null} = -524.9513$, LRt = 9.624205, P = 0.008130748**) (Appendix S10). Multiple threshold GMYC recovered four speciation-coalescence transition events (Appendix S11) and identified 24 species (P = 0.05). The multiple threshold model exhibited significantly better fit than the null model ($\log L_{GMYC} = -530.1472$, $\log L_{null} = -524.9513$, LRt = 10.39178, P = 0.005539288**). A comparison between single and multiple GMYC models revealed that they were not significantly different from each other ($X^2 = 0.1918$, $P_{0.05} = 0.661437$). From 21 species identified by the single threshold GMYC model, only 5 taxa from subsect. *Cembroides* have been treated as separate species based on morphological and molecular evidence (Gernandt et al., 2005; Montes et al., 2019). These taxa (*P. monophylla*, *P. maximartinezii*, *P. quadrifolia*, *P. remota*, and *P. rzedowskii*) also were exclusive lineages in the Bayesian inference tree (Appendix S9) whereas with multiple threshold GMYC only 3 taxa identified in the subsect. *Cembroides* clade correspond to recognized species (*P. monophylla*, *P. maximartinezii*, and *P. remota*). Moreover, multiple threshold GMYC separated individuals of *P. quadrifolia* and *P. rzedowskii* into multiple species. Others were treated as a single entity with both single and multiple threshold models, including the *P. cembroides* subsp. *cembroides* + *P. cembroides* subsp. *orizabensis* + *P. lagunae* clade and the *P. californiarum* + *P. fallax* + *P. edulis* clade.

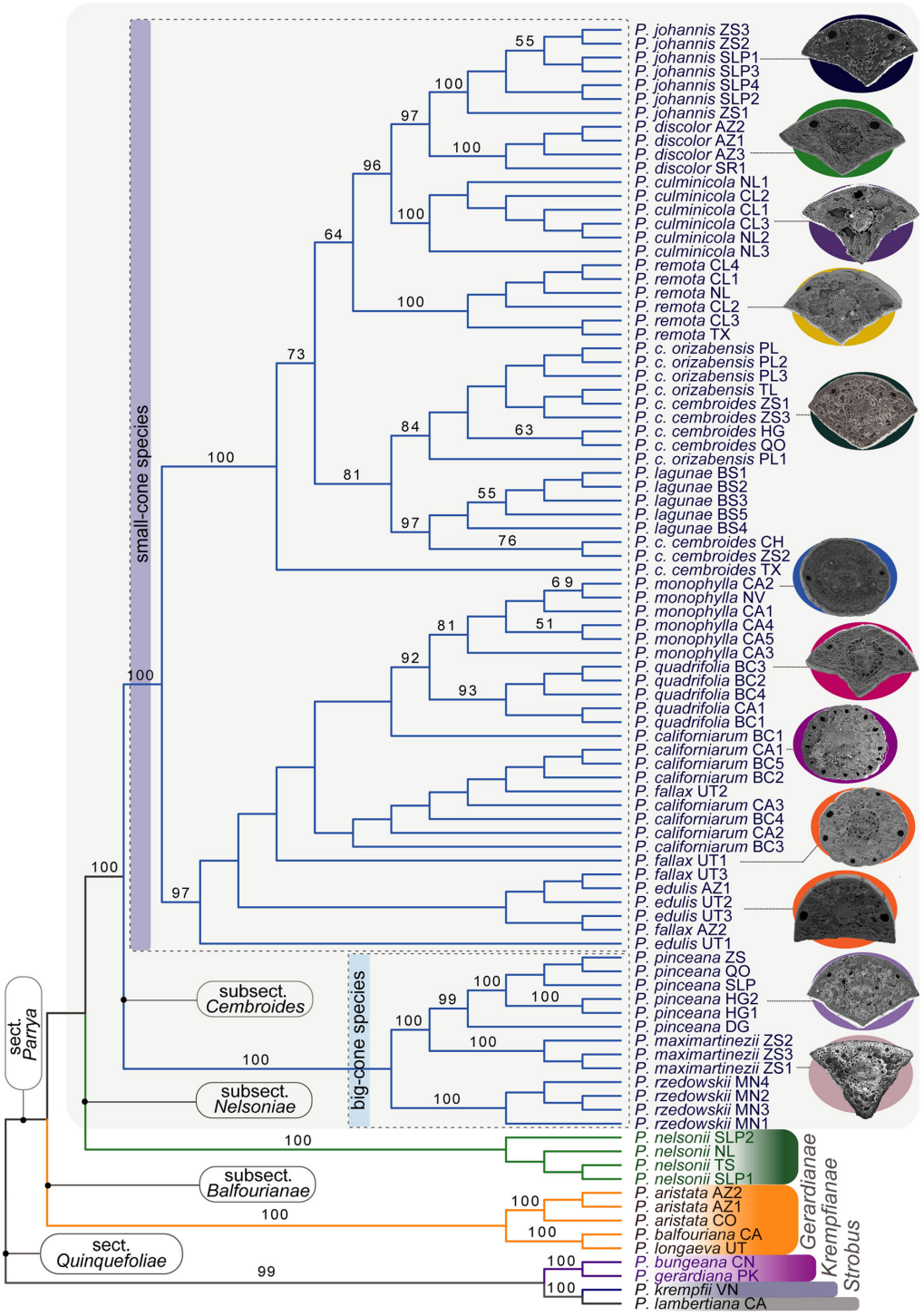


FIGURE 1 SVDquartets lineage tree based on SNPs. The small-cone and big-cone clades of *Pinus* subsection *Cembroides* are indicated. Transverse sections of needles by SEM show the needle shape, number of resin canals, and other internal structures. Bootstrap values >50% are shown on branches.

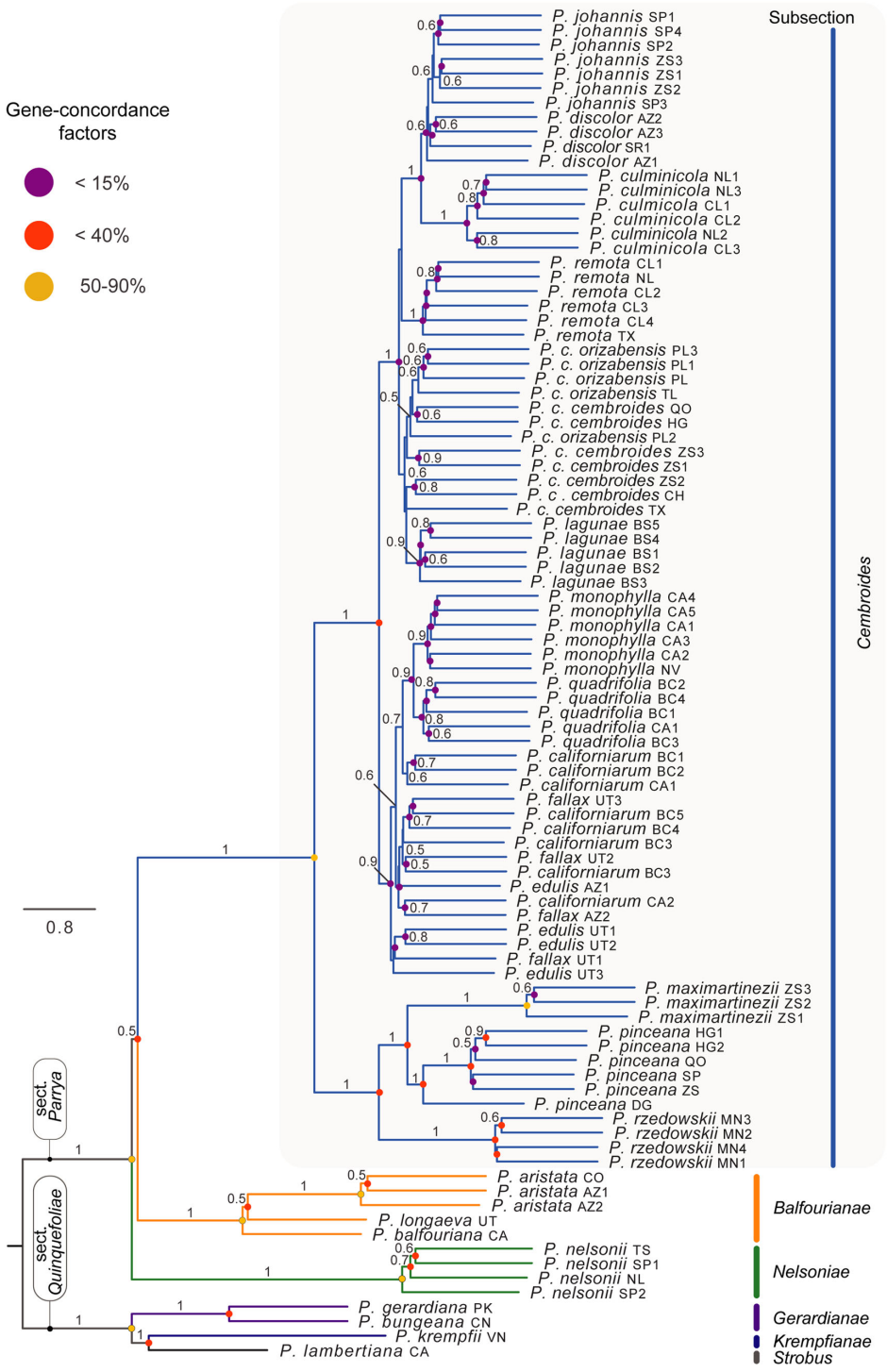


FIGURE 2 ASTRAL lineage tree based on low-copy nuclear genes. Results were inferred based on 207 trees inferred with RAxML. Branch lengths represent coalescent units. Local posterior probability values ≥0.5 are shown on branches. The gene conflict using gene concordance factors is shown on nodes with colored circles.

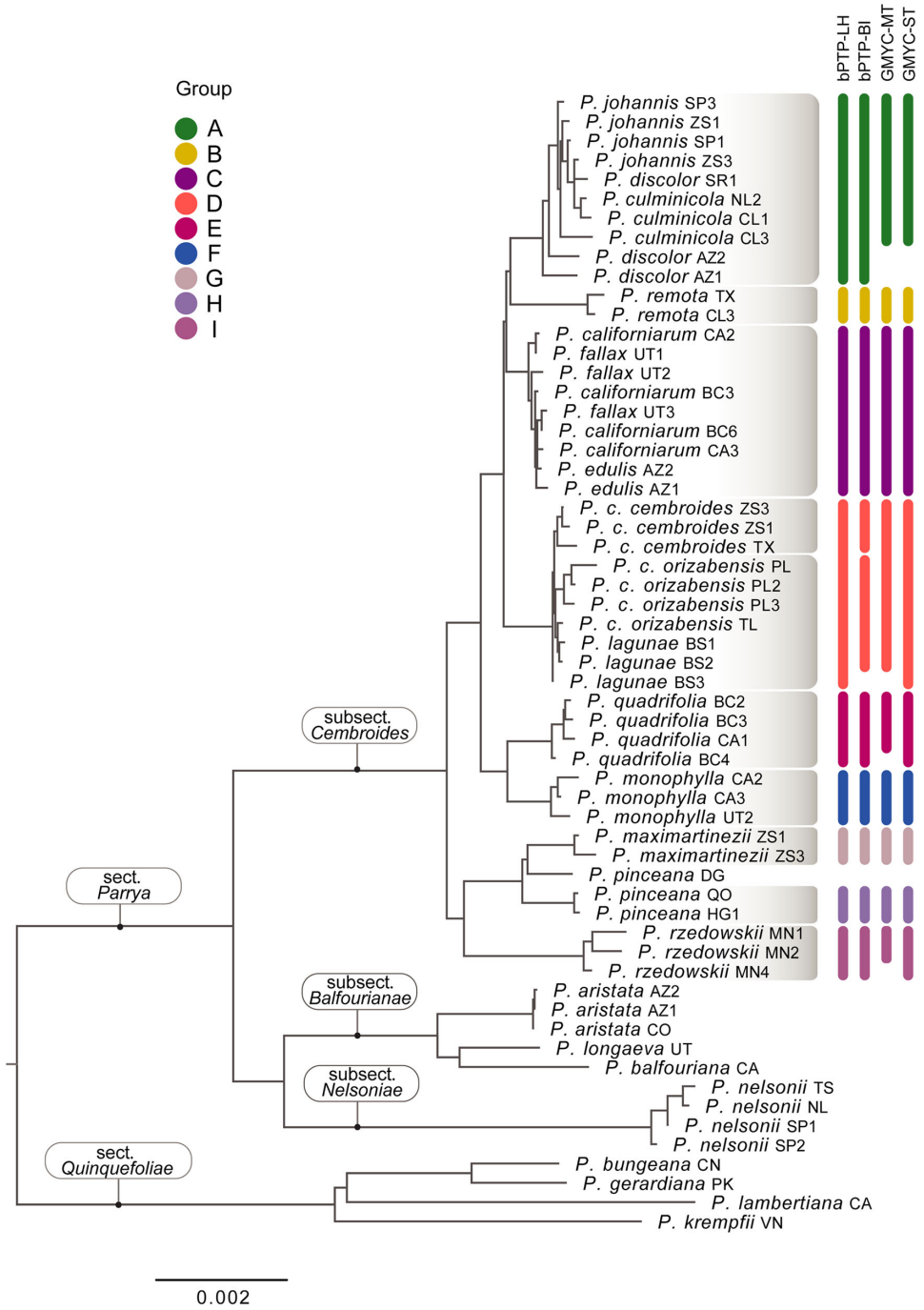


FIGURE 3 Single-locus, coalescent-based species delimitation. Results presented are based on the plastome tree resulting from Bayesian inference (BI). Vertical bars corresponding to each lineage are potential species. Colored vertical bars represent delimited clusters between both bPTP and GMYC methods. The first vertical bars correspond to bPTP with maximum likelihood estimation, the second to bPTP with Bayesian inference, the third to GMYC with multiple thresholds, and the final vertical bars indicate clusters recovered by GMYC with a single threshold.

Poisson tree processes estimated 9 to 10 species within subsect. *Cembroides* (Figure 3). Bayesian and maximum likelihood solutions agreed in all independent coalescent groups identified with the exception of the monophyletic *P. cembroides* group (Figure 3). The number of independent coalescent groups estimated with a moderately to well-supported partition in both analyses represents 50% of all taxa ($Acceptance_{rate} = 0.53$). The likelihood plot indicated convergence of Markovian chains (Appendix S12). The Bayesian solution of PTP estimated more well-supported species than the maximum likelihood solution (Appendices S12 and S13). The bPTP-BI analysis estimated 10 species within subsect. *Cembroides*, and the bPTP-ML analysis estimated 9 (Figure 3). Of 10 species estimated by bPTP-BI, only *P. cembroides* subsp. *cembroides*, *P. monophylla*, *P. maximartinezii*, *P. quadrifolia*, *P. remota*, and *P. rzedowskii* were monophyletic lineages in the Bayesian inference tree (Appendix S9). As with GMYC, bPTP solutions identified *P. californiarum*, *P. fallax*, and *P. edulis* as a single species (Figure 3) and all individuals from the *P. discolor* + *P. johannis* + *P. culminicola* clade as a single species. Based on the bPTP-BI solution, *P. cembroides* is divided into two taxa: *P. cembroides* subsp. *cembroides* and *P. lagunae* + *P. cembroides* subsp. *orizabensis*. However, the bPTP-BI solution did not identify all individuals of *P. lagunae* as part of the same species (*P. lagunae* BS3). This result was not supported by the bPTP-ML solution (Figure 3).

Based on the guide tree obtained with SVDquartets, five different hypotheses were tested with Tr2. The alternative model (Ha = 66208.32) was better than the null model (Ho = 9289676.36). Based on the null hypothesis, Tr2 estimated 8 species in subsect. *Cembroides* (Figure 4) corresponding to *P. culminicola*, *P. maximartinezii*, *P. pinceana*, *P. remota*, *P. rzedowskii*, the *P. discolor* + *P. johannis* clade, *P. cembroides* (together with *P. lagunae* and *P. cembroides* subsp. *orizabensis*), and the one-needle pines + *P. edulis* + *P. quadrifolia*. The best delimitation model was H_3, which identified thirteen species. Ten belonged to the small-cone pinyons clade: *Pinus californiarum*, *P. culminicola*, *P. cembroides* subsp. *cembroides* + *P. cembroides* subsp. *orizabensis*, *P. discolor*, *P. edulis* + *P. fallax*, *P. johannis*, *P. lagunae*, *P. monophylla*, *P. quadrifolia*, and *P. remota*, and three to the large-cone pinyons: *P. maximartinezii*, *P. pinceana*, and *P. rzedowskii*. *Pinus fallax* and *P. cembroides* subsp. *orizabensis* were not identified as independent evolutionary lineages or species. Particularly, *P. fallax* was not recovered as a monophyletic lineage because two samples (UT3, AZ2) grouped with *P. edulis* (Figure 4). These individuals of *P. fallax* had one and two needles on the same tree. Another two samples (UT1, UT2) grouped with *P. californiarum* (Figure 4). All individuals of this latter cluster had predominantly solitary needles but both samples of *P. fallax* also had one and two needles on the same tree.

Identifying genetic clustering from SNP subsets

The first subset of SNPs for the *P. cembroides* complex had 22,500 variants. The first three principal components (PCs) represent 33.1% of the total variation, 16%, 11%, and 6.1%, respectively (Figure 5). The $K = 2$ analysis identified and separated two different clusters, one comprising all individuals of *P. cembroides* subsp. *cembroides* and *P. cembroides* subsp. *orizabensis* and the second comprising individuals of *P. lagunae*. The PC1 had low and positive values, mainly for *P. cembroides* subsp. *cembroides* and *P. cembroides* subsp. *orizabensis*. The PC2 differentiated *P. lagunae* from the other taxa. The second set of SNPs for *P. johannis* and *P. discolor* had 29,601 variants. Similarly, the first three principal components represent 42% of the total variation, 17%, 14%, and 11%, respectively (Figure 5). The $K = 2$ analysis differentiated two clusters, one comprising all individuals of *P. johannis* and the other all individuals of *P. discolor*. The PC1 had negative values for *P. discolor*. The PC2 had mainly positive values and differentiated *P. johannis* from *P. discolor*. The last set of SNPs for solitary needle species, *P. edulis* and *P. quadrifolia* had 30,666 variants. The first three principal components represent 28.8% of the total variation, 17%, 6%, and 5.8%, respectively (Figure 5). The $K = 4$ analysis differentiated four clusters, one composed of individuals of *P. monophylla*, a second of individuals of *P. quadrifolia*, which is nested in *P. monophylla*, a third of individuals of *P. edulis* and only one sample of *P. fallax* (UT1) from Utah, and the last cluster is composed of individuals of *P. californiarum* and the remaining *P. fallax* samples from Utah and Arizona (UT2, UT3, AZ2). The PC1 had positive values mainly for *P. californiarum* and *P. fallax* UT2, UT3, and AZ2, whereas *P. monophylla* and *P. quadrifolia* had negative values. In the PC2 both *P. monophylla* and *P. quadrifolia* were differentiated from the other species. The analysis also separated *P. edulis* and *P. fallax* UT1 from a cluster of all other species. We performed three discriminant analyses of principal components (DAPC) retaining the same principal components of the PCA (3 PCs). The DAPC results coincide with the PCA (Figure 5). Eight clusters were differentiated: (I) *Pinus cembroides* subsp. *cembroides* and *P. cembroides* subsp. *orizabensis* distributed from southwestern USA and north to south-central Mexico; (II) *P. lagunae*, endemic to Baja California Sur, Mexico; (III) *P. discolor* distributed in southwestern USA (Arizona and New Mexico), northwestern Mexico and southern San Luis Potosí; (IV) *P. johannis* distributed in the Sierra Madre Oriental; (V) *P. californiarum*, *P. fallax* UT2, UT3, and AZ2 are distributed in southwestern USA and Baja California, Mexico; (VI) *P. edulis* and *P. fallax* (UT1) distributed in southwestern USA; (VII) *P. monophylla* distributed in southwestern USA; and (VIII) *P. quadrifolia* distributed in southwestern USA (California) and Baja California, Mexico. All individuals of this last cluster, including the only sample of *P. fallax,* share two needles per fascicle, and. All individuals of this last cluster have predominantly solitary needles but two samples of *P. fallax* had both solitary and two-needle fascicles. The posterior membership probability calculated for all clusters was high (>95%), indicating a correct assignment of the individuals to species (Figure 5).

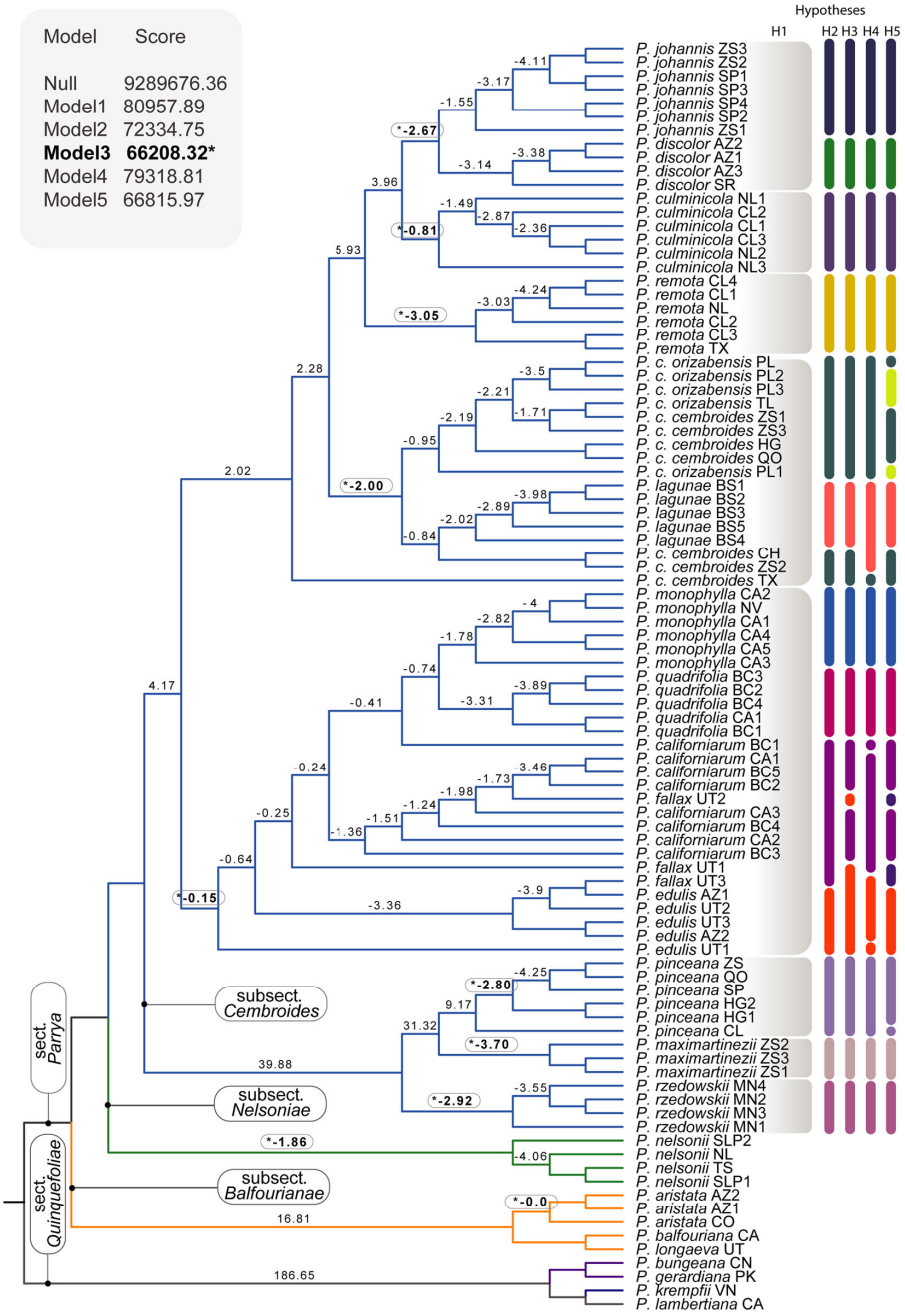


FIGURE 4 Multi-locus coalescent-based species delimitation scenarios. Results presented are based on the SNPs tree resulting from SVDquartets. Vertical bars correspond to each lineage recovered as a potential species. Colored vertical bars represent different hypotheses tested (H#). The gray boxes correspond to the null hypothesis. The best model of delimitation is hypothesis three. The numbers on the branches indicate average differences of posterior probability scores. "*–" indicates the best delimitation according to the null model. Positive values = between-species branches, negative values = within species. Values without symbols do not have enough samples to split.

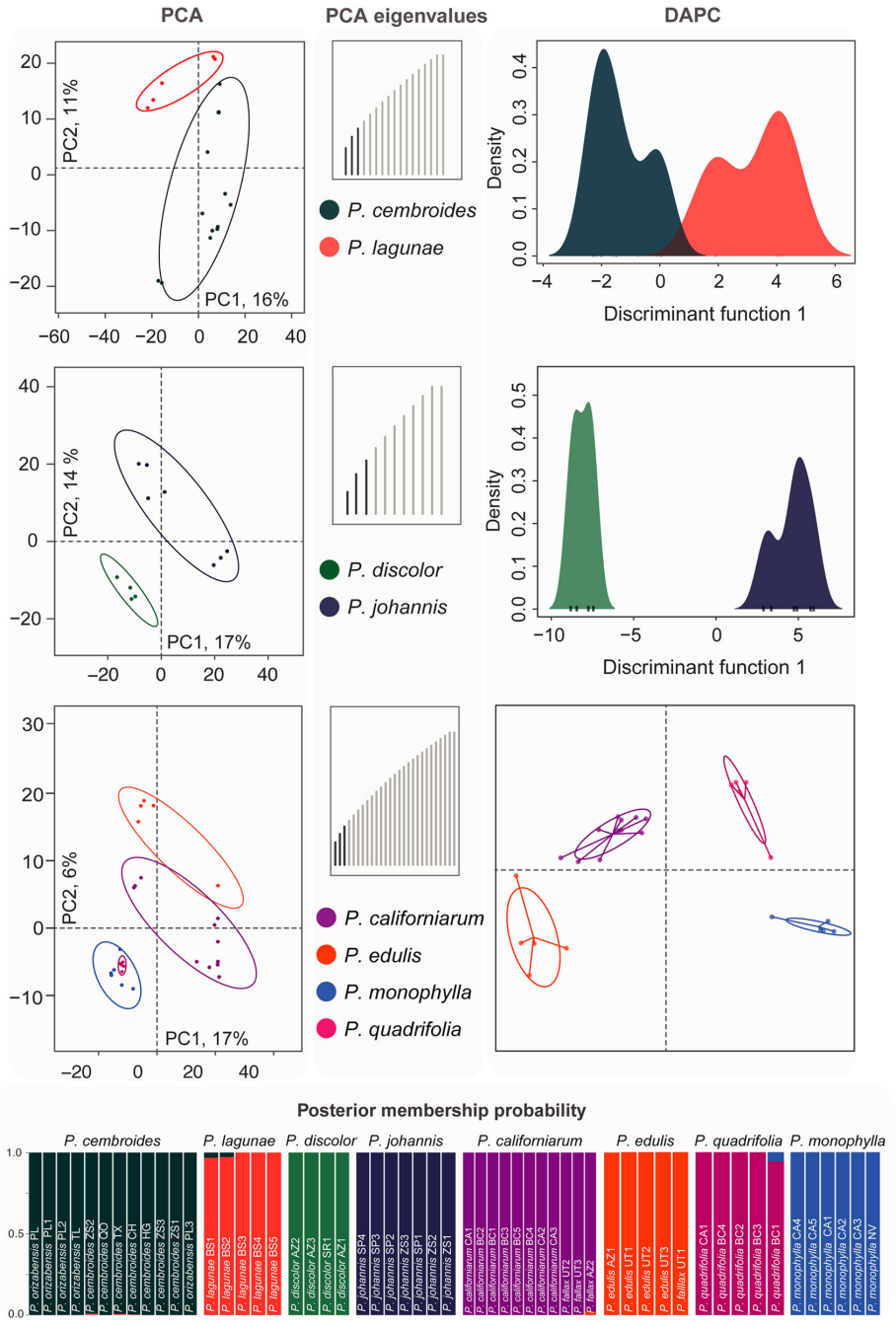


FIGURE 5 PCA and DAPC plots based on SNPs. Plots show PC1 and PC2 for all species and specimens. Each species used in PCA and DAPC graphs is represented by a different color. Specimens and groups of individuals of the same species in DAPC are enclosed by ellipses that included 95% of the data for each group. Bar graphs depict the percentage variance of significant PCs (eigenvalues). The colored bar-graphs represent the posterior membership probability of specimens to species.

Admixture analysis

In total, 29,506 SNPs were obtained for the *P. johannis* and *P. discolor* group, 22,454 for the *P. cembroides* group, and 30,609 for the one-needle pines, *P. edulis*, and *P. quadrifolia* group. The average mean depth of the SNPs for the three sets was 200×. For the three groups analyzed, the optimal clustering was for two subpopulations ($K = 2$); however, a cluster range of 2 to 6 was evaluated to explore the dynamics of the classifications (Figure 6). Overall, the two

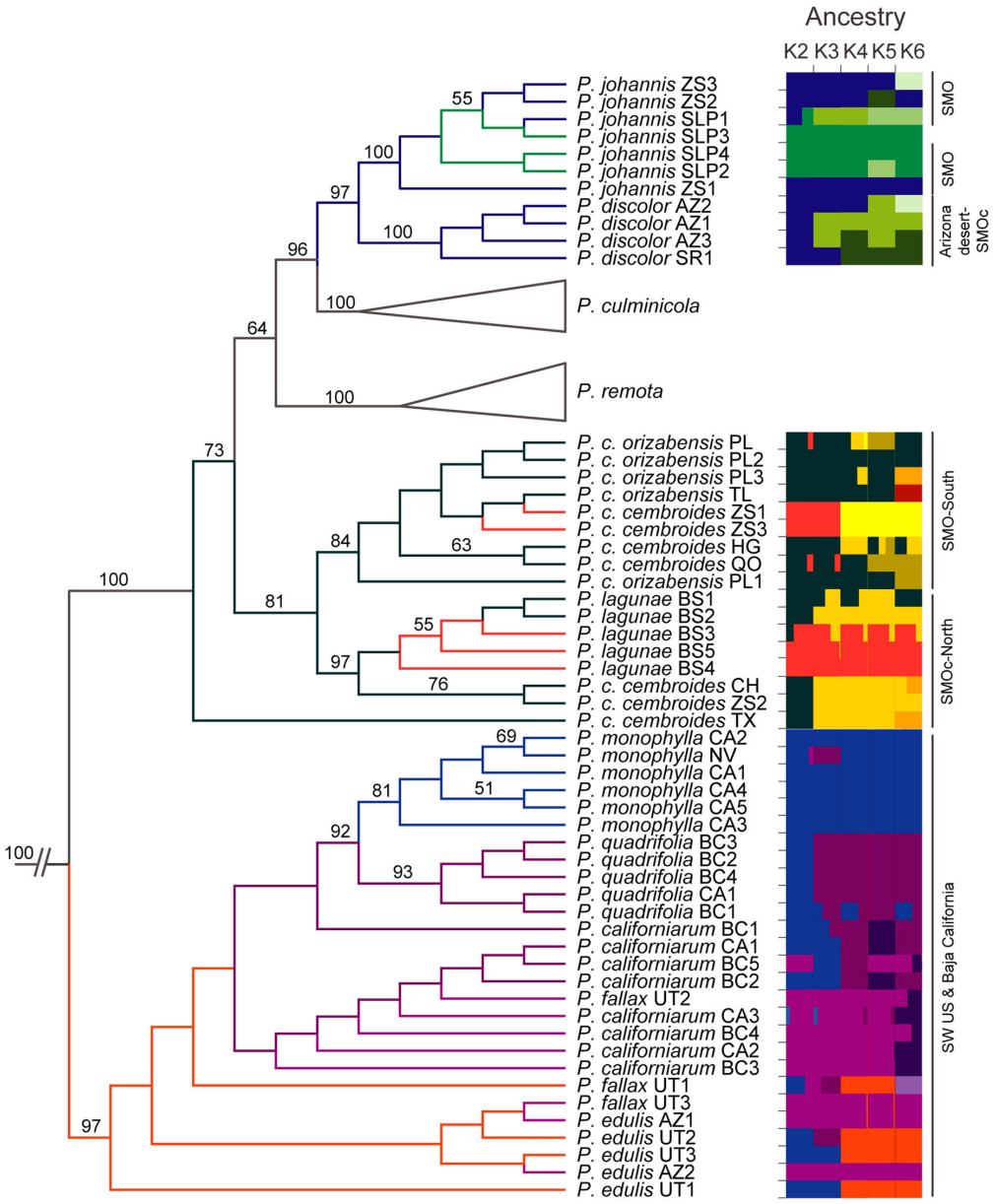


FIGURE 6 Genetic admixture proportions in the three subgroups of small-cone pinyons. Analyses were performed for K values ranging from 2 to 6 with three different matrices containing 22,500 variants for the *P. cembroides* subgroup, 29,601 variants for *P. johannis* and *P. discolor*, and 30,666 variants for single-needle species, *P. edulis*, and *P. quadrifolia*. Different colors represent different clusters. The combination of different colors in a bar indicates the degree of admixture. Samples in the admixture model are in the same topological order as in the SVDquartets tree. Bootstrap values >50% are shown above the branches. SMO = Sierra Madre Oriental and SMOc = Sierra Madre Occidental.

subpopulations partitioned by the optimal clustering in each of the three groups did not match exactly with the clade organization in the SNP tree. In the *P. johannis* and *P. discolor* group, optimal clustering inferred two subpopulation structures for individuals of *P. johannis*, sharing one of these structures with *P. discolor*. Only one individual (*P. johannis* SLP1) in this group showed signs of admixture. For most of the subsequent higher *K* values, the subpopulation composed of only individuals of *P. johannis* was kept, and at least one of the other subpopulations was composed of individuals of the two species. In the *P. cembroides* group, one subpopulation ($K = 2$) included individuals of *P. cembroides* subsp. *cembroides* and *P. lagunae*, whereas the other subpopulation had individuals of all three taxa. Three individuals showed signs of admixture, including one of *P. cembroides* subsp. *orizabensis*. Three of five individuals of *P. lagunae* that do not have the same structure (BS3, BS4, and BS5) group with individuals of *P. cembroides* from Zacatecas, Mexico (ZS1 and ZS3) (SMOc-South; Figure 6). The *P. lagunae* BS3 individual was admixed with the main subpopulation from the north. Individuals predominantly from the south (Sierra Madre Oriental-South), forming one clade, were partitioned in more subpopulations than the individuals from the north, and thus had more admixed individuals in the successive *K* values. The subpopulations of the one-needled pines, *P. edulis*, and *P. quadrifolia* best matched the order of the clades in the tree (Figure 6) across most of the *K* values evaluated. The admixture run of $K = 2$ in this group resulted in a subpopulation composed of all the individuals of *P. monophylla* and *P. quadrifolia*, three of *P. californiarum*, one *P. fallax*, and three *P. edulis*. The other subpopulation comprised individuals of *P. californiarum*, *P. fallax*, and *P. edulis*, suggesting higher genetic variation and admixture among the individuals of these species than those of *P. monophylla* and *P. quadrifolia*. For $K = 3$, a new subpopulation containing mainly individuals of *P. quadrifolia* was observed. For the remaining values of *K*, the individuals of *P. monophylla* comprised one reliable subpopulation. Only one individual from *P. quadrifolia* showed strong admixture proportions with this subpopulation. Also, for different numbers of ancestral clusters, individuals of *P. californiarum* tended to be assigned to two different subpopulations denoting structure within the species. For $K > 3$, certain individuals of *P. fallax* and *P. edulis* were grouped in the same subpopulation as some individuals of *P. californiarum*; however, certain individuals of these two species also constituted an independent subpopulation.

DISCUSSION

Plastid and Nuclear DNA data

Our analyses of plastomes and low copy nuclear genes allowed us to re-evaluate previous studies of *Pinus* subsection *Cembroides* based on much smaller data sets (Gernandt et al., 2001, 2005). We analyzed plastomes for a comprehensive taxonomic sampling from multiple individuals per species within subsect. *Cembroides*. The alignment length (117,210 bp) was shorter, and the number of informative sites (11,281 bp) was fewer in our plastome alignment of *Pinus* subgen. *Strobus* than the genus-wide alignment by Parks et al. (2012), which had a total length of 141,265 aligned sites and 15,151 informative characters but only one individual per species. The plastid DNA relationships among pinyon pines, including those supported by high bootstrap values, changed somewhat between the studies. In contrast to Parks et al. (2012), *P. monophyla* and *P. quadrifolia* formed a well-supported clade (1.0 posterior probability [pp]). Likewise, *P. maximartinezii*, *P. pinceana*, and *P. rzedowskii* formed another well-supported clade (1.0 pp).

For nuclear DNA, we used fewer genes for phylogenetic estimates compared to a previous study (Montes et al., 2019) because we included an additional criterion to eliminate genes with exceptionally high substitution rates. Nonetheless, we added 36 more individuals than Montes et al. (2019). The individuals cover a greater part of the geographical distribution of *Pinus* subsect. *Cembroides*, including several individuals per species from edges of their ranges. For instance, we included eight individuals of *P. californiarum*, three from California, and five from Baja California (including one population from La Asamblea, the southern limit of the species). We also added four more individuals of *P. pinceana* from different populations in Mexico (Durango, Hidalgo, Querétaro, and Zacatecas), and expanded sampling of *P. cembroides* subsp. *cembroides* to include individuals from three more populations (Querétaro, Zacatecas, Mexico and Texas, USA). The phylogenetic position of taxa with expanded sampling is consistent with our previous study (Montes et al., 2019) although there are some differences in poorly supported relationships of closely related taxa.

Coalescent-based species delimitation

Species delimitations were inconsistent among GMYC, PTP, and Tr2 methods in *Pinus* subsection *Cembroides*. Single-locus based coalescent analyses accurately delimited the species of the big-cone pinyons, *P. maximartinezii*, *P. pinceana*, and *P. rzedowskii*. These three Mexican endemics are easily recognized based on their morphologically divergent needles, wood anatomy, cones, and seeds (Malusa, 1992). Also, both GMYC and bPTP methods identified *P. monophylla*, *P. quadrifolia*, and *P. remota* as species. These taxa are consistently recognized in morphological treatments (Price et al., 1998) and have been recovered as distinct lineages in molecular phylogenetic studies (Gernandt et al., 2003; Montes et al., 2019). Our results support treating *P. quadrifolia* from southwestern USA and Baja California, Mexico as separate from the single-needle pinyon pines, *P. monophylla* from southwestern USA and *P. californiarum* distributed in California and Baja California. Lanner (1974) proposed that *P. quadrifolia* originates from recent hybridization

between *P. californiarum* and a five-needled species that he named *P. juarezensis* Lanner (which we treat as a synonym of *P. quadrifolia*). The evidence suggests that *P. quadrifolia* is not a hybrid (Montes et al., 2019), although *P. quadrifolia* show signs of admixture with some individuals of *P. californiarum* and *P. monophylla* (Buck et al., 2020).

GMYC analyses of plastid DNA recovered more lineages (up to 24; Figure 3) than bPTP and Tr2. Nonetheless, GMYC lumped together greater numbers of putative species in the small-cone clade, mainly in the "A" and "C" groups. Our results failed to separate *P. culminicola*, *P. discolor*, and *P. johannis* (Group A). *Pinus culminicola* has consistently been treated as a valid species, whereas *P. discolor* and *P. johannis* have been treated as separate species (Perry, 1991, Price et al., 1998) or as a single variety of *P. cembroides* (Farjon and Styles, 1997). The three taxa have been recovered as a clade in phylogenetic studies (Malusa, 1992; Gernandt et al., 2003; Ortiz-Medrano et al., 2016; Montes et al., 2019).

The single-locus methods GMYC and bPTP did not separate *P. californiarum*, *P. edulis*, and *P. fallax* (Group C). Montes et al. (2019) recognized *P. fallax* as a valid species based on its phylogenetic position in an analysis of nuclear genes but the results from the three species delimitation methods presented do not support that conclusion here (the nuclear results are discussed below). Henceforth, we refer to this taxon as *P. edulis* var. *fallax*, as originally proposed by Little (1968). *Pinus californiarum* and *P. edulis* have been recognized based on morphology, phylogeny, geographic distribution, and distinctive precipitation regimens (Bailey, 1987; Cole et al., 2008; Montes et al., 2019). However, the species of clade "C" cluster together and were identified as only one species by bPTP and GMYC. This result corroborates the study by LaHood (1995), who found that populations of *P. edulis*, *P. edulis* var. *fallax*, and *P. californiarum* share plastid DNA due to introgression.

Our results reflect both the nature of the molecular data and the limitations of both bPTP and GMYC methods. Hernández-León et al. (2013) suggested that including longer sequences of plastid DNA could improve species estimates with GMYC in another pine clade, *Pinus* sect. *Trifoliae* Duhamel, but despite obtaining relatively high resolution among terminals using nearly complete plastomes (~117,000 bp) in this study (Appendix S9), bPTP identified fewer species than expected whereas GMYC identified more species.

Our results suggest that even though sampling was increased and longer plastid DNA sequences were used, both GMYC and bPTP are not accurate for delimiting pine species. Many pine and other tree species have large effective population sizes (Petit and Hampe, 2006), and the discriminatory power of GMYC and bPTP in recognizing species with plastome data could be associated with different biological phenomena, such as plastid capture, gene flow, and population size (Luo et al., 2018).

Multi-locus analysis with the Tr2 method recovered 13 species (*P. californiarum*, *P. cembroides*, *P. culminicola*, *P. discolor*, *P. edulis*, *P. johannis*, *P. lagunae*, *P. maximartinezii*, *P. monophylla*, *P. pinceana*, *P. quadrifolia*, *P. remota*, and *P. rzedowskii*). Moreover, our delimitation results with Tr2 and multivariate analyses agreed with recent phylogenetic analyses (Montes et al., 2019) but failed to identify *P. edulis* var. *fallax* as a distinct species. Although we included one individual from near the type locality in Gila, Arizona, it grouped with *P. edulis* from Utah and Arizona. Thus, our results support the hypothesis by Little (1968) that single-needle pinyon pines from Arizona are a taxonomic variety of *P. edulis*, which predominantly has two needles per fascicle. *Pinus edulis* and *P. edulis* var. *fallax* are parapatric in distribution (Malusa, 1992; Cole et al., 2008) and some populations co-occur in Arizona and New Mexico. Both taxa occupy areas with similar seasonal precipitation (Cole et al., 2008). Needle numbers seem to be associated with distinct precipitation regimens and seasonality, and particularly in *P. edulis* var. *fallax* the needle numbers could be an adaptation to water deficit in summer (Cole et al., 2008). Nevertheless, two individuals of *P. edulis* var. *fallax* with one and two-needles on the same tree (UT1 and UT2) grouped with *P. californiarum*. Bailey (1987) studied the morphological similarities mainly in needles, resin canals, cones, and seeds of single-needle pinyons from Arizona and concluded that they are a taxonomic variety of *P. californiarum* (predominantly one-needle).

The phylogenetic position of *P. edulis* var. *fallax* individuals could be reflecting the presence of shared ancestral polymorphism or introgression from *P. edulis* var. *fallax* into *P. californiarum*. However, the hypothesis of introgression from *P. edulis* var. *fallax* into *P. californiarum* was not recovered using phylogenetic network analysis (Than et al., 2008) of nuclear genes by Montes et al. (2019). The coalescent, paleoclimatic, ecological, and genetic evidence do not reinforce the species boundaries between *P. edulis* var. *fallax* and *P. californiarum*. Increased sampling of *P. edulis* var. *fallax* populations is required for a more complete perspective of this taxon.

In morphology-based classifications, *P. californiarum* and *P. monophylla* are united by possessing solitary needles (Malusa, 1992) and *P. californiarum* has been considered a taxonomic variety of *P. monophylla* (Silba, 1990). Bailey (1987) segregated *P. californiarum* from *P. monophylla* based on differences in fascicle sheath length, the number of leaf resin canals, the shape of the base of the seed cone, and seed size. Our results support Bailey (1987) in recognizing this taxon as a valid species. Coalescent phylogenetic analyses recovered *P. californiarum* as an independent lineage and sister to *P. monophylla* and *P. quadrifolia* (Figures 7 and 8 in Montes et al., 2019). *Pinus californiarum* occurs in southeastern and central California and northern Baja California (Silba, 1990) and co-occurs with *P. quadrifolia* in both regions. Although *P. californiarum* occurs in sympatry with *P. quadrifolia* at some sites, it is easy to distinguish one taxon from the other because *P. quadrifolia* has approximately four needles per fascicle. *Pinus monophylla* and *P. californiarum* both occur in California (Bailey, 1987) but are parapatric or allopatric in distribution (Critchfield and Little,

1966). The geographic distribution of *P. monophylla* extends to western Utah, northwestern Arizona, southern Idaho, and western Nevada (Critchfield and Little, 1966; Bailey, 1987). Both single-needle pinyon pines occur in regions with high winter precipitation (Cole et al., 2008). Morphological, phylogenetic, paleoclimatic, geographical, ecological, and genetic evidence support the recognition of *P. californiarum* and *P. monophylla* as separate species.

Pinus quadrifolia and *P. monophylla* present divergent morphology in the trunk, number of leaves per fascicle, number of rows of stomata on the leaves, and leaf anatomy (Bailey, 1987; Farjon and Styles, 1997). The species are genetically distinct (Montes et al., 2019; Buck et al., 2020) but in our PCA analysis, *P. quadrifolia* is indistinguishable from *P. monophylla*. In contrast, DAPC was capable of separating the two. *Pinus quadrifolia* is sister to *P. monophylla* and likely they share alleles due to shared ancestral polymorphism or introgression (see below).

Pinus cembroides is distributed in the southern USA and widespread in Mexico (Critchfield and Little, 1966). This species was segregated into three taxa (Bailey, 1983; Passini, 1987) based on morphology and distribution. All three taxa share characters including a tree growth form with a monopodial and short trunk, short (4–6 mm) and loosely imbricate fascicle sheaths, ovoid to cylindric vegetative buds, rough shoots with non-decurrent or short decurrent pulvini, and a pinkish megagametophyte (Farjon and Styles, 1997). Our results support the treatment by Passini (1987) of *P. lagunae* as a distinct species from *P. cembroides* but do not support the proposal by Bailey and Hawksworth (1992) to elevate *P. cembroides* subsp. *orizabensis* to specific status. The *P. cembroides* subsp. *cembroides* + *P. cembroides* subsp. *orizabensis* clade is identified as a single species (*P. cembroides*).

The rooted triplets method had discriminatory power to delimit *P. lagunae* and *P. cembroides*. *Pinus lagunae* is endemic to Baja California Sur and is widely separated geographically from the rest of *P. cembroides*, whereas *P. cembroides* subsp. *cembroides* and *P. cembroides* subsp. *orizabensis* are allopatric or parapatric (Bailey, 1983). Based on morphology, *P. lagunae* differs from *P. cembroides* in height, substantially longer leaves (4–7 cm; Passini and Pinel, 1987), and in internal and external leaf cuticular characteristics such as an elliptical to rectangular stomatal apparatus, and circular stomata that usually lack a plug (Whang et al., 2001). *Pinus lagunae* occurs in a subtropical climate on granitic slopes whereas *P. cembroides* inhabits a broad elevational range with vegetation types ranging from semi-desert to montane forest (Bailey, 1983; Passini and Pinel, 1987; Farjon and Styles, 1997). Moreover, climatic changes during glacial or interglacial periods in the Pleistocene may have affected the geographic range and genetic composition of pine species (Moreno-Letelier and Piñero, 2009). We hypothesize that the divergence of *P. cembroides* and *P. lagunae* occurred during the Pleistocene and the current geographic distribution of *P. lagunae* may be the result of climatic fluctuations during that time, resulting in an expansion of its range towards the south and its subsequent persistence in the southern Baja California refuge. A similar history was reported in the columnar cactus *Pachycereus pringlei* (S. Watson) Britton & Rose distributed in the Baja California Peninsula and the Sonoran Desert where the interaction of climatic fluctuations, historical vicariance, and dispersal can explain its current biogeographic pattern (Gutiérrez-Flores et al., 2016).

The rooted triplets method and multivariate analyses also were congruent in recovering *P. discolor* and *P. johannis* as separate species. Our results support previous studies (Flores-Rentería et al., 2013, Ortíz-Medrano et al., 2016; Montes et al., 2019) in treating *P. discolor* and *P. johannis* as distinct. *Pinus discolor* is not a synonym of *P. johannis* as suggested by Passini (1994). *Pinus johannis* and *P. discolor* are morphologically similar, although some differences have been identified. For instance, the number of cotyledons was reported as 6–11 in *P. johannis* and 8–15 in *P. discolor* (Robert, 1978; Little, 1968), height was reported as 2 to 6 m in *P. johannis* and 5 to 12 m in *P. discolor* (Bailey and Hawksworth, 1983), and growth form was described as a multi-stemmed shrub or tree in *P. johannis* compared to a tree in *P. discolor* (Bailey and Hawksworth, 1979) but these claims have caused confusion and disagreement. The two taxa have yet to evolve clear morphological differences although apparently, they are geographically separated. The distribution of *P. johannis* is restricted to the Sierra Madre Oriental whereas *P. discolor* occurs in the Sky Islands of the southwestern USA, the Sierra Madre Occidental, and the southern Sierra Madre Oriental in San Luis Potosí (Bailey and Hawksworth, 1979). Moreover, *P. johannis* develops on sand-textured lithic rendzina or calcareous soils (Robert, 1978) whereas *P. discolor* occurs on arid slopes and ridges (Farjon and Styles, 1997). Populations of *P. discolor* are adapted to a mild winter climate (Little, 1968), whereas *P. johannis* is adapted to a longer winter period from October to March (Robert, 1978).

Pinus discolor and *P. johannis* differ significantly in the concentration of sabinene-related monoterpenes such as sabinene, thujene, γ-terpinene, terpinolene, and *p*-cymene. *Pinus discolor* produces a higher quantity of these monoterpenes compared to *P. johannis*, which has them in trace amounts (Zavarin and Snajberk, 1986). Traditionally, terpenes have been used as a character to differentiate species of pines (Mitić et al., 2017) but the differences are not always significant at the species level because the composition of monoterpenes differs very little among species or there is great variability in composition of monoterpenes within species (Zavarin et al., 1980; Snajberk and Zavarin, 1986; Zavarin and Snajberk, 1987).

Admixture analysis

Our results provide a picture of possible interbreeding in the three groups of small-cone pinyon pines: (1) *Pinus cembroides* subsp. *cembroides*, *P. cembroides* subsp.

orizabensis, and *P. lagunae*; (2) *P. johannis* and *P. discolor*; and (3) *P. californiarum*, *P. edulis* var. *fallax*, and *P. monophylla*. For all values of *K* analyzed (2 to 6), two subpopulations ($K = 2$) were the optimal clustering in this study (Figure 6). The number of subpopulations partitioned by the optimal clustering did not coincide with the number of lineages in the SNPs tree, particularly in the single-needled pinyons, *P. edulis*, and *P. quadrifolia* subgroup where five previously hypothesized taxa are considered (Figure 6) and the best model of *K* in admixture resulted in two subpopulations. Distribution of ancestry fractions indicate that *P. monophylla* (s.s.) shares genetic diversity with individuals of *P. quadrifolia*, *P. californiarum*, *P. edulis*, and *P. edulis* var. *fallax*. Buck et al. (2020) reported from 1868 SNPs that interbreeding does occur between *P. californiarum* and *P. quadrifolia*, as well as between *P. quadrifolia* and *P. monophylla*, and less commonly between *P. monophylla* and *P. californiarum*. Also, Montes et al. (2019) detected gene flow in *P. edulis* from *P. monophylla* using nuclear genes and SNPs. The other subpopulation indicates that individuals of *P. edulis* and *P. edulis* var. *fallax* are introgressed with *P. californiarum* but the direction of gene flow was not determined (Montes et al., 2019). Our results suggest higher genetic variation and admixture among the individuals of *P. edulis*, *P. edulis* var. *fallax*, and *P. californiarum* than those of *P. monophylla* and *P. quadrifolia*. Although interbreeding occurs between *P. monophylla* and *P. quadrifolia* it is less common (Buck et al., 2020). We observed genetic structure in *P. californiarum*, *P. monophylla,* and *P. quadrifolia* for $K = 3$ to $K = 6$, supporting that they are genetically distinct species. Conversely, we did not observe genetic structure between *P. edulis* and *P. edulis* var. *fallax*.

Distribution of ancestry fractions indicates that *P. lagunae* is introgressed with *P. cembroides*. According to Montes et al. (2019), evidence of reticulation between *P. lagunae* and *P. cembroides* was weak but signatures of admixture could be the result of long-distance pollen dispersal. Both species release their pollen in a short interval of time from May to July (Farjon and Styles, 1997). However, *P. cembroides* and *P. lagunae* are sisters and it is more likely that their shared genetic diversity is due to retention of ancestral polymorphism.

Distribution of ancestry fractions indicates that *P. johannis* is introgressed with *P. discolor* but this was not detected by Montes et al. (2019). *Pinus discolor* and *P. johannis* are closely related and share little genetic diversity that can be caused by introgression or ILS. This sign of admixture occurs only in one individual of *P. johannis* from San Luis Potosí where *Pinus discolor* and *P. johannis* come into close contact, separated by a ca. 120 km between San Miguelito Mountains and Las Charcas. Our results showed genetic structure in these species and support that they are genetically distinct. Nonetheless, more field work is needed in San Luis Potosí to explore the evidence of historical contact between *P. discolor* and *P. johannis*.

Perspectives for the use of coalescent-based methods of species delimitation

Employing data from target enrichment and genome skimming (Hyb-Seq; Weitemier et al., 2014) permitted us to delimit species using plastome and nuclear DNA sequences in a group of pines. Moreover, we examined the utility of using coalescent-based models to assess the boundaries with single-locus and multi-locus data. Our study revealed the potential of using single and multi-locus methods to estimate species in the presence of ILS and recent divergence (Gernandt et al., 2008; Montes et al., 2019). The multi-locus method Tr2 provided an estimate that better matched our expectations based on morphology, geography, and previous genetic studies.

AUTHOR CONTRIBUTIONS

J.R.M. performed both field and laboratory work, assembled DNA sequences, performed the phylogenetic, delimitation, and multivariate analyses, and was the primary author for the manuscript. J.R.M., A.M.L., and D.S.G. designed the study and performed fieldwork. P.P. provided SNP data and performed admixture analyses. All authors reviewed and edited the manuscript.

ACKNOWLEDGMENTS

The authors are grateful to Nelly López for assisting in genomic library construction and María Inés Badillo for a valuable revision of a previous version of the manuscript. We also thank José Delgadillo for providing logistical assistance during collecting trips to Baja California. We thank Jorge Pérez de la Rosa and Abisaí García for sharing collections of *Pinus lagunae*, and Laura Figueroa for collecting and providing material of *Pinus pinceana*. We thank Xitlali Aguirre-Dugua, Eng. Mario S. Montes-Montiel, PT. Angélica Castolo P. for their participation in fieldwork. We also thank Dra. Lidia I. Cabrera and the LANABIO of the Instituto de Biología, UNAM. Two anonymous reviewers provided valuable comments on a previous draft of the manuscript. This project was funded by PAPIIT-DGAPA, UNAM Grant IN209816 and CONACyT Grant 221694, and is part of the Ph.D. dissertation of J. R. Montes in the Posgrado de Ciencias Biológicas, Instituto de Biología, UNAM.

DATA AVAILABILITY STATEMENT

The nuclear, SNP, and plastid DNA sequence alignments analyzed in this study can be found in Appendices S14, S15, and S16, respectively.

ORCID

José-Rubén Montes http://orcid.org/0000-0001-6441-5983
David S. Gernandt http://orcid.org/0000-0002-3592-994X

REFERENCES

Aguirre-Dugua, X., and D. S. Gernandt. 2017. Complete plastomes of three endemic Mexican pine species (*Pinus* subsection *Australes*). Mitochondrial DNA part B, Resources 2: 562–565.

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

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Website: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bailey, D. K. 1983. A new allopatric segregate from and a new combination in *Pinus cembroides* Zucc. at its southern limits. Phytologia 54: 89–100.

Bailey, D. K. 1987. A study of *Pinus* subsection *Cembroides*. I: The single-needle pinyons of the Californias and the Great Basin. Notes from the Royal Botanic Garden Edinburgh 44: 275–310.

Bailey, D. K., and F. G. Hawksworth. 1979. Pinyons of the Chihuahuan Desert region. Phytologia 44: 129–133.

Bailey, D. K., and F. G. Hawksworth. 1983. Pinaceae of the Chihuahuan Desert region. Phytologia 53: 226–234.

Bailey, D. K., and F. G. Hawksworth. 1992. Change in status of *Pinus cembroides* subsp. *orizabensis* (Pinaceae) from Central Mexico. Novon 2: 306–307.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Leslin, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455–477.

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

Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury. 2012. Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. Molecular Biology and Evolution 29: 1917–1932.

Buck, R., S. Hyasat, A. Hossfeld, and L. Flores-Rentería. 2020. Patterns of hybridization and cryptic introgression among one- and four-needled pinyon pines. Annals of Botany 126: 401–411.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17: 540–552.

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

Cole, K. L., J. F. Fisher, S. T. Arundel, J. Cannella, and S. Swift. 2008. Geographical and climatic limits of needle types of one-and two-needled pinyon pines. Journal of Biogeography 35: 257–269.

Cole, K. L., J. F. Fisher, K. Ironside, J. I. Mead, and P. Koehler. 2013. The biogeographic histories of *Pinus edulis* and *Pinus monophylla* over the last 50,000 years. Quaternary International 310: 96e110.

Critchfield, W. B., and E. L. Little. 1966. Geographic distribution of the pines of the world, Miscellaneous Publication 991. U.S. Department of Agriculture, Washington, District of Columbia, USA.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. The variant call format and VCFtools. Bioinformatics 27: 2156–2158.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: More models, new heuristics and parallel computing. Nature Methods 9: 772.

de Queiroz, K. 2007. Species concepts and species delimitation. Systematic Biology 56: 879–886.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11–15.

Ence, D. D., and B. C. Carstens. 2011. SpedeSTEM: A rapid and accurate method for species delimitation. Molecular Ecology Resources 11: 473–480.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32: 3047–3048.

Farjon, A., and D. Filer. 2013. An atlas of the world's conifers: An analysis of their distribution, biogeography, diversity and conservation status. Brill Press, Leiden, Netherlands.

Farjon, A., and B. Styles. 1997. *Pinus* (Pinaceae). Flora neotropica monograph 75. New York Botanical Garden, Bronx, New York, USA.

Feng, X., J. Liu, and X. Gong. 2016. Species delimitation of the *Cycas segmentifida* complex (Cycadaceae) resolved by phylogenetic and distance analyses of molecular data. Frontiers in Plant Science. 7: 134.

Flores-Rentería, L., A. Wegier, D. Ortega Del Vecchyo, A. Ortíz-Medrano, D. Piñero, A. V. Whipple, F. Molina-Freaner, et al. 2013. Genetic, morphological, geographical and ecological approaches reveal phylogenetic relationships in complex groups, an example of recently diverged pinyon pine species (subsection *Cembroides*). Molecular Phylogenetics and Evolution 69: 940–949.

Fujisawa, T., A. Aswad, and T. G. Barraclough. 2016. A rapid and scalable method for multilocus species delimitation using Bayesian model comparison and rooted triplets. Systematic Biology 65: 759–771.

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

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

Gernandt, D. S., G. Geada López, S. Ortiz García, and A. Liston. 2005. Phylogeny and classification of *Pinus*. Taxon 54: 29–42.

Gernandt, D. S., A. Liston, and D. Piñero. 2001. Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: Implications for molecular systematic studies of pine species complexes. Molecular Phylogenetics and Evolution 21: 449–467.

Gernandt, D. S., A. Liston, and D. Piñero. 2003. Phylogenetics of *Pinus* subsections *Cembroides* and *Nelsoniae* inferred from cpDNA sequences. Systematic Botany 28: 657–673.

Gernandt, D. S., S. A. Magallón, G. Geada López, O. Zerón Flores, A. Willyard, and A. Liston. 2008. Use of simultaneous analyses to guide fossil-based calibrations of Pinaceae phylogeny. International Journal of Plant Sciences 169: 1086–1099.

Gernandt, D. S., and J. A. Pérez de la Rosa. 2014. Biodiversity of Pinophyta (conifers) in Mexico. Revista Mexicana de Biodiversidad 85: 126–133.

Grünwald, N. J., Z. N. Kamvar, and S. E. Everhart. 2016. Population genetics in R. Website: https://grunwaldlab.github.io/Population_Genetics_in_R/gbs_analysis.html

Gutiérrez-Flores, C., F. J. García-De León, J. L. León-de la Luz, and J. H. Cota-Sánchez. 2016. Microsatellite genetic diversity and mating systems in the columnar cactus *Pachycereus pringlei* (Cactaceae). Perspectives in Plant Ecology, Evolution and Systematics 22: 1–10.

Hernández-León, S., D. S. Gernandt, J. A. Pérez de la Rosa, and L. Jardón-Barbolla. 2013. Phylogenetic relationships and species delimitation in *Pinus* section *Trifoliae* inferred from plastid DNA. PLoS One 8: e70501.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755.

IUCN. 2019. The IUCN red list of threatened species, version 2019-3. International Union for Conservation of Nature, Gland, Switzerland. Website: http://www.iucnredlist.org

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

Jones, G., Z. Aydin, and B. Oxelman. 2015. DISSECT: An assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. Bioinformatics 31: 991–998.

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

Kan, X-Z., S-S. Wang, X. Ding, and X-Q. Wang. 2007. Structural evolution of nrDNA ITS in Pinaceae and its phylogenetic implications. Molecular Phylogenetics and Evolution 44: 765–777.

Kapli, P., S. Lutteropp, J. Zhang, K. Kobert, P. Pavlidis, A. Stamatakis, and T. Flouri. 2017. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. Bioinformatics 33: 1630–1638.

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

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

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

LaHood. E. 1995. A chloroplast DNA phylogeny of nine taxa in *Pinus Cembroides* subsection. MSc. thesis, Northern Arizona University, Flagstaff, Arizona, USA.

Lanner, R. M. 1974. A new pine from Baja California and the hybrid origin of *Pinus quadrifolia*. *Southwestern Naturalist* 19: 75–95.

Lanner, R. M. 1981. The piñon pine: A natural and cultural history. University of Nevada Press, Reno, Nevada, USA.

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27: 2987–2993.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Arxiv 1303.3997. Website: https://arxiv.org/abs/1303.3997

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

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Liston, A., W. A. Robinson, D. Piñero, and E. R. Alvarez-Buylla. 1999. Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. *Molecular Phylogenetics and Evolution* 11: 95–109.

Little, E. L. 1968. Two new pinyon varieties from Arizona. *Phytologia* 17: 329–342.

Long, C., and L. Kubatko. 2018. The effect of the gene flow on coalescent-based species-tree inference. *Systematic Biology* 67: 770–785.

Long, C., and L. Kubatko. 2019. Identifiability and reconstructibility of species phylogenies under a modified coalescent. *Bulletin of Mathematical Biology* 81: 408–430.

López-Giráldez, F., and J. P. Townsend. 2011. PhyDesign: An online application for profiling phylogenetic informativeness. *BMC Evolutionary Biology* 11: 1–4

López-Reyes, A., J. A. Pérez de la Rosa, E. Ortiz, and D. S. Gernandt. 2015. Morphological, molecular, and ecological divergence in *Pinus douglasiana* and *P. maximinoi*. *Systematic Botany* 40: 658–670.

Luo, A., C. Ling, S. Y. W. Ho, and C-D. Zhu. 2018. Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Systematic Biology* 67: 830–846.

Madrigal, S. X., and D. M. Caballero. 1969. Una nueva especie mexicana de *Pinus*. *Boletín Técnico del Instituto Nacional de Investigaciones Forestales* 26: 1–11.

Malusa, J. 1992. Phylogeny and biogeography of the pinyon pines (*Pinus* subsect. *Cembroides*). *Systematic Botany* 17: 42–66.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. Proceedings of the gateway computing environments whorkshop (GCE) 14 November 2010, vol. 14, 1–8. Institute of Electrical and Electronics Engineers, New Orleans, Louisiana, USA. Website: https://doi.org/10.1109/GCE.2010.5676129

Minh, B.Q., H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, and R. Lanfear. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.

Mirarab, S., and T. Warnow. 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31: i44–i52.

Mirov, N. T. 1967. The genus Pinus. Ronald Press, New York, New York, USA.

Mitić, Z. S., B. M. Nikolić, M. S. Ristić, V. V. Tešević, S. R. Bojović, and P. D. Marin. 2017. Terpenes as useful markers in differentiation of natural populations of relict pines *Pinus heldreichii*, *P. nigra*, and *P. peuce*. *Chemistry and Biodiversity* 14: e1700093.

Monaghan, M. T., R. Wild, M. Elliot, T. Fujisawa, M. Balke, D. J. G. Inward, D. C. Lees, et al. 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology* 58: 298–311.

Montes, J. R., P. Peláez, A. Willyard, A. Moreno-Letelier, D. Piñero, and D. S. Gernandt. 2019. Phylogenetics of *Pinus* subsection *Cembroides* Engelm. (Pinaceae) inferred from low-copy nuclear gene sequences. *Systematic Botany* 44: 501–518.

Moreno-Letelier, A., A. Ortíz-Medrano, and D. Piñero. 2013. Niche divergence versus neutral processes: Combined environmental and genetic analyses identify contrasting patterns of differentiation in recently diverged pine species. *PLoS One* 8: e78228.

Moreno-Letelier, A., and D. Piñero. 2009. Phylogeographic structure of *Pinus strobiformis* Engelm. across the Chihuahuan Desert filter-barrier. *Journal of Biogeography* 36: 121–131.

Mort, M. E., J. K. Archibald, C. P. Randle, N. D. Levsen, T. R. O'Leary, K. Topalov, C. M. Wiegand, et al. 2007. Inferring phylogeny at low taxonomic levels: Utility of rapidly evolving cpDNA and nuclear ITS loci. *American Journal of Botany* 94: 173–183.

Neale, D. B., J. L. Wegrzyn, K. A. Stevens, A. V. Zimin, D. Puiu, M. W. Crepeau, C. Cardeno, et al. 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* 15: R59.

Neves, L. G., J. M. Davis, W. B. Barbazuk, and M. Kirst. 2013. Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant Journal* 75: 146–156.

Nobis, M. P., C. Traiser, and A. Roth-Nebelsick. 2012. Latitudinal variation in morphological traits of the genus *Pinus* and its relation to environmental and phylogenetic signals. *Plant Ecology and Diversity* 5: 1–11.

O'Meara, B. C. 2010. New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology* 59: 59–73.

O'Meara, B., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60: 922–933.

Ortíz-Medrano, A., D. P. Scantlebury, A. Vázquez-Lobo, A. Mastretta-Yanes, and D. Piñero. 2016. Morphological and niche divergence of pinyon pines. *Ecology and Evolution* 6: 2886–2896.

Paradis, E., and K. Schliep. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528.

Parks, M., R. Cronn, and A. Liston. 2012. Separating the wheat from the chaff: Mitigating the effects of noise in a plastome phylogenomic data set from *Pinus* L. (Pinaceae). *BMC Evolutionary Biology* 12: 1–17.

Passini, M.-F. 1987. The endemic pinyon of Lower California *Pinus lagunae* M.-F. Passini. *Phytologia* 63: 337–338.

Passini, M.-F. 1994. Synonymie entre *Pinus discolor* Bailey & Hawksworth et *Pinus johannis* M.-F. Robert. *Acta Botanica Gallica* 141: 387–388.

Passini, M.-F., and N. Pinel. 1987. Morphology and phenology of *Pinus lagunae* M.-F. Passini. *Phytology* 63: 331–336.

Perry, J. P. 1991. The pines of Mexico and Central America. Timber Press, Portland, Oregon, USA.

Petit, R. J., and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics* 37: 187–214.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, et al. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* 55: 595–609.

Poulos, H. M. and G. P. Berlyn. 2007. Variability in needle morphology and water status of *Pinus cembroides* across an elevational gradient in the Davis Mountains of west Texas, USA. *Journal of the Torrey Botanical Society* 134: 281–288.

Price, R. A., A. Liston, and S. H. Strauss. 1998. Phylogeny and systematics of *Pinus*. *In* D. M. Richardson [ed.], Ecology and biogeography of Pinus, 49–68. Cambridge University Press, Cambridge, UK.

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

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81: 559–575.

Rambaut, A. 2012. FigTree. Tree figure drawing tool, version 1.4.0. Website: https://github.com/rambaut/figtree/releases

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67: 901–904.

Rieseberg, L. H., and D. E. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5: 65–84.

Robert, M.-F. 1978. Un nouveau pin pignon mexicain: *Pinus johannis*. *Adansonia*, ser. 2,18: 365–373.

Rosenberg, N. A. 2003. The shapes of neutral gene genealogies in two species: Probabilities of monophyly, paraphyly and polyphyly in a coalescent model. *Evolution* 57: 1465–1477.

RStudio team. 2019. RStudio: Integrated development for R. Rstudio Inc., Boston, Massachusetts, USA. Website: http://www.rstudio.com/

Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian, and R. H. Crozier. 2010. Integrative taxonomy: A multisource approach to exploring biodiversity. *Annual Review of Entomology* 55: 421–438.

SEMARNAT. 2010. Norma Oficial Mexicana NOM-059-ECOL-2001. Protección ambiental. Especies nativas de México de flora y fauna silvestres. Categorías de riesgo y especificaciones para su inclusión, exclusión o cambio. Lista de especies en riesgo. Diario Oficial de la Federación, 30 de diciembre de 2010. Mexico City, Mexico.

Silba, J. 1990. A supplement to the international census of the Coniferae, II. *Phytologia* 68: 7–78.

Sites, J. W., and K. A. Crandall. 1997. Testing species boundaries in biodiversity studies. *Conservation Biology* 11: 1289–1297.

Sites J. W., and J. C. Marshall. 2003. Delimiting species: A renaissance issue in systematic biology. *Trends in Ecology and Evolution* 18: 462–470.

Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals conflict, concordance and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15: 150.

Smith S. A., N. Walker-Hale, J. F. Walker, and J. W. Brown. 2020. Phylogenetic conflicts, combinability, and deep phylogenomics in plants. *Systematic Biology* 69: 579–592.

Snajberk, K., and E. Zavarin. 1986. Monoterpenoid differentiation in relation to the morphology of *Pinus remota*. *Biochemical Systematics and Ecology* 14: 155–163.

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

Swofford, D. L. 2002. PAUP* Phylogenetic analysis using parsimony (*and other methods), v. 4.0b10. Sinauer Associates, Sunderland, Massachusetts, USA.

Syring. J., A. Willyard, R. Cronn, and A. Liston. 2005. Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. *American Journal of Botany* 92: 2086–2100.

Than, C., D. Ruths, and L. Nakhleh. 2008. PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9: 322.

Townsend, J. P. 2007. Profiling phylogenetic informativeness. *Systematic Biology* 56: 222–231

Tsutsui, K., A. Suwa, K. Sawada, T. Kato, T. A. Ohsawa, and Y. Watano. 2009. Incongruence among mitochondrial, chloroplast and nuclear gene trees in *Pinus* subgenus *Strobus* (Pinaceae). *Journal of Plant Research* 122: 509–521.

Turna, I., and D. Güney. 2009. Altitudinal variation of some morphological characters of Scots pine (*Pinus sylvestris* L.) in Turkey. *African Journal of Biotechnology* 8: 202–208.

Villanueva, R. A. M., Z. J. Chen, and H. Wickham. 2016. Ggplot2: Elegant graphics for data analysis using the grammar of graphics. Springer-Verlag, New York, New York, USA.

Wang, J., Y. Wu, G. Ren, Q. Guo, J. Liu, and M. Lascoux. 2011. Genetic differentiation and delimitation between ecologically diverged *Populus euphratica* and *P. pruinosa*. *PLoS One* 6: e26530.

Wegrzyn, J. L., J. D. Liechty, K. A. Stevens, L-S. Wu, C. A. Loopstra, H. A. Vasquez-Gross, W. M. Dougherty, et al. 2014. Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* 196: 891–909.

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

Whang, S. S., J-H. Pak, R. S. Hill, and K. Kim. 2001. Cuticle micromorphology of leaves of *Pinus* (Pinaceae) from Mexico and Central America. *Botanical Journal of the Linnean Society, Linnean Society of London* 135: 349–373.

Willyard, A., D. S. Gernandt, B. Cooper, C. Douglas, K. Finch, H. Karemera, E. Lindberg, et al. 2021. Phylogenomics in the hard pines (*Pinus* subsection Ponderosae; Pinaceae) confirms paraphyly in *Pinus ponderosa*, and places *Pinus jeffreyi* with the California big cone pines. *Systematic Botany* 46: 538–561.

Willyard, A., D. S. Gernandt, K. Potter, V. Hipkins, P. Marquardt, M. F. Mahalovich, S. K. Langer, et al. 2017. *Pinus ponderosa*: A checkered past obscured four species. *American Journal of Botany* 104: 161–181.

Willyard, A., J. Syring, D. S. Gernandt, A. Liston, and R. Cronn. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution* 24: 90–101.

Wolfe, J. A., and H. E. Schorn. 1990. Taxonomic revision of the Spermatopsida of the Oligocene Creede flora, Southern Colorado. USGS Bulletin 1923. U.S. Geological Survey, Denver, Colorado, USA.

Yang, Z. 2015. The BPP program for species tree estimation and species delimitation. *Current Zoology* 61: 854–865.

Yang, Z., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences, USA* 107: 9264–9269.

Zavarin, E., and K. Snajberk. 1986. Monoterpenoid differentiation in relation to the morphology of *Pinus discolor* and *Pinus johannis*. *Biochemical Systematics and Ecology* 14: 1–11.

Zavarin, E., and K. Snajberk. 1987. Monoterpene differentiation in relation to the morphology of *Pinus culminicola*, *Pinus nelsonii*, *Pinus pinceana* and *Pinus maximartinezii*. *Biochemical Systematics and Ecology* 15: 307–312.

Zavarin, E., K. Snajberk, and R. Debry. 1980. Terpenoid and morphological variability of *Pinus quadrifolia* and its natural hybridization with *Pinus monophylla* in northern Baja California and adjoining United States. *Biochemical Systematics and Ecology* 8: 225–235.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 15–30.

Zhang, D., T. Xia, M. Yan, X. Dai, J. Xu, S. Li, and T. Yin. 2014. Genetic introgression and species boundary of two geographically overlapping pine species revealed by molecular markers. *PLoS One* 9: e101106.

Zhang, J., P. Kapli, P. Pavlidis, and A. Stamatakis. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 22: 2869–2876.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. List of individual taxa used in this study. The ID locality is based in ISO 3166-2, RENAPO, MX.

Appendix S2. Details for designing target DNA enrichment probes for conifers and processing of Hyb-Seq data.

Appendix S3. Conditions and concentrations of the parallel sequencing per sample. Sequencing depth = unenriched:enriched. The ID locality is based in ISO 3166-2, RENAPO, MX.

Appendix S4. Informativeness profile. (A) Plot of the genes list identified in PhyDesign with unusually high inferred substitution rates. Colored lines represent nuclear genes. (B) List of nuclear genes with unusually high inferred substitution rates. A total of 22 genes were eliminated from sampling.

Appendix S5. List of the individual taxa and statistics for plastome coverage. **Plastomes assembled and used in this study.

Appendix S6. Species delimitation hypotheses in Tr2. The hypothesis was tested according to the species reported for subsection *Cembroides* in several studies (Lanner, 1974; Bailey, 1987; Price et al., 1998; Gernandt et al., 2005; Montes et al., 2019). The numbers represent lineages.

Appendix S7. Maximum likelihood tree based on low-copy nuclear genes. Results presented are based on the concatenated analysis with partitions. Bootstrap support values >50% are shown on branches. The abbreviations of the states are based on ISO3166-2:MX. Subsections are colored, *Cembroides* by blue branches, *Balfourianae* by orange, *Nelsoniae* by green, *Gerardianae* by purple, *Krempfianae* by navy blue, and *Strobus* by gray.

Appendix S8. Maximum likelihood tree based on low-copy nuclear genes. Results presented are based on the concatenated analysis without partitions. Bootstrap support values >50% are shown on branches. The abbreviations of the states are based on ISO3166-2:MX. Subsections are colored, *Cembroides* by blue branches, *Balfourianae* by orange, *Nelsoniae* by green, *Gerardianae* by purple, *Krempfianae* by navy blue, and *Strobus* by gray.

Appendix S9. Bayesian inference (BI) tree based on plastome sequences. We included plastomes with coverage >20× with some exceptions reported in this appendix. Posterior probability (PP) values > 0.5 from the BI analysis are shown on branches. The abbreviations of the states are based on ISO3166-2:MX.

Appendix S10. Generalized Mixed Yule Coalescent using single thresholds models. Results presented are based on the plastome tree resulting from Bayesian inference (BI). Colored areas corresponding to each cluster are identified as potential species. The blue line represents the transition between coalescence and speciation. The corresponding lineage-through-time plot is given on the upper left. The coalescent model of sGMYC exhibited significantly better fit over the null model ($logL_{GMYC} = -529.7634$, $logL_{null} = -524.9513$, LRt = 9.624205, P = 0.008130748**)

Appendix S11. Generalized Mixed Yule Coalescent using multiple thresholds models. Results presented are based on the plastome tree resulting from Bayesian inference (BI). Colored areas corresponding to each cluster identified as potential species. The corresponding lineage-through-time plot is given on the upper left. The coalescent model of mGMYC exhibited significantly better fit than the null model ($logL_{GMYC} = -530.1472$, $logL_{null} = -524.9513$, LRt = 10.39178, P = 0.005539288**).

Appendix S12. Poisson tree processes using Bayesian inference solution. Results presented are based on the plastome tree resulting from Bayesian inference (BI). Colored areas corresponding to each cluster are identified as potential species. The likelihood plot indicated convergence of Markovian chains on the upper left. Bootstrap values (left) and posterior probability (right) are shown on branches.

Appendix S13. Poisson tree processes using maximum likelihood solution. Results presented are based on the plastome tree resulting from Bayesian inference (BI). Colored areas corresponding to each cluster identified as potential species. Bootstrap values are shown on branches.

Appendix S14. Concatenated alignment of nuclear genes.

Appendix S15. Concatenated alignment of SNPs.

Appendix S16. Plastome alignment with partitions, coding, and noncoding blocks.

How to cite this article: Montes, J.-R., P. Peláez, A. Moreno-Letelier, and D. S. Gernandt. 2022. Coalescent-based species delimitation in North American pinyon pines using low-copy nuclear genes and plastomes. American Journal of Botany 109(5): 1–21. https://doi.org/10.1002/ajb2.1847

3.3. Timing of diversification in North American pinyon pines using multigene molecular clocks

Montes, J.-R., A. Benítez-Villaseñor, R. Hernández-Gutiérrez, D. S. Gernandt (*In prep.*)

Timing of diversification in North American pinyon pines using multigene molecular clocks

José-Rubén Montes[1], Adriana Benítez-Villaseñor[1], Rebeca Hernández-Gutiérrez[1], David S. Gernandt[2, 3]

[1]Posgrado en Ciencias Biológicas, Instituto de Biología, Departamento de Botánica, Universidad Nacional Autónoma de México, A.P. 70-153, C.P. 04510, Ciudad de México, México;

[*]ruben.montes@st.ib.unam.mx; adriana.bv@outlook.com; rebecahdezgtz@gmail.com

[2]Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Apartado Postal 70-233, Ciudad de México, 04510, México.

[3]Author for correspondence; dgernandt@ib.unam.mx

***Abstract*—**We studied the timing of diversification in a clade of North American pinyon pines to understand its evolutionary dynamics using different calibration strategies. A total-evidence analysis was performed including with all extant species and three fossil taxa to evaluate the phylogenetic position of fossils and the reconstruction of the phylogenetic relationships of the extant groups. We used phylogenetic analyses and molecular clocks under the node-dating approach and lognormal relaxed-clock model from Hyb-Seq data (target enrichment and genome skimming). A total of 207 low-copy nuclear genes were used to estimate both phylogeny and divergence times in *Pinus* subsection *Cembroides* and close relatives. The clock-tree suggested that the origin of *Pinus* subsection *Cembroides* was in the Late Oligocene about 25.5 MYA. In contrast, diversification of the North American pinyon pines occurred in the Middle Miocene about 14 MYA. *Pinus discolor*, *P. johannis*, *P. monophylla*, and *P. quadrifolia* are the most recent diverging species of *Pinus* subsection *Cembroides.* Macroevolutionary analysis detected an increase in the speciation rate in *Pinus* subsection *Cembroides* and a decrease in close

relatives. In general, the extinction rates have remained constant throughout the time of the phylogeny.

Keywords— Fossils; macroevolutionary analysis; multi-locus; node-dating; *Pinus*; relaxed molecular clocks; target enrichment.

Species of the genus *Pinus* L. (Pinaceae) are widely distributed in the northern hemisphere (Critchfield and Little 1966). *Pinus* has major economic value due to production of timber, cellulose for wood pulp and cardboard, and resin (Farjon and Styles, 1997; Lanner, 1981). Pines are also of exceptional biological, ecological, and biogeographic importance due to their distribution and morphological diversity (Richardson 1998). *Pinus* is a monophyletic group and comprises two taxonomic subgenera (*Pinus* L. and *Strobus* Lemmon) and each subgenera subdivided into two sections: *Pinus* and *Trifoliae* Duhamel (*Pinus*) and *Quinquefoliae* Duhamel and *Parrya* Mayr (*Strobus*). *Pinus* section *Parrya* is divided into three subsections from North America: *Balfourianae* Engelm. (W U.S.A.), *Cembroides* Engelm. (SW North America, Mexico), and *Nelsoniae* Van Der Burgh from Mexico (Gernandt et al. 2005). *Pinus* subsection *Cembroides*, the focal group of this research, comprises fifhteen taxa (Montes et al. 2019). These pines are small to medium-sized trees or shrubs with heterogeneous morphology mainly in characters related to needles, cones, and seeds (Malusa 1992; Farjon and Styles 1997), but are united in possessing secondary leaves with deciduous fascicle sheaths, cone scale apices with dorsal umbos, and enlarged seeds that are functionally wingless, except for *P. rzedowskii* Madrigal & M.Caball wich is 20-30 mm length (Madrigal and Caballero 1969; Malusa 1992; Farjon and Styles 1997). *Pinus* subsection *Cembroides* also includes trees with extraordinary ecological and economical values in Mexico and the United States (Lanner 1981; Farjon and Styles 1997).

The timing of diversification in *Pinus* using fossils for calibration has been estimated across distinct taxonomic ranges such as genus-level (Willyard et al. 2007; Gernandt et

al. 2008; Saladin et al. 2017; Jin et al. 2021). The timing of diversification also has been studied at the section level; Hernández-León et al. (2013) employed molecular clocks to calculate the divergence time between species of three subsections of North American hard pines (*Australes* Loudon, *Contortae* Little and Critchfield, and *Ponderosae* Loudon) with a comprehensive taxonomic sampling. The understanding of divergence time in pines relays? on location of the fossil in the phylogeny, taxonomic samples, and calibration methods (Willyard et al. 2007; Gernandt et al. 2008; Saladin et al. 2017; Jin et al. 2021). Unrecognized diversity and unsampled species have a negative impact on divergence time estimates (Cusimano and Renner 2010) in terms of accuracy (Heath et al. 2008). Similarly, accurate divergence time estimates depend on robust phylogenetic hypotheses. If the phylogenies are inferred from a small number of loci they can be affected by incomplete lineage sorting (ILS; Degnan and Rosenberg 2009). High levels of gene discordance by ILS have been reported in all branches within *Pinus* subsection *Cembroides* (Montes et al., 2019).

Pinus subsection *Cembroides* is the second-most species-rich subsection in subgenus *Strobus* (Price et al. 1998; Montes et al. 2022). Nonetheless, its origin and diversification are still controversial. Axelrod (1986) suggests that its origin goes back to the Middle Eocene according to fossil record of *P*. *balli* Brown from Green River, Colorado (46 MYA) and appears to have evolved in refugia during the Eocene (Axelrod 1986; Millar 1993). In contrast, Wolfe and Schorn (1990) cast doubt on Axelrod´s hypothesis and they suggested that *Pinus* subsection *Cembroides* first appears in an Oligocene fossil deposit (27.2 MYA). Subsequently, in a series of molecular studies to understand the relationship between fossil and extant species, as well as to estimate absolute divergence times and molecular rates of the *Pinus*, different ages were reported for the origin of *Pinus* subsection *Cembroides*, although it was not the focal group. For instance, Gernandt et al., (2008) reported that the stem-age of pinyon pines from North America occurred in the early Oligocene, whereas Saladin et al. (2017) estimated a Middle Eocene age and Jin et al. (2021) estimated an Early Eocene age.

Regarding the crown-age of pinyon pines from North America, Jin et al. (2021) reported that *Pinus* subsection *Cembroides* diverged during the Oligocene or in the Miocene (23-5.3 MYA: Willyard et al. 2007; Gernandt et al. 2008; Saladin et al. 2017) Although they have radiated recently, the pinyon pines include *P*. *maximartinezii* Rzed., *P*. *pinceana* Gordon, and *P*. *rzedowskii* (big-cone clade), a clade with high levels of genetic variation (Delgado et al. 1999; Ledig et al. 2001) and divergent morphology in needles, cones, seeds, and wood anatomy (Malusa 1992; Farjon and Style 1997). However, the diversification dynamics of the sister group (small-cone clade) is still not well understood because the phylogenetic relationships had not been resolved or remained controversial, furthermore molecular dating phylogenies have not included all extant species (Table 1) (see Silba 1990; Price et al. 1998; Little 1968; Silba 1990; Farjon and Styles 1997). A recent study included all extant taxa reported for *Pinus* subsection *Cembroides* and the results with multiple genes provided satisfactory phylogenetic resolution and supported the recognition of the species *Pinus californiarum* D.K.Bailey, *P*. *discolor* D.K.Bailey & Hawksw., *P*. *johannis* M.-F.Robert, and *P*. *lagunae* (Montes et al. 2019).

Pinus subsection *Cembroides* does not have an abundant fossil record compared to the subsections of subgenus *Pinus* (see Millar 1993). The fossil record of pinyon pines extends possibly to the Late Oligocene (Wolfe and Schorn 1990). *Pinus lindgrenii* Knowlton from Idaho (Late Miocene) and *P*. *sanjuanensis* Axelrod discovered in Colorado (Late Oligocene) have been used to calibrated molecular phylogenies for *Pinus* (Saladin et al. 2017; Jin et al. 2021) and their phylogenetic placement has been based on possible morphological synapomorphies among extant species from *Pinus* subsection *Cembroides*. In *P*. *lindgrenii* and *P*. *sanjuanensis* the fossil record is incomplete and the close affinities with extant species are poorly understood. *Pinus lindgrenii* was described from ovulate cones and seeds whereas *P*. *sanjuanensis* was described from to ovoulate cones, seeds, and needles. *Pinus sanjuanensis* has been reported as allied to *P*. *edulis* Engelm. whereas *P*. *lindgrenii* has been identified as allied to *P*. *cembroides* and *P*. *edulis*

(Axelrod 1986). Another hypothesis suggested that *P. lindgrenii* is very similar to the modern species *P. monophylla* and *P. edulis* due to the range of cone size and seed size (Miller 1992). For calibration, *P. sanjuanensis* has been used to estimate the stem-age to *Pinus* subsection *Cembroides* while *P. lindgrenii* has been used as MRCA of *P. johannis*–*P. edulis* clade (see Saladin et al. 2017), but a critical evaluation with morphology and genetic data from extant species is required as a strategy to identify the position of *P. sanjuanensis* and *P. lindgrenii* within *Pinus*.

The use of a comprehensive taxonomic sampling within *Pinus* subsection *Cembroides* should increase the accuracy of divergence time estimates and the inclusion of multiple genes will increase the number of informative sites to estimate the divergence time and should compensate for rate heterogeneity (Magallón and Sanderson 2005). We used multiple fossils to calibrate molecular clocks in North American pinyon pines under the node-dating approach. This study is a contribution to understanding the timing of origin and diversification of trees adapted to arid and semiarid environments. Here our objectives are 1) to assess the phylogenetic position of *Pinus* fossil taxa based on a total-evidence matrix, 2) to re-estimate origin and divergence times of *Pinus* subsection *Cembroides* using low-copy nuclear genes and all known extant species, and 3) to estimate the diversification rates in *Pinus* subsection *Cembroides*.

MATERIALS AND METHODS

Sampling and Morphology— A morphological data set was compiled in Mesquite ver. 3.6.1 (Maddison and Maddison 2019) for a total of 22 taxa, including all fifteen extant taxa of *Pinus* subsection *Cembroides*, and *P. nelsonii* Shaw (subsection *Nelsoniae*), *P. aristata* Engelm., *P. longaeva* D.K.Bailey, and *P. balfouriana* A.Murray bis (subsection *Balfourianae*) as close relatives, and three fossil taxa: *P. lindgrenii*, *P. sanjuanensis*, assigned to *Pinus* subsection *Cembroides* by Axelrod (1986), and *P. crossii* Knowlton, assigned to subsection *Balfourianae* by Axelrod (1986). Other? assignments to *Pinus* subsection *Cembroides* by Axelrod (1986) and Millar (1998) are considered dubious

because these are based on “matching” extant plants organs with fossils and might gives little taxonomic significance of the "matching" characters in modern pine species (Wolf and Schorn 1990). Fossils assigned to subsection *Balfourianae* were chosen because they are available and represent the sister group to *Pinus* subsection *Cembroides* together with *Pinus* subsection *Nelsoniae* (Montes et al. 2019; Jin et al. 2021).

Primarily, we build a matrix with a total of 37 morphological characters coded previously reported, 11 from Malusa (1992), five from Gernandt et al. (2016), and the other 21 characters were incorporated from other sources (Shaw 1914; Little 1968; Robert 1978; Klaus 1980; Floyd 1983; Bailey 1987; Farjon and Styles 1997; Cobb et al. 2002; Eckenwalder 2009; Flores-Rentería et al. 2013). (Appendix S1; see Supplemental Data with this article). Later, we reviewed 300 specimens deposited in the Herbario Nacional (MEXU) and from field collections to corroborate some character states of 37 previously reported characters. The character states of fossil taxa were taken from previous morphological descriptions (Axelrod 1986; Wolfe and Schorn 1990; Miller 1992; Erwin and Schorn 2005). The list of characters included 20 multistate and 17 treated as binary. Eleven multistate characters were ordered and the rest were treated as unordered (Appendix S1). The characters were from branching and growth (2), foliar (8), vegetative buds (3), anatomy needles (6), pollen cones (2), cotyledons (1), phenology and sex distribution (2), seed cones (8), and seeds (5) (Appendix S1). We retained morphological characters with missing data for some species (mainly from the fossils) following the recommendation of Wiens (2006

Molecular Data — We used an alignment of 207 low-copy nuclear genes for 14 species of *Pinus* subsection *Cembroides* and 8 close relatives previously applied by Montes et al. (2022). The 207 nuclear genes resulted from a series of filters from 969 nuclear genes, including the exclusion of paralogs and genes with exceptionally high substitution rates (Montes et al. 2022). A total of forty individuals were included, representing 23 species. Fifteen terminals correspond to *Pinus* subsection *Cembroides*

species, three species of subsection *Balfourianae*, two species of subsection *Gerardianae* Loudon, and one species each of subsections *Nelsoniae* (section *Parrya*), , *Krempfianae* Little and Critchfield, and *Strobus* (section *Quinquefoliae*). This dataset represents a comprehensive taxonomic and genetic sampling from multiple individuals per species within *Pinus* subsection *Cembroides*.

Fossil Phylogenetic Placement— A total-evidence analysis was performed to evaluate the phylogenetic position of fossils, for this reason we did not use all 207 nuclear genes published by Montes et al. (2022). We used a combined matrix with 37 morphological characters and a concatenated alignment of 20 low-copy selected nuclear genes, which represent 10% of the genes analyzed by Montes et al. (2022). The selection criterion was the length of genes in base pairs, we choose the longest genes. The 20 nuclear genes were previously aligned with MAFFT ver. 7.3 (Katoh et al. 2002; Katoh and Standley 2013). We inferred a phylogenetic tree under Bayesian inference in MrBayes ver. 3.2.6 (Huelsenbeck and Ronquist 2001; Miller et al. 2010) from the combined matrix with 13,331 characters and 21 partition blocks, 1 for all morphological characters and 20 for nuclear genes. The analysis was conducted using a mixed model. For morphology, we used a fixed-rates continuous-time Markov model for the evolution of a discrete trait (Lewis, 2001), which corrects the coding bias resulting from only variable characters being present in the matrix.

For molecular data, each gene we assign both rate and frequency variation combined with invariant sites (I+G). The analysis was run using four chains, three heated and one cold, and heating of 0.2. Two independent runs for 300 millions of generations were performed with sampling every 1,000 generations, discarding 0.25 as a burn-in fraction. We used Tracer ver. 1.7.1. (Rambaut et al. 2018) to corroborate chain convergence. Tree topologies were summarized in a 50% majority rule tree. The consensus tree was imported into FigTree ver. 1.4.0 for further editing (Rambaut 2012).

Evolutionary Relationships — We generated a concatenated nuclear alignment from 207 individual genes and 40 individuals with the Perl script BeforePhylo.pl ver. 0.9.0 deposited in Github (https://github.com/qiyunzhu/BeforePhylo). We evaluated partition schemes at the level of genes in PartitionFinder ver. 2.1.1 (Lanfear et al. 2014, 2016). Model selection was estimated using BIC (Bayesian Information Criterion) and we used the unlinked option to calculate the branch lengths. The analysis was executed under option --rcluster-max for 207 genes.

We inferred a phylogenetic tree with maximum likelihood in RAxML ver. 8.2.10 (Stamatakis 2014) with partition blocks for 207 nuclear genes. The analysis was conducted using the General Time Reversible model (GTR) with the gamma parameter (G). The analysis was run using 1,000 heuristic searches and 1,000 bootstrap replicates. The best tree was imported into FigTree ver. 1.4.0 for further editing (Rambaut 2012).

Calibrations and Divergence Time Estimates— We simultaneously estimated both divergence times and phylogeny in BEAST ver. 2.4.8 (Bouckaert et al. 2014; Miller et al. 2010). First, we imported an alignment with 207 nuclear genes and ten partition subsets into BEAUti ver. 2.4.8 (Bouckaert et al. 2014). We assigned the general time-reversible model with the Gamma parameter (GTR+G) to each partition scheme. We estimated the divergence times under lognormal (uncorrelated) relaxed-clock model (Drummond et al. 2006), assuming that the number of bins is equal to the number of branches in the phylogeny. A birth-death process was designated for the tree topology (Yule 1924; Kendall 1948). The priors of the shape Gamma, substitution parameters, and topology were estimated with default restrictions. We also used a Gamma distribution for each nucleotide exchange per partition (α = 0.05, β = 10.0). A uniform distribution for the mean of the lognormal relaxed clock and Gamma distribution for standard deviation (α = 0.5396, β = 0.3819 by default) of the lognormal relaxed clock were selected. We assigned 0.5 for relative death rate and 1.0 for birth differential rate and selected a uniform distribution (0 up to 1, *offset* = 0) for both speciation and extinction rates.

Based on both synapomorphies and the position of fossils in a total-evidence analysis, we calibrated two nodes within *Pinus* subsection *Cembroides* and one node of subsection *Balfourianae*. We assigned taxa to each clade for dating and forced them to be monophyletic based on previously reported relationships (Montes et al. 2019; Montes et al. 2022). Within *Pinus* subsection *Cembroides*, *Pinus lindgrenii* from the Late Miocene Chalk Hills Formation of Idaho (5-6 MYA; Miller 1992) was recovered as sister to *P. monophylla* in the total evidence analysis, but was used to estimate the crown-age of pines from SW USA and Baja California (*P. californiarum*, *P. edulis*, *P. edulis* var. *fallax*, *P. monophylla*, and *P. quadrifolia* Parry ex Parl.) due to its sub-globose seed cone, ovoid-ellipsoid seed, and seed length (16-20 mm). The phylogenetic position of *Pinus sanjuanensis* from the Late Oligocene Creede Flora of Colorado (26.5 MYA; Axelrod 1986) was not resolved in our analysis, but the few morphological details of needles, seed cone, and enlarged, wingless seeds are distinctive of *Pinus* subsection *Cembroides*. This interpretation allowed us to assign *Pinus sanjuanensis* to the crown-group of *Pinus* subsection *Cembroides* (Fig. 1), which is the most inclusive group of taxa that contains all extant members of the subsection. *Pinus crossii* from the Late Oligocene Creede Flora of Colorado (26.5 MYA; Axelrod 1986) was used to calibrate the most-recent common ancestor node (MRCA) between subsection *Balfourianae* and *Pinus* subsection *Cembroides* (section *Parrya*) due to the shape and morphology of its needles. We also calibrated the crown-node of section *Quinquefoliae* with *P. florissantii* Lesquereux from the Early Oligocene (34 MYA; Axelrod 1986) due to the terminally-positioned umbo on the ovulate cone scale considered a morphological character of subsection *Strobus*. Finally, we used the age *Pinuxylon* spp. Meijer from the Late Cretaceous Aachen Formation of Belgium (Santonian: 83.5-85.8 MYA; Meijer 2000) only for calibrating the root-node. This fossil has been employed to calibrate the divergence between *Pinus* subgenus *Strobus* and *Pinus* subgenus *Pinus* (Willyard et al. 2007; Saladin et al. 2017; Jin et al. 2021).

We employed a uniform distribution for all calibrated nodes. Both minimum and maximum ages of all fossils were derived from stratigraphic intervals using the

International Chronostratigraphic Chart ver. 2020/03 (Cohen et al. 2013). In *P. lindgrenii* the minimum boundary was 5.3 MYA and the maximum boundary was 7.2 MYA from the Messinian stage. We designated a minimum age of 23.03 MYA to *Pinus sanjuanensis* and *P. crossii* from the Chattian stage and a maximum age of 28.1 MYA while in *P. florissantii* the minimum age was 27.8 MYA and the maximum was 33.9 MYA from the Rupelian stage. Based on the *Pinuxylon* sp. the minimum age of the root-node was 83.6 MYA from the Santonian stage and the maximum age was 86.3 MYA.

Two independent runs for 300 M generations of MCMCs were performed with sampling every 100,000 trees. We used Tracer to confirm chain convergence and an adequate effective sample size (ESS) and LogCombiner ver. 2.4.8 (Bouckaert et al. 2014) to integrate the resulting trees sampled from the MCMCs. We employed TreeAnnotator ver. 2.4.8 (Bouckaert et al. 2014) to annotate the maximum credibility tree with mean values. The tree was imported into FigTree for further editing.

Estimated evolutionary dynamics— We assessed the diversification rates in *Pinus* subsection *Cembroides* using Bayesian Analysis of Macroevolutionary Mixtures ver. 2.5.0 (BAMM; Rabosky et al. 2013). BAMM is a framework that infers complex evolutionary dynamics of diversification through time and among lineages. BAMM explores different rate configurations in proportion to their posterior probability. We used the time-calibrated tree as input to BAMM for assessing diversification rates. The analysis was conducted using a control file obtained with BAMMtools ver. 2.1.7 (Rabosky et al. 2014), implemented in Rstudio ver. 4.1.0 (RStudio team, 2019; http://www.rstudio.com). The control file includes the priors, model type (speciation/extinction), and MCMC conditions. The analysis was run using four chains and heating of 0.1. Fifty million generations were performed with sampling every 1,000 generations, discarding 0.1 as a burn-in fraction (Recommended by authors). We analyzed the MCMC convergence and effective sample sizes using the R package “coda” ver. 0.19-4 (Plummer et al. 2006). We used Bayes factors package ver. 0.9.2 (Morey et al. 2018) for comparing the sampled

macroevolutionary models. We computed the posterior distributions of the best global rates diversification and plotted the rate of speciation onto a phylogram using BAMMtools. Finally, we assessed the common inherited evolutionary rate regimen using the posterior distribution of each sample. The evolutionary rate regimen (0 up to 1) was compared between species using a macroevolutionary cohort analysis in BAMM (Rabosky et al. 2014).

RESULTS

Molecular Data— A total of 207 low-copy nuclear genes were used to estimate both phylogeny and divergence times in *Pinus* subsection *Cembroides* and close relatives. The mean length of low-copy nuclear genes was 705.6 bp with a range of 297-1248 bp. We recorded an average pairwise identity of 98.30% with a range of 93%-99.8% and an average of 69.5% of identical sites with a range of 51.7%-95.9%. The concatenation of 207 nuclear genes resulted in an alignment with 40 sequences and 140,895 sites, 7868 which were parsimony informative.

Phylogenetic Position of Cembroides Fossils— The 22-taxon combined matrix with morphology and molecular characters included 13,331 total characters. The alignment included three fossil taxa, 19 extant species, 37 morphological characters, and 13,294 molecular sites. For the Bayesian analysis, the average standard deviation of split frequencies was 0.001215 at the end of the MCMC run. All nodes were resolved in the Bayesian majority-rule tree except the position of *P. sanjuanensis* and *P. remota*, and most relationships were moderately well-supported (pp >0.8; Fig. 1). The branch support of all three fossils was lower than the recovered values for the extant species in the Bayesian tree (<0.8 pp). *Pinus lindgrenii* and *P. sanjuanensis* retrieved within *Pinus* subsection *Cembroides* and *P. crossi* occurred within the subsection *Balfourianae* (Fig. 1). *Pinus crossi* was recovered as sister to *P. longaeva* (0.6 pp), whereas *P. lindgrenii* was sister to *P. monophylla* and formed a monophyletic group with *P. californiarum*, *P. edulis*,

P. *edulis* var. *fallax*, and *P*. *quadrifolia* (0.8 pp). *Pinus maximartinezii*, *P*. *pinceana*, and *P. rzedowskii* (big-cones clade) formed a monophyletic sister group to the small-cones clade (0.8 pp). *Pinus discolor* was sister to *P. johannis* and *P*. *culminicola* and the three taxa formed a monophyletic sister group to *P*. *cembroides*, *P*. *lagunae*, and *P*. *cembroides* subsp. *orizabensis* clade (0.7 pp) (Fig. 1).

Evolutionary Relationships— Maximum likelihood analysis of the concatenated alignment produced a well-supported phylogeny (Fig. 2). *Pinus* subsection *Cembroides* was monophyletic and subsects. *Balfourianae* from W USA and *Nelsoniae* (Mexico) were sister to *Pinus* subsection *Cembroides* (100% bs). Phylogenetic relationships among close relatives were well-supported (Fig. 2) where all species were recovered as monophyletic lineages except *P*. *edulis* and *P*. *edulis* var. *fallax* (Fig. 2). In *Pinus* subsection *Cembroides* two main clades were inferred, *P*. *maximartinezii*, *P*. *pinceana*, and *P*. *rzedowskii* distributed exclusively in Mexico and a second clade comprising the remaining species, species distributed from SW US to South-Central of Mexico (Critchfield and Little 1966). Not all relationships in *Cembroides* were well-supported (>90% bs). For instance, *P*. *remota* and sister group (58% bs) and the position of *P*. *edulis* (UT3) in the phylogeny (62% bs) (Fig. 2). Furthermore, the phylogenetic positions of *P*. *californiarum*, *P*. *edulis*, and *P*. *edulis* var. *fallax* were unclear. This last result was consistent with other phylogenetic analyses with Bayesian and maximum likelihood that do not consider ILS (Montes et al. 2019).

Estimated Divergence Ages—To infer the crown-age, stem-age, and diversification dynamics in *Pinus* subsection *Cembroides*, we included all extant taxa and generated time-calibrated trees based on morphological synapomorphies and total-evidence analysis (Fig. 1). All relationships in the clock-tree were well-supported (0.9 pp). All estimated intervals corresponding to the 95% highest posterior density (HPD). Calibration with *Pinuxylon* sp. fossil on the root-node (85 MYA) suggests that the origin of

pinyon pines from North America was dated to the Late Oligocene (about 25.5 MYA; 27.7–23.2 MYA) whereas their diversification occurred during the Middle Miocene about 14.0 MYA (19.0–9.9 MYA). This is the stem-age recorded for the small-cone and big-cone species of *Pinus* subsection *Cembroides* (Fig. 3). The age estimate for the MRCA of subsections *Balfourianae* and *Nelsoniae* was 22.1 MYA during the Early Miocene (26.9–13.7 MYA) and the diversification of *Balfourianae* occurred during the Middle Miocene about 13.3 MYA (20.0–3.4 MYA) whereas the age estimate for the origin of *Gerardianae*, *Krempfianae*, and *Strobus* subsections was 30.9 MYA during the Oligocene (33.8–28.0 MYA) (Fig. 3).

The clock-trees estimated using node-dating showed that all extant species in *Pinus* subsection *Cembroides* diverged during the Neogene period (Fig. 3). Specifically, eleven pinyon pines species from North America diverged during the Miocene epoch and *P. discolor*, *P. johannis*, *P. monophylla*, and *P. quadrifolia* during the Last Pliocene (Fig. 3). Two main clades have diversified within *Pinus* subsection *Cembroides*: one composed of 3 species, *P. maximartinezii*, *P. pinceana*, and *P. rzedowskii* (big-cone species) about 9.3 MYA (13.3–6.1 MYA) and another of 12 taxa (small-cone species) about 8.8 MYA (12.1–7.0 MYA). Within the small-cone clade, two groups of pinyons have diverged, the pinyons from SW US and Baja California and the pinyons from North to South-Central of Mexico (Fig. 3). The most recent diverging species of *Pinus* subsection *Cembroides* were *P. discolor* and *P. johannis* about 4.6 MYA (6.6–2.8 MYA) and *P. monophylla* and *P. quadrifolia* from SW US and Baja California about 5.0 MYA (6.0–3.3 MYA). The divergence-time recorded for *P. cembroides* and *P. lagunae* was about 5.6 MYA during the Miocene (7.2–4.0 MYA). Also, in the Miocene *P. californiarum* diverged from *P. monophylla* and *P. quadrifolia* about 6.1 MYA (6.8–4.9 MYA). For the first time, this work includes individuals of *P. californiarum*, *P. lagunae*, *P. cembroides* subsp. *orizabensis*, and *P. edulis* var. *fallax* in molecular clocks and diversification analysis (Fig. 3).

Estimated evolutionary dynamics— We assessed the convergence of MCMC and plotting the log-likelihood trace of Markov chains (Appendix S2). The effective sample size of the log-likelihood was -23314.09. The analysis of rate shifts recovered 13 potential shift configurations (set of locations through the phylogeny that were sampled) based on posterior probabilities (Appendix S3A). The best model recovered with the highest Bayes factor compared to the null model (no rate shifts) was in favor of the single rate shift (Appendix S3B). The probability of this single rate implied strong evidence over another model with a Bayes factor value of 36.2 (Appendix S4). The phylorate plot showed an increment in the speciation rate in *Pinus* subsection *Cembroides* and a deceleration in speciation rates in close relatives (Fig. 4).

Additionally, BAMM recovered two distinct shift configurations (Fig. 5). The results showed that 44% of the posterior distribution was assigned to a single major increase in speciation rate in *Pinus* subsection *Cembroides*, whereas 56% of the posterior distribution was assigned to zero shifts (Fig. 5).

All speciation and extinction rates estimated were non-significant but the extinction rates were lower than speciation rates (Table 2). In general, the extinction rates have remained constant throughout time in the phylogeny as well as within *Pinus* subsection *Cembroides* (Fig. 6). In comparison, the speciation rate showed a slight increase throughout time (Fig. 6). The speciation rate for the crown-node of *Pinus* subsection *Cembroides* was 0.160 species/MYA (CI = 0.109–0.221) whereas extinction rate was 0.034 species/MYA (CI = 0.002–0.099). The speciation rate for subsections *Nelsoniae* and *Balfourianae* was 0.106 species/MYA (CI = 0.026–0.189) whereas extinction rate was 0.044 species/MYA (CI = 0.002–0.122) (Table 2).

The patterns of the evolutionary rate regimes between *Pinus* subsection *Cembroides* and close relatives can be visualized in cohort analysis of macroevolutionary regimes (Fig. 7). All three major cohorts (*Cembroides*, *Nelsoniae* + *Balfourianae*, and *Gerardianae* + *Krempfianae* + *Strobus* clades) were well-supported. Only *Cembroides* and *Nelsoniae* + *Balfourianae* clades share a common macroevolutionary rate regime but

with a different pairwise probability (Fig. 7). *Pinus* subsection *Cembroides* species are characterized by dynamics that are more correlated with *Nelsoniae* and *Balfourianae* species whereas *Cembroides* and *Gerardianae* + *Krempfianae* + *Strobus* cohort do not share a common macroevolutionary rate regime and dynamics in this group are not correlated (Fig. 7).

DISCUSSION

Phylogenetic Position of Cembroides Fossils— There is an extensive collection available of *Pinus* fossils (Axelrod 1986, 1987; Millar 1993) but most reported specimens are incomplete or poorly preserved. Fossils that have been assigned to *Pinus* subsection *Cembroides*, lack morphological information and the close affinities with extant species are poorly understood. For instance, *P. ballii* Brown from Green River, Colorado (47 MYA) was interpreted by Axelrod (1986) as the oldest record for *Pinus* subsection *Cembroides* but Wolfe and Schorn (1990) argued that the putative ovulate cone of *P. ballii* is not a pine. Despite incomplete fossils, the total-evidence analysis allowed us to explore the phylogenetic position of both *P. lindgrenii* and *P. sanjuanensis* within *Cembroides* and *P. crossii* within *Balfourianae* (Fig. 1). On the one hand, our results support the hypothesis by Wolfe and Schorn (1990) that *P. lindgrenii* is related to *P. monophylla* (0.7 pp). *Pinus lindgrenii* has similar traits to extant species within *Pinus* subsection *Cembroides* such as seed cone length < 6 cm, scales and enlarged wingless seeds, whereas it shares synapomorphies with *P. monophylla*, *P. californiarum* (single-needle pinyons), and *P. edulis* (two-needle pinyon), including the ovoid-globose seed cones, prominent apophysis, and ovoid-ellipsoid seed shape. Although our knowledge of morphological characters in *P. lindgrenii* is limited, this fossil can be used to calibrate the MRCA of the *P. edulis*-single-needle pinyons clade. Some studies have used *P. lindgrenii* to calibrate the MRCA of a *P. edulis*-*P. johannis* clade (Saladin et al. 2017) but the relationship between *P. edulis* and *P. johannis* as sister has not been supported by recent coalescent phylogenetic analyses with multilocus data (Montes et al. 2019; Montes et al. 2022). In addition, *P. lindgrenii* was

described from Idaho and *P. monophylla* is also distributed in southern Idaho (Critchfield and Little 1969) ~200 km from locality of *P. lindgrenii* in Glenns Ferry Formation.

On the other hand, the phylogenetic position of *P. sanjuanensis* was not resolved in this study. Nonetheless, Saladin et al. (2017) has used to *P. sanjuanensis* to calibrate the MRCA of *Pinus* subsection *Cembroides* and following this philosophy, *P. sanjuanensis* was assigned to the crown-group of pinyon pines from North America (Fig. 1), suggesting that the kind of needles, seed cones, and enlarged wingless seeds, representative of *Cembroides* (except for *P. rzedowskii*) are present from the Oligocene (Axelrod 1986).

Estimated Divergence Ages for Pinyon Pines—Divergence times in pines have been estimated using small number of loci (Willyard et al. 2007; Gernandt et al. 2008; Hernández-León et al. 2013; Saladin et al. 2017) and a larger data set of transcriptome sequences (Jin et al. 2021). Although the focal group in those previous studies was not pinyon pines from North America, the estimates of the divergence ages for *Pinus* subsection *Cembroides* varies widely. Furthermore, the number of species sampled in this subsection has varied, as well as the approaches for estimating the divergence ages. Therefore, our results on the divergence dynamics in *Pinus* subsection *Cembroides* are partially comparable.

Our results do not support the hypothesis by Axelrod (1986) who suggests that the origin of the *Pinus* subsection *Cembroides* goes back to the Middle Eocene according to fossil record (46 MYA). Likewise, the hypothesis defended by Millar (1993) who suggests that *Pinus* subsection *Cembroides* have evolved in refugia during the Eocene is not supported because our estimates are younger than proposes previously mentioned and are closer with small number of loci results than genomic data, suggesting that both numbers of fossils and calibrated nodes have more impact than molecular data. This behavior also has been reported in pines by Willyard et al. (2007). Other causes that impact the time-calibrated phylogenies in pines are the approaches. Saladin et al (2017) compared the results between node-dating and fossilized birth-death model and they

observed differences in the estimated ages in pines due mainly to placement of fossils and definitions of priors. However, it is important to clarify that the taxonomic range in this study was at the subsection level whereas the studies of Saladin et al (2017) and Jin et al (2021) were at the genus level. The different results obtained in our study could be due to multiple causes. For instance, a greater species number could be a reason for obtaining older ages to the divergence of *Pinus* subsection *Cembroides* from its extant sister and species divergence within small-cone clade in studies at the genus level (Table 3). According to Poux et al. (2008), estimated ages in a phylogeny tend to get older when increase taxon sampling. Evolutionary histories also play an important role in the estimated ages and the interpretation of morphological evolution. Saladin et al. (2017) recovered *P*. *rzedowskii* as sister to *P*. *monophylla* and *P*. *quadrifolia*, and not as sister to *P*. *pinceana* and *P*. *maximartinezii* (Montes et al. 2019; Jin et al., 2021; Montes et al. 2022). *Pinus rzedowskii* is endemic from Michoacan, Mexico and is the only one species with functional wing seed in *Cembroides*. *Pinus rzedowskii*, *P*. *pinceana*, and *P*. *maximartinezii* are restricted to Mexico and together comprise the well-support big-cone pinyon pines clade (Montes et al. 2019; Jin et al., 2021; Montes et al. 2022). *Pinus monophylla* and *P*. *quadrifolia* are small-cone species with enlarged seeds that are functionally wingless and restricted to Baja California and SW US.

Particularly, our origin and diversification estimated ages in *Pinus* subsection *Cembroides* tend to be younger, similar to ages reported by Gernandt et al. (2008), who reported a stem-age of about 27 MYA in the Oligocene and a crown-age of about 11 MYA in the Miocene. Respecting genomic data, both stem and crown-ages reported in *Pinus* subsection *Cembroides* are older using the fossilized birth-death approach (see Jin et al. 2021). However, regardless of the data set most of the species *Pinus* subsection *Cembroides* diverged in the Miocene except for *Pinus discolor*, *P. johannis*, *P*. *monophylla*, and *P*. *quadrifolia*, which diverged in the Pliocene about 5.0 MYA.

The recent divergence of *Pinus discolor* and *P*. *johannis* around 5.0 MYA could explain the taxonomic confusion when treating these taxa as independent species

because *P*. *discolor* and *P*. *johannis* have not yet developed clear morphological differences (Robert, 1978; Little, 1968; Bailey and Hawksworth, 1983). This suggests that *P*. *discolor* and *P*. *johannis* are in the "gray zone" of speciation where morphology is almost invariable between lineages and genetic information is shared between lineages (Mayr 1963). Recently, a high level of gene conflict due to ILS have been reported in *P*. *discolor* and *P*. *johannis* (Montes et al. 2022). Nonetheless, the genetic evidence based on coalescent methods indicates that the two taxa present sufficient molecular divergence to be treated as independent lineages. Similarly, *P*. *monophylla* and *P*. *quadrifolia* also have recently diverged but unlike the two taxa mentioned above, *P*. *monophylla*, and *P*. *quadrifolia* are morphologically distinct (Bailey 1987). Studies indicate that *P*. *quadrifolia* show signs of admixture with some individuals of *P*. *monophylla* (Montes et al. 2019; Buck et al., 2020). However, evidence based on coalescent delimitation analyses suggest that *P*. *monophylla* and *P*. *quadrifolia* are two independent species (Montes et al. 2022). Another species that has not yet developed many morphological differences and shares similar traits with *P*. *monophylla* is *P*. *californiarum* (another one-needle species). Our results support *P*. *californiarum* as a lineage that first diverged from *P*. *monophylla* and *P*. *quadrifolia* about 6.1 MYA during the Miocene. *P*. *californiarum* and *P*. *monophylla* are adapted to similar precipitation conditions (Cole et al. 2008). This ecological trait together with their recent divergence could explain the morphological overlapping with some variants (i.e. *P. edulis* var. *fallax*) in their needle anatomy (i.e. the number of resin canals) in both single-needle pinyon pines (see Bailey, 1987). Nevertheless, phylogenetic analyses with multilocus data revealed that *P*. *californiarum* is distinct from *P*. *monophylla* (Montes et al. 2019; Montes et al. 2022). Although the focus groups in this study were not *Pinus* subsection *Balfourianae* and *Nelsoniae*, the estimates of the divergence ages reported are closer to those inferred with a small number of loci by Gernandt et al. (Figs 3 and 4; 2008).

Pinus cembroides *hypothesis*— Our results do not support the idea that *P. lagunae* is the most ancient living member of *P. cembroides* or a relict taxon as proposed by Axelrod (1986). These two species diverged during the late Miocene about 5.6 MYA (Fig. 3). Robert-Passini (1981) and Bailey (1983) considered *P. lagunae* a subspecies or a variety of *P. cembroides* with a disjunct distribution). Montes et al. (2022) concluded that *P. lagunae* is an independent sister species of *P. cembroides*. *Pinus lagunae* is restricted to the Cape Region of Baja California whereas *P. cembroides* is widely distributed from SW US to south-central of Mexico (Critchfield and Little 1969; Malusa 1992). The presence of *P. lagunae* in the Cape Region could be associated with the geographical history of Baja California Peninsula that occurred during the late Miocene to Middle Pleistocene 5.5 MYA ago. The divergence times between *P. lagunae* and *P. cembroides* seem to exhibit a geographic pattern that occurred before the southern gulf vicariance event about 4.0 MYA (Riddle et al. 2000). This geographic pattern has also been reported in other Baja California Penninsula Desert biota species (Riddle et al 2000; McCauley et al. 2010). Two other species, *P. quadrifolia* and *P. californiarum*, are also distributed in the peninsula farther north, but these are not sister to *P. lagunae* (Montes et al. 2019; Montes et al. 2022).

***Dynamics of diversification in* Pinus *subsection* Cembroides—** Studies have reported a robust phylogenetic hypothesis with all extant species and clear delimitation of species in *Pinus* subsection *Cembroides* (Montes et al. 2019; Montes et al. 2022). Species delimitation studies have recognized as independent lineages some species that previously had not been included in time-calibrated phylogenies of North American pinyon pines (Montes et al. 2022). The incorporation of all extant species allowed us to explore the diversification dynamics in *Cembroides* and close relatives. Our results are congruent in reporting an accumulation of species (a net diversification rate) of pinyon pines during the Miocene (Gernandt et al. 2008; Saladin et al. 2017; Jin et al. 2021). Accelerated diversification of coniferous species adapted to arid land in North America during the

Miocene to Pliocene has also occurred in genera such as *Juniperus* L. (Mao et al. 2010) and *Ephedra* L. (Loera et al. 2012). The Miocene has been considered an epoch where terrestrial ecosystem experienced drying (Herbert et al. 2016). According to Jin et al. (2021), the aridity index may have played a decisive role for the rates shifts of niche evolution in pines during the Miocene. This suggests that the aridity index might play an important role in the diversification of *Pinus* subsection *Cembroides* since all species are distributed in arid and semiarid environments (Critchfield and Little 1966; Malusa 1992).

The speciation rates in *Cembroides* show a single shift configuration with BAMM (Fig. 4) suggesting a key innovation. Notably, our results suggest that the greater number of lineages in *Cembroides* compared to the only two subsections within section *Parrya* (*Nelsoniae* and *Balfourianae*) depended on low extinction rates (Fig. 6) causing an accumulation of lineages. This behavior of rates has been recently reported for all *Pinus* genus using the same approach by Jin et al. (2021). In contrast, the level speciation rates to subsection-level are high in the early evolutionary history of pinyon pines but are not higher compared to angiosperms (Crepet and Niklas 2009). The results of rates through time indicate a non-accelerated speciation rate in *Cembroides* and a decelerated extinction rate during the Miocene (Fig. 6). This pattern is consistent with the results reported recently for the entire genus. However, speciation rates were lower in this study than those reported by Jin et al. (Fig. S11B; 2021). In general, in conifers, the diversification rate has been slowing since the late Paleogene (Fig. 2B; Condamine et al. 2020).

Finally, we need to incorporate ecological characters to understand more about the diversification of pinyon pines. Our results provide a temporal context in which to interpret the diversification dynamics of *Pinus* subsection *Cembroides.*

ACKNOWLEDGMENTS

The authors are very grateful to María Inés Badillo for participating in the valuable revision of manuscript. We also would like to thank Cristian Cervantes for providing computational

assistance. We are very grateful to Jorge Pérez de la Rosa and Abisaí García for collecting and providing *Pinus lagunae* specimens, and Laura Figueroa for collecting and providing *Pinus pinceana* specimens. We also thank Dra. Lidia I. Cabrera and the LANABIO of the Instituto de Biología, UNAM. This project was funded by PAPIIT-DGAPA, UNAM Grant IN209816 and CONACyT Grant 221694.

Author Contributions

JRM performed the phylogenetic analyses and diversification analysis, and was the primary author for the manuscript. JRM, ABV, and RHG inferred total-evidence and the molecular clocks analyses. JRM, ABV, RHG, and DSG designed the study. JRM and DSG participated in field work. All authors reviewed and edited the manuscript.

Literature Cited

Axelrod, D. I. 1986. Cenozoic history of some western American pines. *Annals of the Missouri Botanical Garden* 73: 565–641.

Axelrod, D. 1987. The late Oligocene Creede Flora, Colorado. Berkeley, California: University of California Press.

Bailey, D. K. 1983. A new allopatric segregate from and a new combination in *Pinus cembroides* Zucc. at its southern limits. *Phytologia* 54: 89–100.

Bailey, D. K. 1987. A study of *Pinus* subsection *Cembroides* I: The single needle pinyons of the Californias and the Great Basin. *Notes from the Royal Botanic Garden Edinburgh* 44: 275–310.

Bailey, D. K., and F. G. Hawksworth. 1983. Pinaceae of the Chihuahuan Desert region. *Phytologia* 53: 226–234.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* 10: e1003537.

Buck, R., S. Hyasat, A. Hossfeld, and L. Flores-Rentería. 2020. Patterns of hybridization and cryptic introgression among one- and four-needled pinyon pines. *Annals of Botany* 126: 401–411.

Critchfield, W. B. and E. L. Little. 1966. *Geographic Distribution of the Pines of the World*. Washington, D.C: Department of Agriculture Miscellaneous Publication 991.

Cobb, N. S., R. T. Trotter, and T. G. Whitham. 2002. Long-term sexual allocation in herbivore resistant and susceptible pinyon pine (*Pinus edulis). Oecologia* 130: 78–87.

Cohen, K.M., S. C. Finney, P. L. Gibbard, and J.-X. Fan. 2013. International Chronostratigraphic Chart. *The International Commissions on Stratigraphy* 36: 199-204.

Cole, K. L., J. F. Fisher, S. T. Arundel, J. Cannella, and S. Swift. 2008. Geographical and climatic limits of needle types of one-and two-needled pinyon pines. *Journal of Biogeography* 35: 257–269.

Condamine, F. L., D. Silvestro, E. B. Koppelhus, and A. Antonelli. 2020. The rise of angiosperms pushed conifers to decline during global cooling. *Proceedings of the National Academy of Sciences of the United States of America* 117: 28867–28875

Cusimano, N. and S. S. Renner. 2010. Slowdowns in diversification rates from real phylogenies may not be real. *Systematic Biology* 59: 458-64.

Crepet, W. L. and K. J. Niklas. 2009. Darwin's second “abominable mystery”: Why are there so many angiosperm species? *American Journal of Botany* 96: 366–381.

Degnan, J. H. and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* 24: 332–40.

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

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: 0699–0710.

Eckenwalder, J. E. 2009. *Conifers of the World: The Complete Reference*. Portland, Oregon: Timber Press.

Erwin, D. M. and H. E. Schorn. 2005. Revision of the conifers from Eocene Thunder Mountain flora, Idaho, U.S.A. *Review of Palaeobotany and Palynology* 137: 125–145.

Farjon, A. and B. Styles. 1997. *Flora Neotropica Monograph*: Pinus (*Pinaceae*). New York: The New York Botanical Garden.

Flores-Rentería, L., A. Wegier, D. Ortega Del Vecchyo, A. Ortíz-Medrano, D. Piñero, A. V. Whipple, F. Molina-Freaner, and C. A. Domínguez. 2013. Genetic, morphological, geographical and ecological approaches reveal phylogenetic relationships in complex groups, an example of recently diverged pinyon pine species (subsection *Cembroides*). *Molecular Phylogenetics and Evolution* 69: 940–949.

Floyd, M. 1983. Dioecy in five *Pinus edulis* populations in the southwestern United States. *The American Midland Naturalist* 110: 405–411.

Gernandt D. S., G. Gaeda López, S. Ortiz García, and A. Liston. 2005. Phylogeny and classification of *Pinus*. *Taxon* 54: 29–42.

Gernandt, D. S., G. Holman, C. Campbell, M. Parks, S. Mathews, L. A. Raubeson, A. Liston, R. A. Stockey, and G. W. Rothwell. 2016. Phylogenetics of extant and fossil Pinaceae: Methods for increasing topological stability. *Botany* 94: 863–884.

Gernandt, D. S., S. A. Magallón, G. Geada López, O. Zerón Flores, A. Willyard, and A. Liston. 2008. Use of simultaneous analyses to guide fossil-based calibrations of Pinaceae phylogeny. *International Journal of Plant Sciences* 169: 1086–1099.

Heath, T. A., S. M. Hedtke, and D. M. Hillis. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46: 239–257.

Herbert, T., K. Lawrence, A. Tzanova, L. C. Peterson, R. Caballero-Gill, and C. S. Kelly. 2016. Late Miocene global cooling and the rise of modern ecosystems. *Nature Geoscience* 9: 843–847.

Hernandez-León, S., D. S. Gernandt, J. A. Pérez de la Rosa, and L. Jardón-Barbolla. 2013. Phylogenetic relationships and species delimitation in *Pinus* section *Trifoliae* inferred from plastid DNA. *PLoS ONE* 8: e70501.

Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

Jin, W. T., D. S. Gernandt, C. Wehenkel, X. M. Xia, X. X. Wei, and X. Q. Wang. 2021. Phylogenomic and ecological analyses reveal the spatiotemporal evolution of global pines. *Proceedings of the National Academy of Sciences of the United States of America* 118.

Katoh, K. and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–80.

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

Kendall, D. G. 1948. On the generalized "birth-and-death" process. *The Annals of Mathematical Statistics* 19: 1–15.

Klaus, W. 1980. Neue beobachtungen zur morphologie des zapfens von *Pinus* und ihre bedeutung für die systematik, fossil-bestimmung, arealgestaltung und evolution der gattung. *Plant Systematics and Evolution* 134: 137–171.

Lanfear, R., B. Calcott, D. Kainer, C. Mayer, and A. Stamatakis. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology* 14: 82.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2016. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34: 772–773.

Lanner, R. M. 1981. *The Piñon Pine: A Natural and Cultural History*. Reno, Nevada: University of Nevada Press.

Ledig, F. T., M. A. Capó-Arteaga, Paul D. Hodgskiss, H. Sbay, C. Flores-López, M. T. Conkle, and B. Bermejo-Velázquez. 2001. Genetic diversity and the mating system of a rare Mexican piñon, *P. pinceana*, and a comparison with *P. maximartinezii* (Pinaceae). *American Journal of Botany* 88: 1977–1987.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic biology* 50: 913-925.

Little, E. L. 1968. Two new pinyon varieties from Arizona. *Phytologia* 17: 329–342.

Loera, I., V. Sosa, and S. M. Ickert-Bond. 2012. Diversification in North American arid lands: Niche conservatism, divergence and expansion of habitat explain speciation in the genus *Ephedra*. *Molecular Phylogenetics and Evolution* 65: 437–450.

Maddison, W. P. and D. R. Maddison. 2019. Mesquite: A molecular system for evolutionary analysis. Version. 3.61. http://www.mesquiteproject.org.

Madrigal, S. X. and D. M. Caballero. 1969. Una nueva especie mexicana de *Pinus*. *Boletín Técnico del Instituto Nacional de Investigaciones Forestales* 26: 1–11.

Malusa, J. 1992. Phylogeny and biogeography of the pinyon pines (*Pinus* subsect. *Cembroides*). *Systematic Botany* 17: 42–66.

Mao, K., G. Hao, J. Liu, R. P. Adams, and R. I. Milne. 2010. Diversification and biogeography of *Juniperus* (Cupressaceae): Variable diversification rates and multiple intercontinental dispersals. *New Phytologist* 188: 254–272.

Mayr, E. 1963. *Animal Species and Evolution*. Cambridge, Massachusetts: Harvard University Press.

McCauley, R. A., A. C. Cortés-Palomec, and K. Oyama. 2010. Distribution, genetic structure, and conservation status of the rare microendemic species, *Guaiacum unijugum* (Zygophyllaceae) in the Cape Region of Baja California, Mexico. *Revista Mexicana de Biodiversidad* 81: 745–758.

Meijer, J. J. 2000. Fossil woods from the Late Cretaceous Aachen Formation. *Review of Palaeobotany and Palynology* 112: 297–336.

Millar, C. I. 1993. Impact of the Eocene on the evolution of *Pinus* L. *Annals of the Missouri Botanical Garden* 80: 471–498.

Millar, C. I. 1998. Early evolution of pines. In D. M. Richardson [ed.], Ecology and biogeography of *Pinus*, 69–91. Cambridge: Cambridge University Press.

Miller, C. N. 1992. Structurally preserved cones of *Pinus* from the Neogene of Idaho and Oregon. *International Journal of Plant Sciences* 153: 147–154.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. *In* Proceedings of the gateway computing environments whorkshop (GCE), 1–8. Institute of Electrical and Electronics Engineers, New Orleans, Louisiana, USA. doi: 10.1109/GCE.2010.5676129.

Montes, J.-R., P. Peláez, A. Willyard, A. Moreno-Letelier, D. Piñero, and D. S. Gernandt. 2019. Phylogenetics of *Pinus* subsection *Cembroides* Engelm. (Pinaceae) inferred from low-copy nuclear gene sequences. *Systematic Botany* 44: 501–518.

Montes, J.-R., P. Peláez, A. Moreno-Letelier, and D. S. Gernandt. 2022. Coalescent-based species delimitation in North American pinyon pines using low-copy nuclear genes and plastomes. *American Journal of Botany* 109:1–21. DOI: 10.1002/ajb2.1847

Morey, R. D., J. N. Rouder, T. Jamil, S. Urbanek, K. Forner, and A. Ly. 2018. BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor

Plummer, M., N. Best, K. Cowles, K. Vines. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* 6:7–11.

Poux, C., O. Madsen, J. Glos, W. De Jong, and M. Vences. 2008. Molecular phylogeny and divergence times of Malagasy tenrecs: Influence of data partitioning and taxon sampling on dating analyses. *BMC Evolutionary Biology* 8: 1-16.

Price, R. A., A. Liston, and S. H. Strauss. 1998. Phylogeny and systematics of *Pinus*. In D. M. Richardson [ed.], Ecology and biogeography of *Pinus*, 49–68. Cambridge: Cambridge University Press.

Rabosky, D., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications* 4:1958.

Rabosky, D., M. Grundler, C. Anderson, P. Title, J. J. Shi, J. W. Brown, H. Huang, and J. G. Larson. 2014. BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* 5: 701–707.

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

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67: 901–904.

Richardson, D. M. 1998. *Ecology and Biogeography of* Pinus. United Kingdom: Cambridge University Press.

Riddle, B. R., D. J. Hafner, L. F. Alexander, and J. R. Jaeger. 2000. Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. *Proceedings of the National Academy of Sciences of the United States of America* 97: 14438-14443

Robert-Passini, M.-F. 1981. Deux nouveaux pins pignons du Mexique. *Adansonia* 1: 61–73

Robert, M. F. 1978. Un nouveau pin pignon Mexicain: *Pinus johannis* M.-F Robert. *Adansonia* 18: 365–373.

Saladin, B., A. B. Leslie, R. O. Wüest, G. Litsios, E. Conti, N. Salamin, and N. E. Zimmermann. 2017. Fossils matter: Improved estimates of divergence time in *Pinus* reveal older diversification. *BMC Evolutionary Biology* 17: 95.

Silba, J. 1990. A supplement to the international census of the coniferae, II. *Phytologia* 68: 7–78.

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

Swofford, D. L. 2002. PAUP* Phylogenetic analysis using parsimony (*and other methods), v. 4.0b10. Sunderland, Massachusetts: Sinauer Associates.

Wiens, J. J. 2006. Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics* 39: 34–42.

Wolfe, J. A. and H. Schorn. 1990. *Taxonomic revision of the Spermatopsida of the Oligocene Creede flora, Southern Colorado. U.S*. Geological Survey Bulletin 1923. Denver, Colorado.

Willyard, A., J. Syring, D. S. Gernandt, A. Liston, and R. Cronn. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution* 24: 90–101.

Yule, G. U. 1924. II.—A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis*. Philosophical Transactions of the Royal Society* 213: 21–87.

TABLE 1. Species recognized for *Pinus* subsection *Cembroides* and taxon sampling included in molecular clock studies.

All extant species	**Gernandt et al. 2008**	**Saladin et al. 2017**	**Jin et al. 2021**
Pinus californiarum D.K.Bailey	*P. cembroides*	*P. cembroides*	*P. cembroides*
Pinus cembroides Zucc.	*P. johannis*	*P. culminicola*	*P. culminicola*
Pinus cembroides subsp. *orizabensis* D.K.Bailey	*P. monophylla*	*P. discolor*	*P. discolor*
Pinus culminicola Andresen & Beaman	*P. pinceana*	*P. edulis*	*P. edulis*
Pinus discolor D.K.Bailey & Hawksw	*P. rzedowskii*	*P. johannis*	*P. maximartinezii*
Pinus edulis Engelm.		*P. maximartinezii*	*P. monophylla*
Pinus edulis var. *fallax* Little		*P. monophylla*	*P. pinceana*
Pinus johannis M.F.Robert		*P. pinceana*	*P. quadrifolia*
Pinus lagunae (Passini) D.K.Bailey		*P. quadrifolia*	*P. remota*
Pinus maximartinezii Rzed.		*P. remota*	*P. rzedowskii*
Pinus monophylla Torr. & Frém.		*P. rzedowskii*	
Pinus pinceana Gordon			
Pinus quadrifolia Parl. ex Sudw.			
Pinus remota (Little) D.K.Bailey & Hawksw.			
Pinus rzedowskii Madrigal, Caball.M.,			

TABLE 2. Estimated evolutionary rates in all pines of phylogeny, *Pinus* subsection *Cembroides* and sister group (rates/MYA). Speciation and extinction rates were estimated with 95% credibility.

Node	**Number of species**	**Age**	**Crown age**	**Speciation**		**Extinction**	
				Mean	95% CI	Mean	95% CI
Root	23	Cretaceous	84.6	0.103	0.072-0.143	0.045	0.009-0.100
Cembroides	15	Miocene	14	0.160	0.109-0.221	0.034	0.002-0.099
Nelsoniae+Balfourianae	4	Miocene	22.1	0.106	0.026-0.189	0.044	0.002-0.122

TABLE 3. Sampling strategies reported in several molecular clock studies.

	Gernandt et al. 2008	**Saladin et al. 2017**	**Jin et al. 2021**	**This study**
Focal group	Genus	Genus	Genus	Subsection
Species number	49	115	112	23
Fossils	6	21	16	5
Individual per species	1	1	2 to 5	1 to 2
Genes	2	8	1662	207
Nucleotide sites	2838	5866	2,146,621	140,835
Divergence of *Cembroides* node	11 MYA	~23 MYA	~25 MYA	14
Divergence of *Nelsoniae+Balfourianae* node	22 MYA	~32 MYA	~35 MYA	22.1
Divergence of *Parrya* node	27 MYA	~43 MYA	~39 MYA	25.5

FIG. 1. Total-evidence tree based on nuclear genes and morphology. Results were inferred based using Bayesian inference. Bold type represents the fossils. Posterior probability values > 0.5 are shown on branches. Vertical bars correspond to each taxonomic subsection of section *Parrya*. Blue: *Cembroides*, green: *Balfourianae*, and orange: *Nelsoniae*.

FIG. 2. Maximum likelihood tree based on 207 low-copy nuclear genes. Results were inferred based on 40 samples. Bootstrap values > 50% are shown on branches. Vertical bars correspond to taxonomic subsections of sections *Parrya* and *Quinquefoliae*. Blue: *Cembroides*, green: *Balfourianae* (*Balf*), orange: *Nelsoniae* (*Ns*), red: *Gerardianae* (*Gd*), violet: *Krempfianae* (*Kp*), and gray: *Strobus* (*St*). The sample names indicate the state of collection. The locality ID is based in ISO 3166-2, RENAPO, MX.

FIG. 3. Inferred molecular dating phylogeny using 207 low-copy nuclear genes. The values shown on nodes represent the mean divergence and the purple bars represent 95% confidence interval of node age. Black circles correspond to nodes with 1.0 posterior probability and red circles to nodes with 0.8-09 pp. The lower scale delimits era and periods in millions of years. Quat = Quaternary.

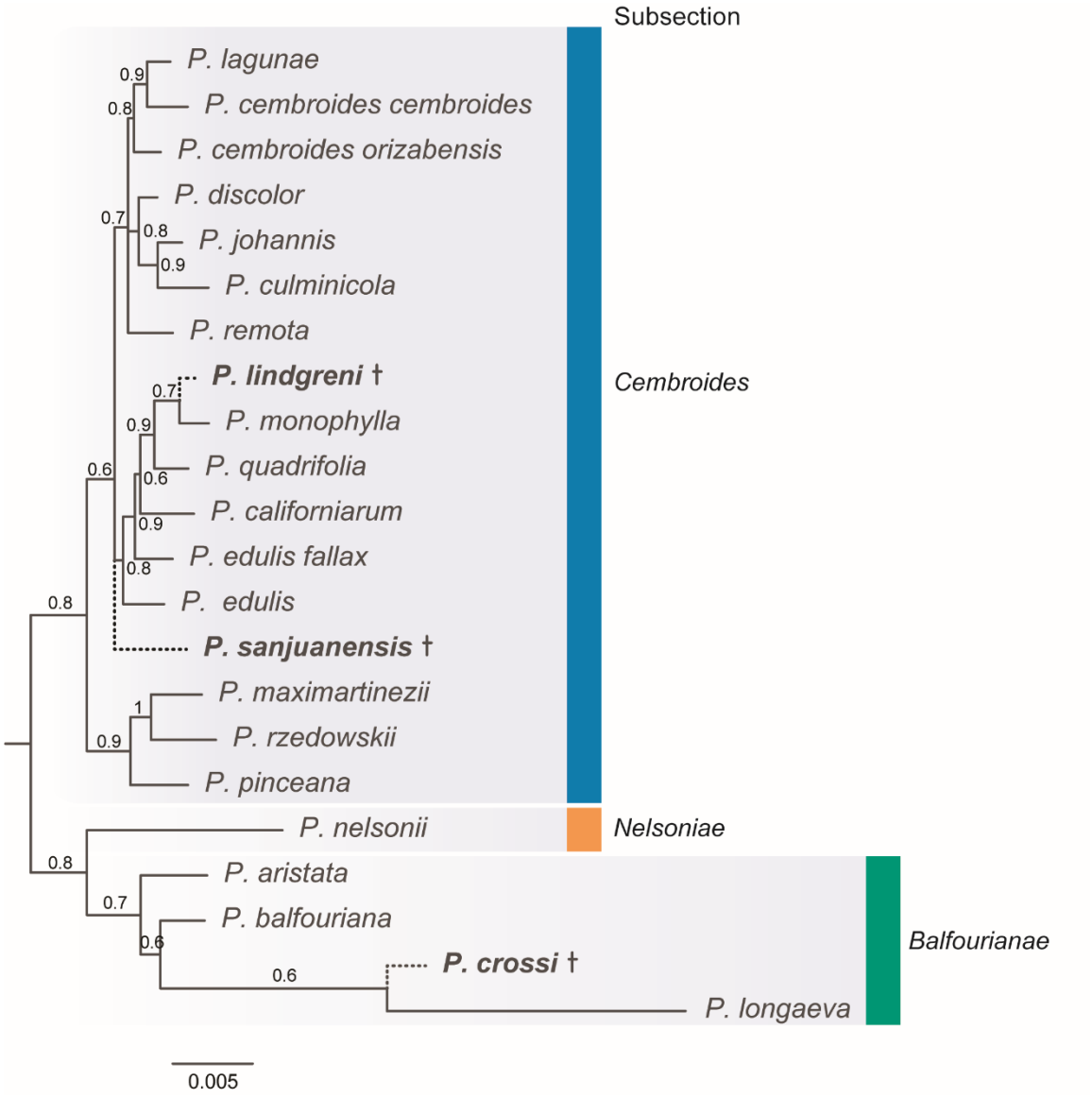
FIG. 4. Inferred phylorate plot under Bayesian Analysis of Macroevolutionary Mixtures framework. This shift configuration was sampled multiple times during simulations of posterior (f = 44%) with 95% credibility. Colored branches showing the level of speciation rate (cool color = slow, warm = fast). Each distinct color section of branch denotes the median diversification rate as summarized from its BAMM posterior probability. *Pinus* subsection *Cembroides* is characterized by a significant increase in the rate speciation (grey circle). The lower scale delimits era and periods in millions of years. Quat = Quaternary.

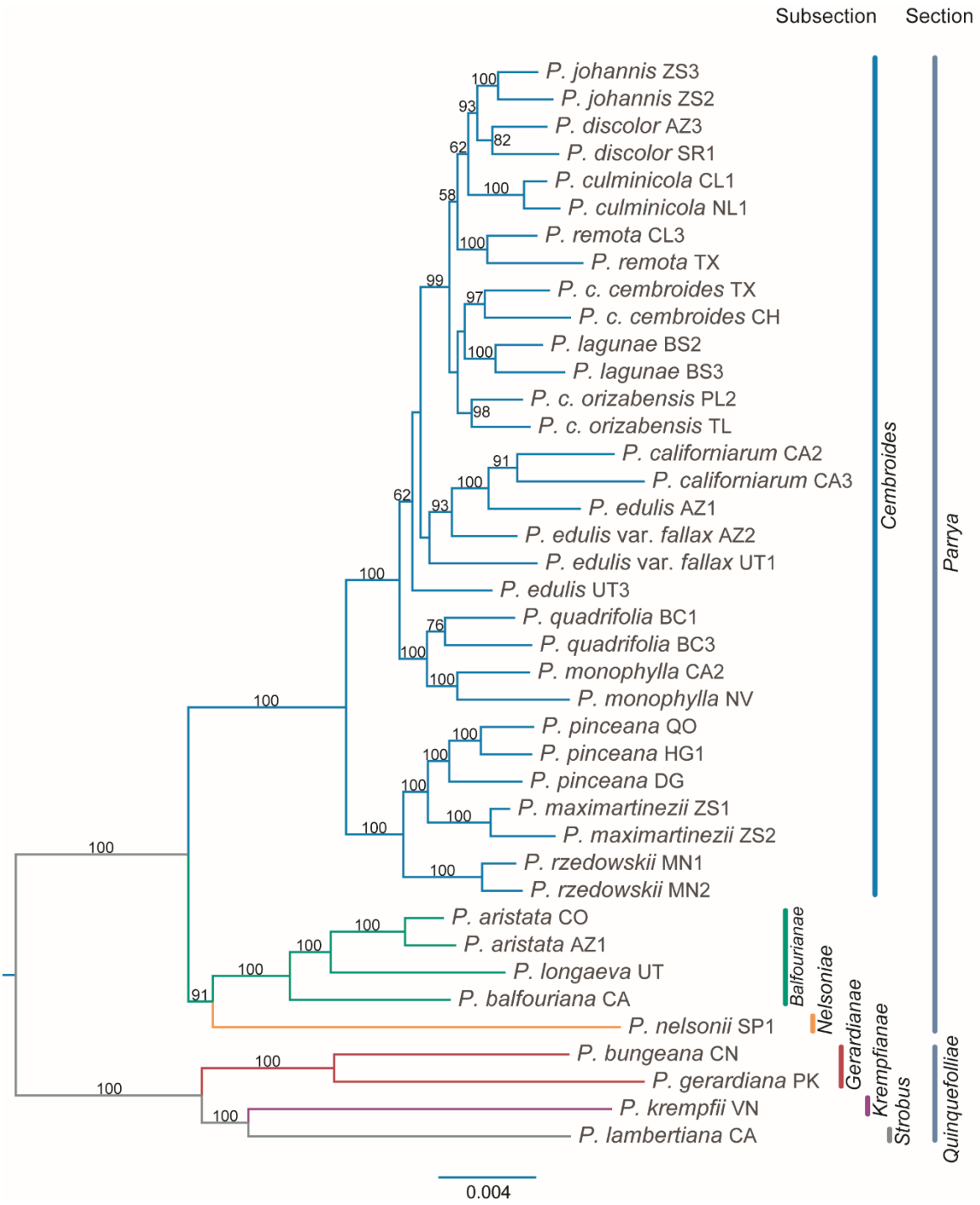
FIG. 5. Distinct shift configurations inferred using BAMM analysis. The chronograms represent the two most probable shift configurations of 13 sampled hypotheses. Each shift configuration was sampled multiple times during simulations of posterior (frequency) with 95% credibility A) Chronogram without increase of speciation rate in *Pinus* subsection

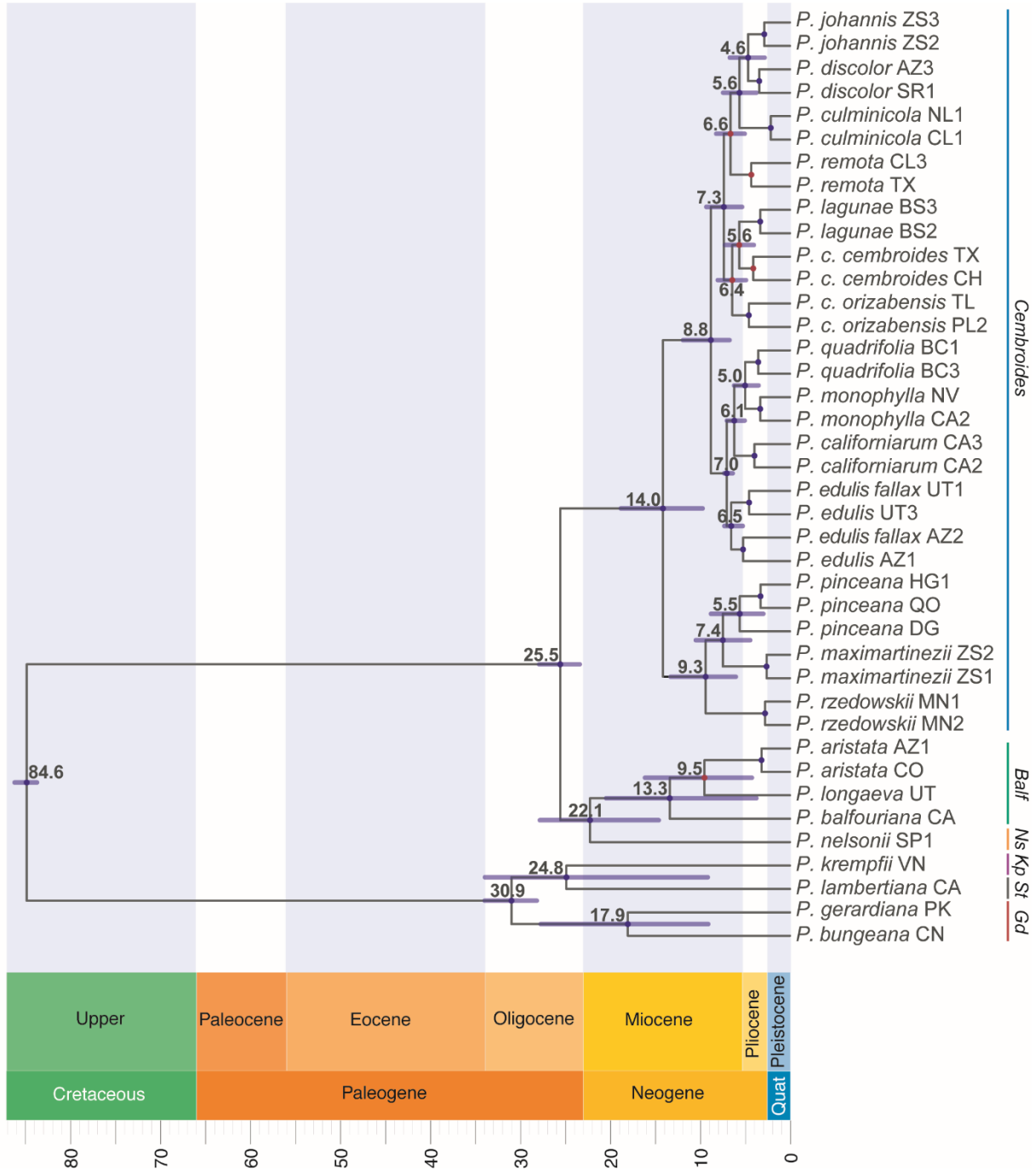
Cembroides (null model). B) Chronogram with one significant increase of speciation rate in *Pinus* subsection *Cembroides*.

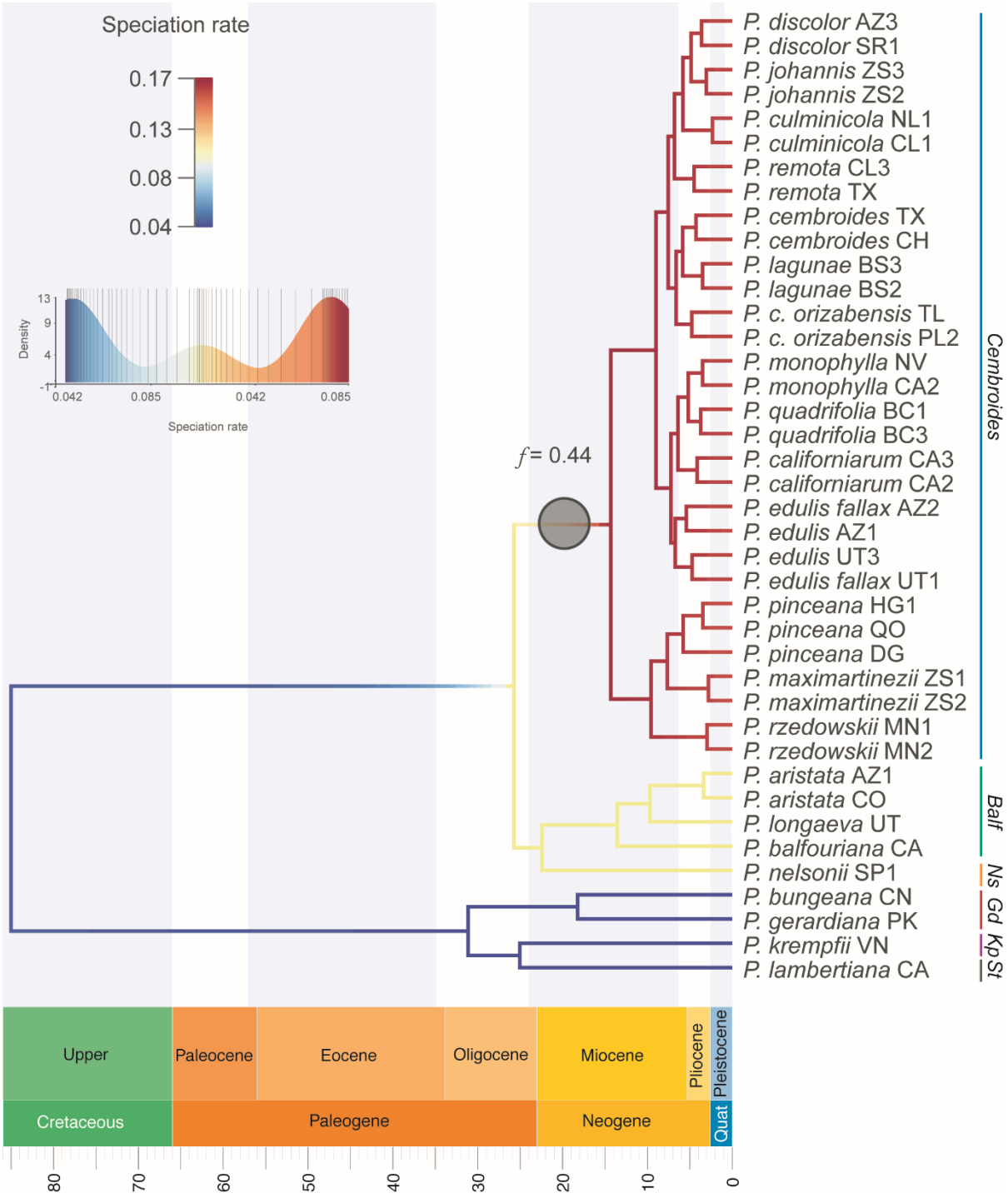
FIG. 6. Evolutionary rates through time for *Pinus* subsection *Cembroides* and close relatives. The lines give median values from posterior distribution, while color density shading denotes the confidence intervals with 95% on evolutionary rate reconstructions. Each distinct plot shows both speciation rates and extinction rates. The green lines denote the speciation rates, and the pink lines denote the extinction rates. Paleo = Paleocene, Oligo = Oligocene, Plio = Pliocene, Pleisto = Pleistocene, and Quat = Quaternary.

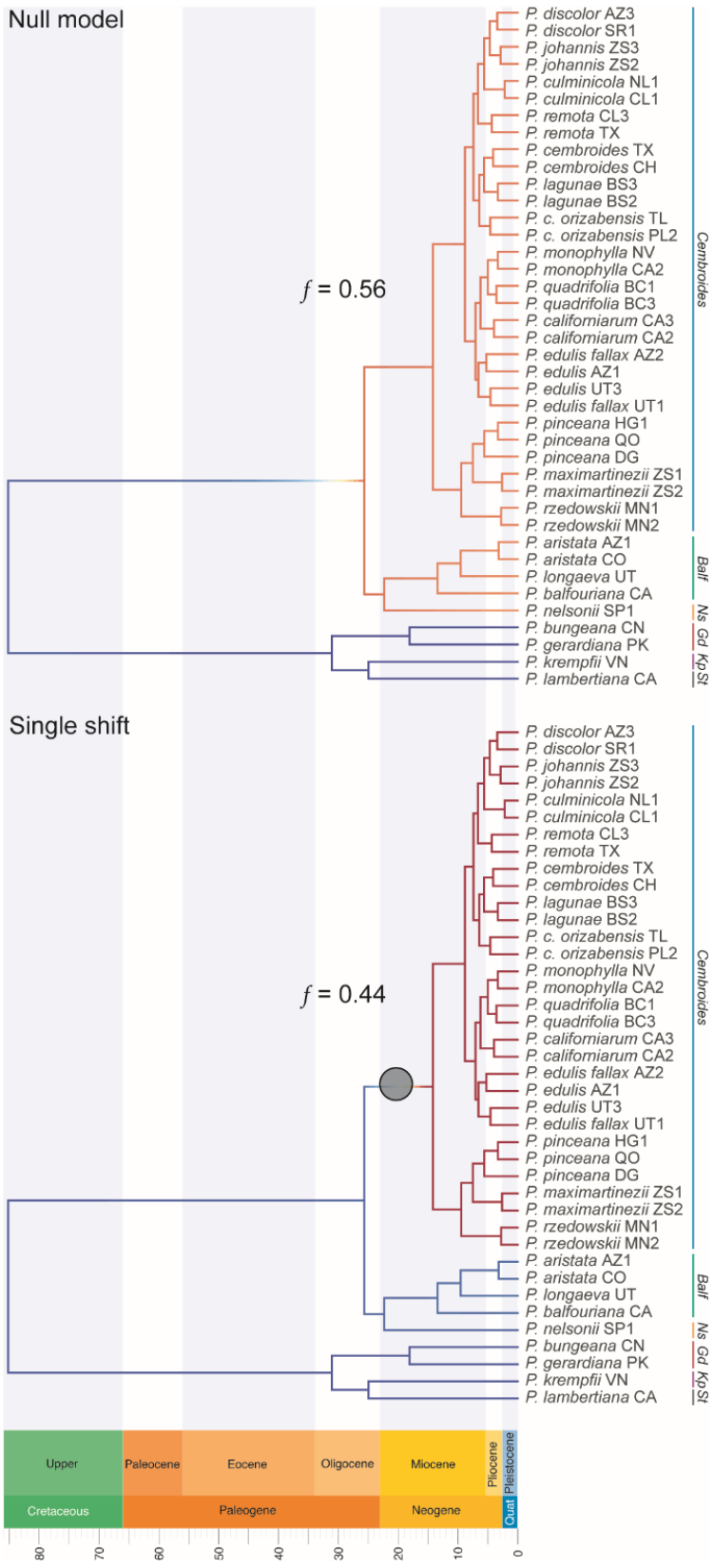
Fig. 7 Macroevolutionary cohort analysis for *Pinus* subsection *Cembroides*. For speciation three major macroevolutionary rate regimes were identified clearly. Only subsections *Cembroides* and *Nelsoniae* + *Balfourianae* share a common macroevolutionary rate regime but with different pairwise probabilities. The scale is pairwise probability (0.0 to 1.0). *Balfourianae* (*Balf*), *Nelsoniae* (*N*), *Gerardianae* (*G*), *Krempfianae*, and *Strobus* (*S*).

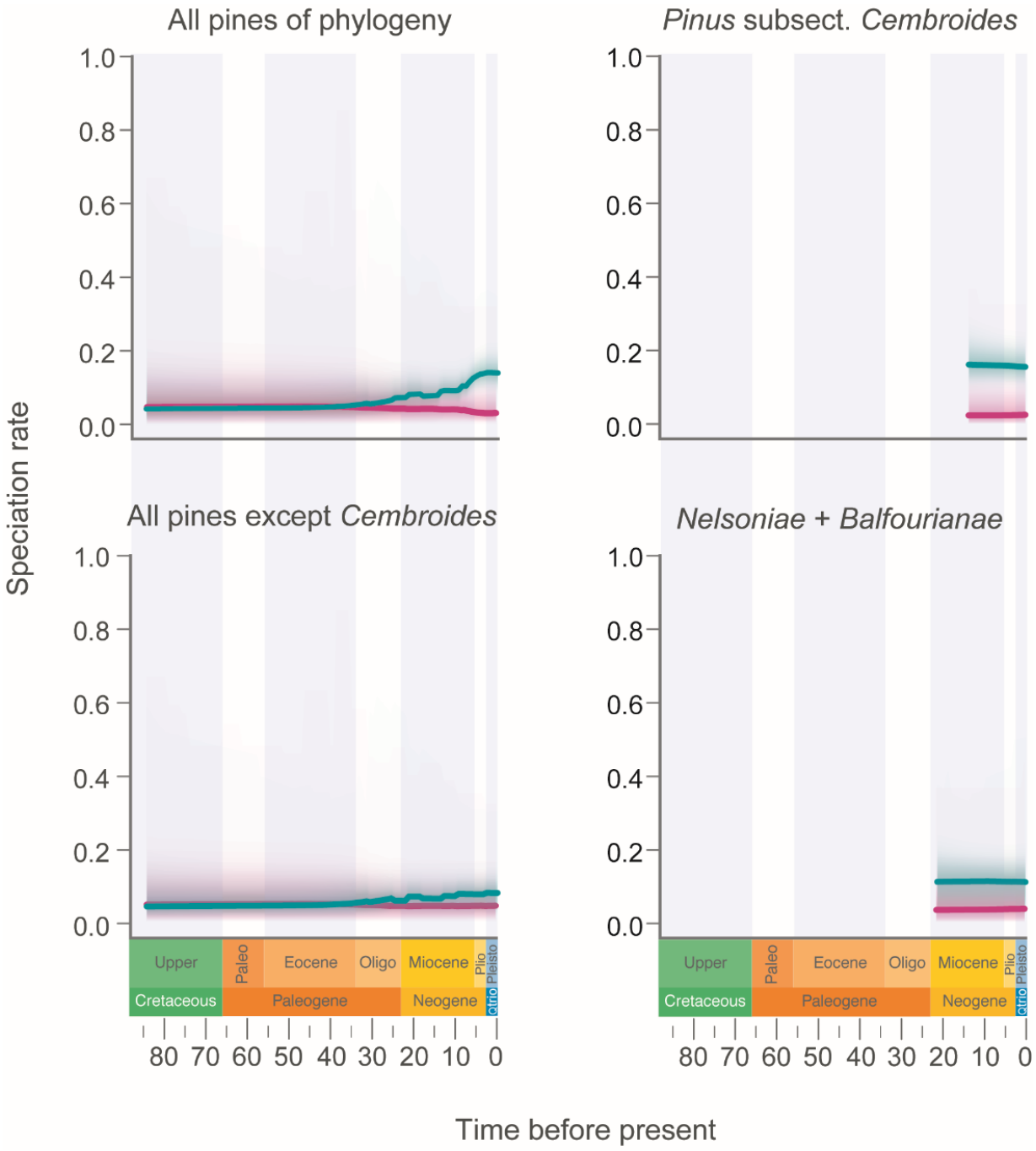

Subsection
P. lagunae
P. cembroides cembroides
P. cembroides orizabensis
P. discolor
P. johannis
P. culminicola
P. remota
P. lindgreni †
P. monophylla
P. quadrifolia
P. californiarum
P. edulis fallax
P. edulis
P. sanjuanensis †
P. maximartinezii
P. rzedowskii
P. pinceana
Cembroides
P. nelsonii
Nelsoniae
P. aristata
P. balfouriana
P. crossi †
P. longaeva
Balfourianae
0.9
0.8
0.7
0.8
0.9
0.7
0.9
0.6
0.6
0.9
0.8
0.8
1
0.9
0.8
0.7
0.6
0.6
0.005

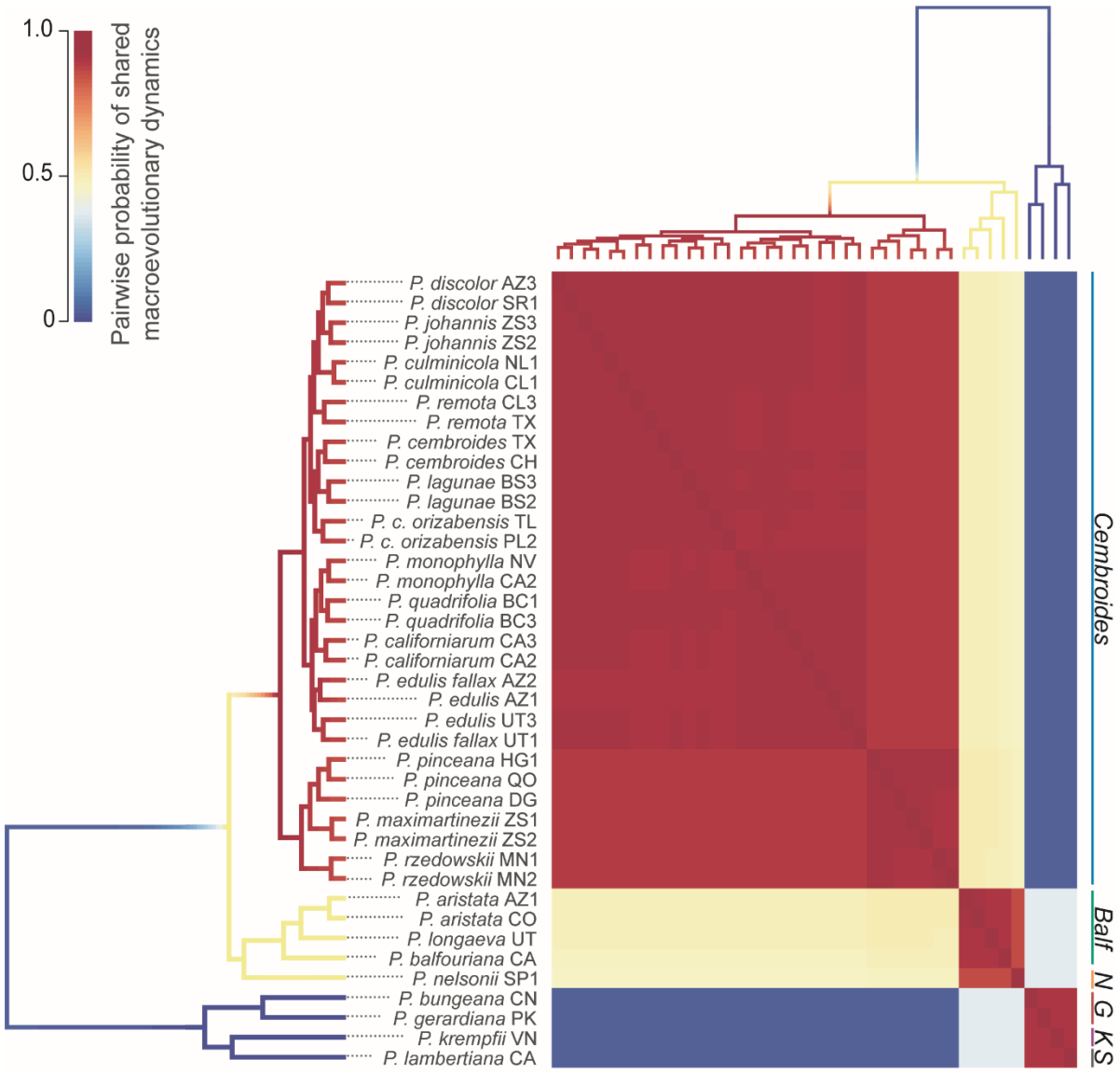

Subsection
Section
P. johannis ZS3
P. johannis ZS2
P. discolor AZ3
P. discolor SR1
P. culminicola CL1
P. culminicola NL1
P. remota CL3
P. remota TX
P. c. cembroides TX
P. c. cembroides CH
P. lagunae BS2
P. lagunae BS3
P. c. orizabensis PL2
P. c. orizabensis TL
P. californiarum CA2
P. californiarum CA3
P. edulis AZ1
P. edulis var. fallax AZ2
P. edulis var. fallax UT1
P. edulis UT3
P. quadrifolia BC1
P. quadrifolia BC3
P. monophylla CA2
P. monophylla NV
P. pinceana QO
P. pinceana HG1
P. pinceana DG
P. maximartinezii ZS1
P. maximartinezii ZS2
P. rzedowskii MN1
P. rzedowskii MN2
P. aristata CO
P. aristata AZ1
P. longaeva UT
P. balfouriana CA
P. nelsonii SP1
P. bungeana CN
P. gerardiana PK
P. krempfii VN
P. lambertiana CA
Cembroides
Balfourianae
Nelsoniae
Gerardianae
Krempfianae
Strobus
Parrya
Quinquefoliae
0.004


P. johannis ZS3
P. johannis ZS2
P. discolor AZ3
P. discolor SR1
P. culminicola NL1
P. culminicola CL1
P. remota CL3
P. remota TX
P. lagunae BS3
P. lagunae BS2
P. c. cembroides TX
P. c. cembroides CH
P. c. orizabensis TL
P. c. orizabensis PL2
P. quadrifolia BC1
P. quadrifolia BC3
P. monophylla NV
P. monophylla CA2
P. californiarum CA3
P. californiarum CA2
P. edulis fallax UT1
P. edulis UT3
P. edulis fallax AZ2
P. edulis AZ1
P. pinceana HG1
P. pinceana QO
P. pinceana DG
P. maximartinezii ZS2
P. maximartinezii ZS1
P. rzedowskii MN1
P. rzedowskii MN2
P. aristata AZ1
P. aristata CO
P. longaeva UT
P. balfouriana CA
P. nelsonii SP1
P. krempfii VN
P. lambertiana CA
P. gerardiana PK
P. bungeana CN
Cembroides
Balf
Ns
Kp
St
Gd
4.6
5.6
6.6
7.3
5.6
6.4
8.8
5.0
6.1
7.0
14.0
6.5
5.5
7.4
9.3
25.5
84.6
9.5
13.3
22.1
24.8
30.9
17.9
Upper
Paleocene
Eocene
Oligocene
Miocene
Pliocene
Pleistocene
Cretaceous
Paleogene
Neogene
Quat
80
70
60
50
40
30
20
10
0


Speciation rate
0.17
0.13
0.08
0.04
Density
13
9
4
-1
0.042
0.085
0.042
0.085
Speciation rate
f = 0.44
P. discolor AZ3
P. discolor SR1
P. johannis ZS3
P. johannis ZS2
P. culminicola NL1
P. culminicola CL1
P. remota CL3
P. remota TX
P. cembroides TX
P. cembroides CH
P. lagunae BS3
P. lagunae BS2
P. c. orizabensis TL
P. c. orizabensis PL2
P. monophylla NV
P. monophylla CA2
P. quadrifolia BC1
P. quadrifolia BC3
P. californiarum CA3
P. californiarum CA2
P. edulis fallax AZ2
P. edulis AZ1
P. edulis UT3
P. edulis fallax UT1
P. pinceana HG1
P. pinceana QO
P. pinceana DG
P. maximartinezii ZS1
P. maximartinezii ZS2
P. rzedowskii MN1
P. rzedowskii MN2
P. aristata AZ1
P. aristata CO
P. longaeva UT
P. balfouriana CA
P. nelsonii SP1
P. bungeana CN
P. gerardiana PK
P. krempfii VN
P. lambertiana CA
Cembroides
Balf
Ns
Gd
Kp
St
Upper
Paleocene
Eocene
Oligocene
Miocene
Pliocene
Pleistocene
Cretaceous
Paleogene
Neogene
Quat
80
70
60
50
40
30
20
10
0


Null model
f = 0.56
P. discolor AZ3
P. discolor SR1
P. johannis ZS3
P. johannis ZS2
P. culminicola NL1
P. culminicola CL1
P. remota CL3
P. remota TX
P. cembroides TX
P. cembroides CH
P. lagunae BS3
P. lagunae BS2
P. c. orizabensis TL
P. c. orizabensis PL2
P. monophylla NV
P. monophylla CA2
P. quadrifolia BC1
P. quadrifolia BC3
P. californiarum CA3
P. californiarum CA2
P. edulis fallax AZ2
P. edulis AZ1
P. edulis UT3
P. edulis fallax UT1
P. pinceana HG1
P. pinceana QO
P. pinceana DG
P. maximartinezii ZS1
P. maximartinezii ZS2
P. rzedowskii MN1
P. rzedowskii MN2
P. aristata AZ1
P. aristata CO
P. longaeva UT
P. balfouriana CA
P. nelsonii SP1
P. bungeana CN
P. gerardiana PK
P. krempfii VN
P. lambertiana CA
Cembroides
Balf
Ns
Gd
Kp
St
Single shift
f = 0.44
Upper
Paleocene
Eocene
Oligocene
Miocene
Pliocene
Pleistocene
Cretaceous
Paleogene
Neogene
Quat
80
70
60
50
40
30
20
10
0


All pines of phylogeny
Pinus subsect. *Cembroides*
All pines except *Cembroides*
Nelsoniae + *Balfourianae*
Speciation rate
1.0
0.8
0.6
0.4
0.2
0.0
Upper Cretaceous
Paleo
Eocene
Oligo
Paleogene
Miocene
Neogene
Plio
Pleisto
Otrio
80 70 60 50 40 30 20 10 0
Time before present


1.0
0.5
0
Pairwise probability of shared macroevolutionary dynamics
P. discolor AZ3
P. discolor SR1
P. johannis ZS3
P. johannis ZS2
P. culminicola NL1
P. culminicola CL1
P. remota CL3
P. remota TX
P. cembroides TX
P. cembroides CH
P. lagunae BS3
P. lagunae BS2
P. c. orizabensis TL
P. c. orizabensis PL2
P. monophylla NV
P. monophylla CA2
P. quadrifolia BC1
P. quadrifolia BC3
P. californiarum CA3
P. californiarum CA2
P. edulis fallax AZ2
P. edulis AZ1
P. edulis UT3
P. edulis fallax UT1
P. pinceana HG1
P. pinceana QO
P. pinceana DG
P. maximartinezii ZS1
P. maximartinezii ZS2
P. rzedowskii MN1
P. rzedowskii MN2
P. aristata AZ1
P. aristata CO
P. longaeva UT
P. balfouriana CA
P. nelsonii SP1
P. bungeana CN
P. gerardiana PK
P. krempfii VN
P. lambertiana CA
Cembroides
Balf
N
G
KS

APPENDIX 1. Morphological matrix of *Pinus* subsection *Cembroides.* The characters were treated as binary and multistate. The species list includes three fossils* and 19 extant species.

Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
P. monophylla	0	2	0	2	1	2	¿	0	1	0	1	0	0	0	2,3	0	1	1	2	2	1	2	0	1	0	1	1	1	1	0	2	0	0	1	0	1	0
P. californiarum	0	2	0	1	¿	¿	¿	0	1	0	1	0	¿	0	4	0	¿	¿	¿	2	2	2	0	1	0	1	1	¿	¿	0	1	0	0	1	¿	1	0
P. fallax	0	2	0	1	¿	¿	¿	0	1	0	1	0	¿	0	2,3	0	¿	¿	¿	2	2	2	0	1	0	1	1	¿	¿	0	1	0	0	1	¿	1	¿
P. edulis	0	2	0	2	¿	¿	¿	1	1	0	0	0	0	0	1	0	3	2	1	2	2	2	0	1	0	1	1	1	¿	0	1	0	1	1	0	1	0,1
P. culminicola	1	2	0	2	1	0,2	0	4	0	0	2	0	0	1	0	0	0	1	1	3	2	2	0	1	0	1	1	1	¿	0	0	1	1	1	0	0	0
P. remota	0	2	0	1	1	0	0	1	1	0	0,2	0	1	0	2	0	0,2	0,1	1	3	2	0	0	1	0	1	1	0	1	0	1	0	0	1	¿	1	0
P. quadrifolia	0	2	0	2	1	¿	¿	3,4	1	0	1	0	0	1	1	0	1	1	1	3	1	2	0	1	0	1	1	1	1	0	2	1	0	1	0	1	0
P. lagunae	1	2	0	2	1	0,2	0	1,2	1	1	0,2	0	0	0	1	0	3	0,1	1	3	2	1	0	1	1	1	1	0	1	0	1	1	1	0	0	0	0
P. orizabensis	1	2	0	2	1	0,2	0	1,2	1	0	0,2	0	0	0	1	0	3	0,1	1	3	2	1	0	1	1	1	1	0	1	0	1	1	1	0	0	0	0
P. cembroides	0	2	0	2	1	0,2	0	1,2	1	0	0,2	0	0	0	1	0	3	0,1	1	3	2	0	0	1	1	1	1	0	1	0	1	1	1	0	0	1	0
P. discolor	0	2	0	2	*1*	0,2	0	2	1	0	2	0	0	1	1	0	2	0,1	1	3	2	1	0	1	1	1	1	0	1	0	1	1	1	1	0	0	0
P. johannis	0	2	0	2	1	0,2	0	2	1	0	2	0	0	1	1	0	3	0,1	1	3	2	1	0	1	1	1	1	0	1	0	1	1	1	1	0	0	0,1
P. pinceana	0	2	0	0	1	1,2	0	2	1	1	2	0	0	0,1	1	0	1	0	1	4	1	0	1	1	0	1	1	1	1	0	1	2	1	1	¿	1	0
P. maximartinezii	0	2	0	2	1	2	0	4	1	1	3	0	0	1	1	0	1	2	1	1	1	¿	2	1	2	1	1	1	¿	0	3	1	2	1	1	0	0
P. rzedowskii	1	2	0	2	1	0	0	2,3	0	1	2	0	0	1	2	0	3	0	1	0	1	¿	1	0	0	1	1	1	0	2	0	2	¿	1	0	0	0
P. nelsonii	0	1	1	¿	1	0	1	2	0	1	2	0	0	0	1	0	0	2	1	5	0	0	1	1	0	1	0	1	¿	0	1	1	1	1	0	1	0
P. aristata	1	2	0	2	1	0	1	4	1	0	2	0	¿	1	0	0	1	2	1	0	1	2	1	1	1	1	0	1	0	2	0	1	¿	¿	¿	¿	0
P. balfouriana	1	2	0	2	1	0	1	4	1	0	2	0	¿	1	1	0	1	2	1	0	0	2	1	1	1	1	0	1	0	2	0	0	¿	¿	¿	¿	0
P. longaeva	1	2	0	2	1	0	1	4	1	0	2	0	¿	1	1	0	1	2	1	0	0	2	1	1	1	1	0	1	0	2	0	0	¿	¿	¿	¿	0
*P. lindgrenii**	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	2	1	?	0	1	0	1	?	?	?	0	2	0	?	?	?	?	?
*P. sanjuanensis**	?	?	0	?	?	?	?	1,2	?	0	0,2	?	?	?	?	?	?	?	?	2	?	?	0	?	1	1	0	?	?	?	?	?	?	?	?	?	?
*P. crossi**	?	?	0	?	?	?	?	4	?	0	?	?	?	?	?	?	?	?	?	4	0	?	1	1	1	1	0	?	?	2	1	2	?	?	?	?	?

1 Habit / 0=shrub 1=tree

2 Bark / 0=exfoliating 1=smooth 2=rough

3 Fascicle_retention / 0=deciduous 1=persistent

4 Fascicle_sheath / 0=stramineous 1=weak_rosette 2=rosette

5 Cataphylls / 0=absent 1=present

6 Cataphyll_shape / 0=subulate 1=ligulate 2=triangular

7 Cataphyll_length / 0=one_to_five_mm 1=six_to_ten_mm,

8 Needles_per_fascicle / 0=one 1=two 2=three 3=four 4=five,

9 Needle_margin / 0=serrulate 1=entire,

10 Needle_length / 0=lower_six_cm 1=upper_six_cm,

11 Shape_transverse_section / 0=semi_circular 1=circular 2=transversally_triangular 3=widely_triangular,

12 Vascular_bundle / 0=single 1=doble,

13 Wall_of_endodermal_cells / 0=not_thicked 1=slightly_thick,

14 Stomata_position / 0=amphistomatic 1=epistomatic,

15 Resin_ducts / 0=one 1=two 2=two_to_five 3=six_to_eight 4=nine_to_seventeen,

16 Position_of_resin_ducts / 0=external 1=internal 2=intermediate 3=septal,

17 Tridimensional_shape_vegetative_buds / 0=ovoid 1=ovoid_conical 2=ovoid_cylindric 3=ovoid_oblong,

18 Resin_on_vegetative_buds / 0=not_resinous 1=slightly_resinous 2=resinous,

19 Vegetative_buds_length / 0= one_to_five_mm 1=six_to_ten_mm 2=upper_ten_mm,

20 Seed_cones / 0=ovoid_conical 1=ovoid_truncate 2=ovoid_globose 3=subglobose 4=ovoid_oblong
5=cylindric,

21 Seed_cone_base / 0=conical 1=rounded 2=flat,

22 Resin_on_seed_cone / 0=rarely_resinous 1=intermediate 2=very_resinous,

23 Seed_cone_length / 0=lower_six_cm 1=six_to_fifteen_cm 2=upper_fifteen_cm,

24 Seed_scales_apophysis / 0=thin 1=thick,

25 Apophysis / 0=prominent 1=raised 2=very_prominent,

26 Umbo / 0=terminal 1=dorsal,

27 Mucro / 0=centromucronato 1=excentromucronato,

28 Pollen_cone_lenght / 0=lower_five_cm 1=upper_six_cm,

29 Microsporophylls / 0=peltate 1=subpeltate,

30 Seed_wing / 0=absent 1=rudimentary 2=long_effective,

31 Seed_length / 0=lower_ten_mm 1=ten_to_fifteen_mm 2=sixteen_to_twenty_mm 3=upper_twenty_mm,

32 Seed_shape / 0=ovoid_ellipsoid 1=ovoid_oblique 2=ovoid,

33 Integument / 0=0.1_to_0.5_mm 1=0.5_to_1.0_mm 2=1.1_to_2.0_mm,

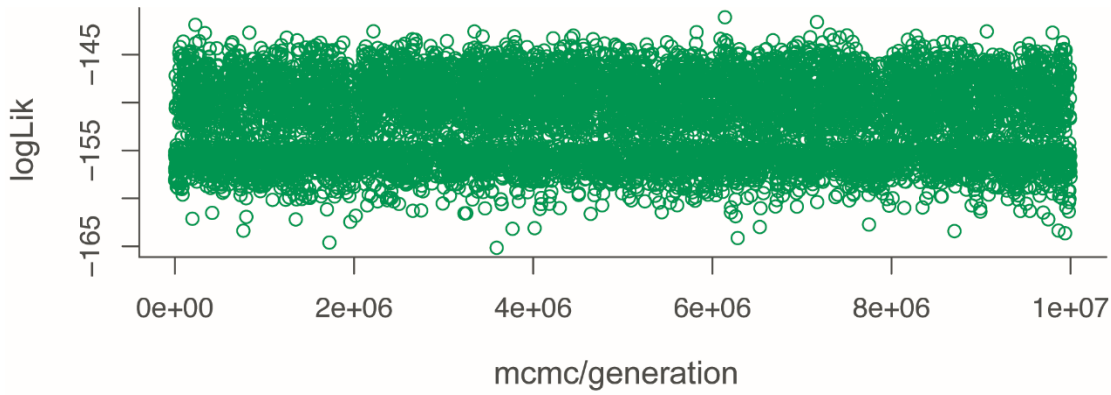
34 Color_of_gametophyte / 0=pink_in_fresh 1=white_in_fresh,

35 Cotyledons / 0=lower_fourteen 1=upper_fifteen,

36 Phenology / 0=spring 1=summer,

37 Sex_distribution / 0=monoecious 1=nearly_dioecious.

APPENDIX 2. Markovian chains results. The likelihood plot indicated the convergence of Markovian chains



Appendix 3. Shift posterior distributions with 95% of credibility. BAMM analyzed 9001 posterior samples: A) BAMM recorded 13 shift configurations but only 6 were reported based on posterior. B) Frequency of the 2 best shift configurations with highest posterior probability.

A

Shift number	Posterior
0	0.032
1	0.580
2	0.240
3	0.092
4	0.033
5	0.013
6	0.005
∑	0.995

B

Rank	Probability	Cumulative	Core_shifts
1	0.5638319	0.5638319	0
2	0.4361681	1.0000000	1

Appendix 4. Prior distribution using Bayes factor. Each column represents a shift rate configuration. Zero shift configuration represents the null model (no diversification rate). BAMM recovered the best model with the highest Bayes factor respecting to null model. The best model was in favor of the single rate shift.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	1	0.027660	0.033139	0.043432	0.060245	0.077481	0.108922	0.134208	0.165786	0.234863	0.128107	0.140918	0.352295	0.176147
1	36.15385	1	1.198098	1.570250	2.178106	2.801224	3.937953	4.852121	5.993796	8.491211	4.631570	5.094727	12.73682	6.368408
2	30.17602	0.834656	1	1.310619	1.817969	2.338058	3.286836	4.049851	5.002757	7.087240	3.865767	4.252344	10.63086	5.315430
3	23.02426	0.636841	0.762998	1	1.387108	1.783935	2.507850	3.090030	3.817096	5.407552	2.949574	3.244531	8.111328	4.055664
4	16.59875	0.459114	0.550064	0.720925	1	1.286082	1.807971	2.227679	2.751838	3.898438	2.126420	2.339063	5.847656	2.923828
5	12.90644	0.356987	0.427705	0.560559	0.777555	1	1.40580	1.73214	2.13971	3.03125	1.65341	1.81875	4.54688	2.273438
6	9.18087	0.253939	0.304244	0.398748	0.553106	0.711340	1	1.232143	1.522059	2.156250	1.176136	1.293750	3.234375	1.617188
7	7.45114	0.206095	0.246923	0.323621	0.448898	0.577320	0.811594	1	1.235294	1.75	0.954545	1.05	2.625	1.3125
8	6.03188	0.166839	0.199890	0.261979	0.363393	0.467354	0.657005	0.809524	1	1.416667	0.772727	0.85	2.125	1.0625
9	4.25780	0.117769	0.141099	0.184927	0.256513	0.329897	0.463768	0.571429	0.705882	1	0.545455	0.6	1.5	0.75
10	7.80596	0.215910	0.258681	0.339032	0.470274	0.604811	0.850242	1.047619	1.294118	1.833333	1	1.1	2.75	1.375
11	7.09633	0.196281	0.235164	0.308211	0.427522	0.549828	0.772947	0.952381	1.176471	1.666667	0.909091	1	2.5	1.25
12	2.83853	0.078513	0.094066	0.123284	0.171009	0.219931	0.309179	0.380952	0.470588	0.666667	0.363636	0.4	1	0.5
13	5.67706	0.157025	0.188132	0.246569	0.342017	0.439863	0.618357	0.761905	0.941176	1.333333	0.727273	0.8	2	1

4. DISCUSIÓN GENERAL

Pinus subsección *Cembroides* ha sido objeto de estudio en múltiples trabajos taxonómicos (Malusa, 1992; Farjon y Styles, 1997; Gernandt et al., 2005; Eckenwalder, 2009; Farjon y Filer, 2013), y la circunscripción y rango taxonómico de algunas especies sigue siendo controversial (Perry, 1991; Farjon y Styles, 1997; Gernandt et al., 2003; Farjon y Filer, 2013). La persistente incertidumbre taxonómica y la falta de límites claros entre algunas especies en los pinos piñoneros de Norte América representó una buena oportunidad para estudiar su delimitación. Las especies de *Pinus* subsección *Cembroides* también representan un modelo excelente para explorar causas de incongruencia filogenética como sorteo incompleto del linajes y evolución reticulada, debido a sus largos tiempos generacionales, grandes tamaños efectivos poblacionales (Syring et al., 2007), polinización por viento y barreras débiles al flujo de genes interespecífico (Mirov, 1967; Hong et al., 1993). El interés por estudiar delimitación de linajes y explorar causas que promueven conflicto entre árboles de genes y árboles de especies ha aumentado en los últimos años gracias a la creciente disponibilidad de datos moleculares, nuevos enfoques y el desarrollo de métodos coalescentes (Schlick-Steiner et al., 2010; Fujita et al., 2012; Zhang et al., 2013). Esta era de datos genómicos y métodos coalescentes hicieron posible alcanzar el objetivo central de este estudio que fue inferir las relaciones evolutivas, así como estudiar eventos de incongruencia filogenética y aclarar la delimitación de especies en *Pinus* subsección *Cembroides*, a partir de secuencias de genes nucleares de baja copia.

4.1. Relaciones evolutivas bajo métodos de coalescencia

Con base en los genes nucleares de baja copia y el empleo de métodos coalescentes como ASTRAL y SVDquartets que toman en cuenta ILS, se infirieron las relaciones evolutivas de los pinos piñoneros de Norte América. Este estudio cuenta con un muestreo taxonómico completo y la mayor cantidad de genes reportados para *Pinus* subsección *Cembroides*. Los análisis permitieron generar una hipótesis más clara y robusta de la filogenia de *Pinus* subsección *Cembroides.* También fue posible generar información

sobre la validez taxonómica de especies como *P*. *californiarum*, *P*. *johannis*, *P*. *discolor*, *P*. *quadrifolia* y *P*. *fallax.* Con los análisis fue posible soportar la hipótesis de que *P*. *rzedowskii* es el único piñonero que posee una semilla con ala funcional dentro de *Pinus* subsección *Cembroides*, lo cual no apoya la idea de ubicar a *P*. *rzedowskii* dentro de la subsección monotípica *Rzedowskianae* Carvajal como sugieren Price et al. (1998) y Farjon y Filer (2013). *Pinus rzedowskii* siempre se recuperó como la especie hermana de *P*. *maximartinezii* y *P*. *pinceana*. Juntos estos tres taxa forman un clado conocido como los piñoneros de conos grandes. Esta hipótesis ha sido apoyada previamente con secuencias de ADN de plasto (Gernandt et al., 2007). La separación de *P*. *nelsonii* de los demás pinos piñoneros de Norte América también fue apoyada con genes nucleares de baja copia. Esta separación ha sido soportada previamente con secuencias de ITS de ADNr (Gernandt et al., 2001). Ahora la evidencia apoya la hipótesis de que *Pinus* subsección *Nelsoniae* y *Pinus* subsección *Balfourianae* son el grupo hermano de *Pinus* subsección *Cembroides*.

Los análisis de coalescencia permitieron recuperar a *P. discolo*r y *P*. *johannis* como linajes independientes y no como subespecies de *P*. *cembroides* como lo reconocen Farjon y Filer (2013) por su similitud morfológica. Nuestros resultados coinciden con Gernandt et al. (2003) donde *P*. *cembroides* y sus variedades se recuperaron como un grupo monofilético. No obstante, fue necesario contar con mayor evidencia para confirmar el rango taxonómico de los tres taxa, debido a que *P. lagunae* se recuperó como la especie hermana de *P*. *cembroides* subsp. *cembroides,* mientras que *P*. *cembroides* subsp. *orizabensis* se colocó como el linaje hermano de los últimos dos taxa. Esto llamó la atención porque *P. lagunae* es una especie endémica de Baja California Sur y está separada de las poblaciones de *P*. *cembroides* (Farjon y Styles, 1997), mientras que las poblaciones más sureñas de *P*. *cembroides* subsp. *cembroides* en el estado de Hidalgo se encuentran a 290 km de distancia de las poblaciones de *P*. c*embroides* subsp. *orizabensis* en el estado de Puebla. Es decir, se esperaría que *P*. *cembroides* subsp. *cembroides* fuera el linaje hermano de *P*. c*embroides* subsp. *orizabensis* debido a que

sus poblaciones están más cercanas entre sí respecto a *P*. *lagunae*. *Pinus californiarum* fue una de las especies con más información generada a partir de los análisis de ASTRAL y SVDquartets. Validamos su estatus taxonómico como linaje independiente y no como subespecie, variedad o sinónimo de *P*. *monophylla* por la característica de poseer acículas solitarias (Silba, 1990; Price et al., 1998). Los análisis coalescentes recuperaron a *P*. *quadrifolia*, un taxón de 4 hojas por fascículo (predominantemente) como la especie hermana de *P*. *monophylla*. Nuestros resultados no soportan la hipótesis basada en morfología de agrupar filogenéticamente a *P*. *californiarum*, *P*. *fallax* y *P*. *monophylla* por compartir una sola acícula por fascículo (Malusa, 1992). Los resultados también difieren de las hipótesis planteadas por Gernandt et al. (2003; 2005), donde recuperaron a *P*. *californiarum* como la especie hermana de *P*. *edulis*, un taxón con dos acículas por fascículo predominantemente. Hasta ahora la evidencia indica que los pinos piñoneros de una sola acícula no forman un grupo monofilético pero junto con *P*. *edulis* y *P*. *quadrifolia* forman un clado bien soportado de los pinos piñoneros del suroeste de Estados Unidos y Baja California. Los resultados indican que las acículas solitarias en *Pinus* tienen un solo origen pero con dos reversiones, una en *P*. *edulis* y otra en *P*. *quadrifolia*. Finalmente, *P*. *culminicola* y *P*. *remota* son dos especies que consistentemente se recuperaron como linajes exclusivos en todos los análisis. *Pinus remota* siempre agrupó como la especie hermana del clado *P*. *culminicola*, *P. discolo*r y *P*. *johannis*, mientras que *P*. *culminicola* se ubicó como la especie hermana de *P. discolo*r y *P*. *johannis*. Esto coincide con Gernandt et al. (2003, 2005), quienes reportaron que *P*. *culminicola* está más cercanamente emparentado con *P. discolo*r y *P*. *johannis* que con *P*. *cembroides* y que *P*. *remota* es un linaje completamente distinto del resto de los pinos piñoneros de Norte América. Hasta esta etapa del estudio se reconocerían 13 especies en *Pinus* subsección *Cembroides*. Estos resultados forman parte del primer artículo de la tesis, logrando cumplir el primer objetivo específico de nuestra investigación.

4.2. Incongruencia filogenética en *Pinus* subsección *Cembroides*

La discordancia nuclear y citoplasmática es común en eucariotas y puede ser causada por procesos como el sorteo incompleto de linajes (Degnan y Rosenberg, 2009), hibridización (Rieseberg y Soltis, 1991) y transferencia lateral de genomas citoplasmáticos (Stegemann et al. 2012). La discordancia citonuclear en la era de los datos genómicos comenzó a retomar interés en plantas herbáceas (*Heuchera*: Folk et at., 2017; *Lachemilla*: Morales-Briones et al., 2018) y más recientemente en árboles (*Pinus*: Gernandt et al., 2018; *Juniperus*: Uckele et al., 2021; *Quercus*: Zhou et al., 2021). Particularmente, un estudio con datos Hyb-Seq en el género *Pinus* documentó la discordancia entre árboles de genes de ADN nuclear, ADN mitocondrial y ADN de plasto, la cual fue atribuida a eventos de introgresión más que a sorteo incompleto de linajes (Gernandt et al., 2018). El proceso de introgresión por captura de cloroplasto en pinos es moderadamente frecuente entre especies simpátricas o parapátricas (Delgado et al., 2007), mientras que, en especies alopátricas, la introgresión es atribuida a eventos de contacto secundario histórico (Liston et al., 2007).

Nuestros análisis de árboles de ADN nuclear y árboles de ADN de plasto presentaron incongruencia inesperadas, principalmente dentro del grupo de los pinos piñoneros de conos pequeños. La posición inesperada pero bien sustentada de *P. culminicola* dentro del clado de *P. discolor* y *P. johannis* sugiere que las tres especies comparten el mismo plastidio. Lo mismo ocurre con *P. californiarum* dentro del clado *P. edulis* y *P. fallax* y finalmente a *P. quadrifolia* dentro de *P. monophylla.* No obstante, los resultados de los cuatro eventos de reticulación con redes filogenéticas coalescentes sin asignación previa (sin probar hipótesis de reticulación basadas en las incongruencias entre árboles de ADN nuclear y ADN de plasto e hipótesis reportadas), solo coincidieron en reconocer a *P. quadrifolia* como un taxón introgresado con *P. culminicola*. Los tres eventos restantes detectaron únicamente escenarios de introgresión en especies inesperadas como *P. edulis*, *P. lagunae* y *P. cembroides* subsp. *cembroides*. Es decir, ningúno de los taxa introgresados detectados por Phylonet estuvieron involucrados entre

los desacuerdos de los árboles de ADN nuclear y árboles de ADN de plasto. De los escenarios de introgresión inesperados, nuestros resultados coincidieron con algunas hipótesis propuestas por Lanner (1974). Por ejemplo, se detectó que *P*. *edulis*, un taxón con predominantemente dos acículas por fascículo, fue introgresado con *P*. *monophylla* (predominantemente con acículas solitarias). Ambas especies se encuentran en simpatría en el Desierto de Mojave al Suroeste de California (Critchfield y Little, 1966), al Oeste de Utah, Nevada (Farjon y Filer, 2013) y en el Este de la "Gran Cuenca" donde árboles de *P*. *edulis* con una o dos acículas por fascículo han sido observados por Lanner (1974). Por tanto, no sería inusual que existiera flujo de genes entre *P*. *edulis* y *P*. *monophylla*, especies de piñoneros más ampliamente distribuidas en Estados Unidos (Critchfield y Little, 1966). Lo cierto es que este fue el único evento que involucró especies en simpatría, los tres eventos de introgresion restantes involucraron especies alopátricas, lo que indicaría que la frecuencia de introgresión por captura de cloroplasto en especies de pinos simpátricas es más baja que moderada. Entre los eventos de introgresión con especies alopátricas se encuentran *P*. *lagunae* introgresado con *P*. *edulis* o *P*. *fallax*, *P*. *cembroides* subsp. *cembroides* introgresado con *P*. *fallax* y *P*. *quadrifolia* introgresado con *P*. *culminicola*. Estos escenarios de introgresión solo podrían ocurrir por tres razones: por polinización a larga distancia (Mirov, 1967; Hong et al., 1993), por contacto secundario histórico y expansión de rangos de distribución (Liston et al., 2007; Gernandt et al., 2018) o por una inferencia incorrecta por parte del criterio de Coalescencia Profunda Mínima (MDC por sus siglas en inglés) implementado en el programa Phylonet. Sin embargo, la polinización por viento presenta un inconveniente entre especies alopátricas porque su fenología no siempre empata entre especies. De acuerdo con reportes de Farjon y Styles (1997) y con observaciones propias de herbario, ninguna de las especies alopátricas posiblemente introgresadas sobrelapa sus tiempos de producción de polen. *Pinus lagunae* produce polen de mayo a junio, mientras que *P*. *edulis* y *P*. *fallax* producen polen de abril a mayo. Asimismo, *P*. *cembroides* subsp. *cembroides* produce polen en un periodo que inicia en mayo y transcurre hasta finales de julio, mientras que *P*. *fallax* inicia

en abril y culmina en mayo. Finalmente, *P*. *quadrifolia* presenta reportes de producir polen a principios de abril y hasta finales de mayo, mientras que en *P*. *culminicola* el proceso ocurre de junio a julio. Por tanto, el aislamiento reproductivo fenológico en pinos juega un papel importante en los eventos de introgresión.

Ahora bien, existe una alta probabilidad de que las hipótesis de introgresión se deban a una falta de precisión por parte del criterio de parsimonia MDC implementado en Phylonet. De hecho, se ha reportado que las inferencias estimadas por el criterio de MDC no siempre son correctas (Alanzi y Degnan, 2021). Por último, sería importante estudiar y modelar la distribución potencial y la demografía histórica de *P*. *lagunae*, *P*. *cembroides* subsp. *cembroides*, *P*. *quadrifolia* y *P*. *culminicola* para probar si los eventos inferidos de introgresión podrían ser causados por contacto secundario histórico y una consiguiente expansión de rangos de distribución. No obstante, también existe la hipótesis de que la diversidad genética compartida entre especies alopátricas podría estar más influenciada por la retención de polimorfismo ancestral que por reticulación.

Cuando probamos las hipótesis de hibridización basadas en reportes previos, los resultados fueron no significativos. Phylonet detectó que *Pinus quadrifolia* no fue introgresado por *P*. *californiarum* como reportó Lanner (1974). En realidad, *P*. *californiarum* se recuperó como la especie hermana de *P*. *monophylla* y *P*. *quadrifolia* pero ningún evento de reticulación fue detectado entre *P*. *quadrifolia* y *P*. *californiarum*. Nuestros resultados tampoco probaron la hipótesis de Little (1968) quien reportó que *P*. *remota* era un híbrido.

Finalmente, los resultados de esta sección sobre incongruencia filogenética forman parte del primer artículo de la tesis, logrando cumplir el segundo objetivo específico del proyecto.

4.3. Delimitación de especies en *Pinus* subsección *Cembroides*

Con base en los análisis coalescentes a partir de genes nucleares y plastomas (GMYC, PTP y Tr2) se determinó que el número de especies reconocidas para los pinos piñoneros

de Norte América varía entre métodos y conjunto de datos. Primero, los resultados con plastomas mostraron inconsistencias entre GMYC y PTP, pero ambos fueron capaces de estimar a *Pinus maximartinezii*, *P*. *monophylla*, *P*. *quadrifolia*, *P*. *remota* y *P*. *rzedowskii* como especies válidas. Sin embargo, el método de GMYC reconoció el mayor número de especies (hasta 24), difiriendo demasiado del número de especies reconocidas por varios tratamientos morfológicos (Malusa, 1992; Perry, 1991; Price et al., 1998; Eckenwalder, 2009), mientras que PTP estimó solo 10 especies. Anteriormente en pinos, se pensó que un muestreo limitado a unos cuantos haplotipos por especie, un bajo número de secuencias y una baja variación en las secuencias de ADN de plasto podrían causar la disparidad entre el número de especies reconocidas y el número de especies estimadas por GMYC (Hernández et al., 2013). En este estudio, se incrementó el número de secuencias por especie, así como el número de sitios informativos al obtener los plastomas casi completos y los resultados indicaron que GMYC es un método que sobrestimó el número de especies en *Pinus* subsección *Cembroides.* Por el contrario, PTP subestimó un poco el número de especies respecto a los tratados taxonómicos pero los estudios han reportado que PTP es más preciso que el método GMYC (Tang et al., 2014).

Otro factor importante que afecta el poder discriminatorio de los métodos de delimitación de especies que utilizan un árbol filogenético como guía es la falta de monofilia de las especies (Carstens et al., 2013). Nuestro árbol guía estimado con plastomas solo recuperó como linajes exclusivos a *P*. *cembroides* subsp. *cembroides*, *P*. *maximartinezii*, *P*. *monophylla*, *P*. *quadrifolia*, *P*. *remota* y *P*. *rzedowskii*. Estos taxa son consistentemente reconocidos en tratamientos morfológicos como especies independientes (Price et al., 1998) y han sido recuperados como linajes distintos en estudios filogenéticos moleculares (Gernandt et al., 2003; Montes et al., 2019). Éstos mismos taxa fueron reconocidos como linajes independientes por GMCY y PTP, pero especies como *P*. *culminicola*, *P*. *discolor*, *P*. *johannis* y *P*. *pinceana* que han sido reconocidas en varios tratamientos morfológicos como linajes independientes (Perry,

1991; Price et al., 1998) no fueron recuperados como monofiléticos en el árbol de plastomas y por tanto tampoco fueron delimitadas por GMYC y PTP. Del mismo modo, los métodos de GMYC y PTP no lograron separar a *P*. *californiarum*, *P*. *edulis* y *P*. *fallax*. *Pinus californiarum* y *P*. *edulis* han sido reconocidos por su morfología, filogenia, distribución geográfica y regímenes de precipitaciones distintivos (Bailey, 1987; Cole et al., 2008; Montes et al., 2019), mientras que *P*. *fallax* ha sido tratado como una subespecie de *P*. *californiarum* o *P*. *edulis* (Little, 1968; Bailey, 1987). Sin embargo, los tres taxa agruparon juntos en el árbol de plastomas y por tanto fueron identificadas como una sola especie por PTP y GMYC. Este resultado sugiere que las poblaciones de *P*. *edulis*, *P*. *fallax* y *P*. *californiarum* podrían compartir el ADN de plasto debido a la introgresión, tal como lo reporta LaHood (1995). Lo que indica que aún utilizando el plastoma completo, incrementando el número de individuos por especie y utilizando métodos formales para delimitar especies bajo el modelo coalescente, los resultados se verán sesgados cuando tratamos de delimitar especies simpátricas o parapátricas que experimentan introgresión por captura de cloroplasto (Lou et al., 2018).

Segundo, los resultados de delimitación con el método de datos multilocus (Tr2) y análisis multivariados coincidieron con análisis filogenéticos recientes (Montes et al., 2019) y respaldaron algunas hipótesis taxonómicas. Por ejemplo, los resultados apoyaron la hipótesis de Little (1968) quien describió a *P*. *fallax* como una variedad taxonómica de *P*. *edulis*, que predominantemente tiene dos acículas por fascículo. También apoyaron la hipótesis de Bailey (1987) al reconocer a *P*. *californiarum* como una especie válida, que anteriormente había sido considerada una variedad taxonómica de *P*. *monophylla* (Silba, 1990). Similarmente, los análisis respaldaron el tratamiento de Passini (1987) en coincidir que *P*. *lagunae* es una especie distinta de *P*. *cembroides*, pero no respaldaron la propuesta de Bailey y Hawksworth (1992) de elevar a *P*. *cembroides* subsp. *orizabensis* a un estatus específico. Por último, el método de tripletes enraizados y los análisis multivariados también fueron congruentes en estimar a *P*. *discolor* y *P*. *johannis* como especies separadas, respaldando la hipótesis de Passini (1994) en sugerir que *P*. *discolor*

no es sinónimo de *P*. *johannis*. En definitiva, los análisis multilocus recuperaron 13 especies dentro de *Pinus* subsección *Cembroides* (*P*. *californiarum*, *P*. *cembroides*, *P*. *cembroides* subsp. *orizabensis*, *P*. *culminicola*, *P*. *discolor*, *P*. *edulis*, *P*. *edulis* var. *fallax*, *P*. *johannis*, *P*. *lagunae*, *P*. *maximartinezii*, *P*. *monophylla*, *P*. *pinceana*, *P*. *quadrifolia*, *P*. *remota*, y *P*. *rzedowskii*). Estos resultados forman parte del segundo artículo de la tesis, logrando cumplir el tercer objetivo específico de nuestra investigación.

4.4. Tiempos de divergencia de los pinos piñoneros de Norte América

En los últimos años, las estimaciones de los tiempos de divergencia en el género *Pinus* se han discutido cada vez más en estudios moleculares (Willyard et al., 2007; Gernandt et al., 2008; Saladin et al., 2017; Jin et al., 2021). Aunque se han utilizado para detallar la evolución del grupo, estos estudios a veces han inferido fechas sustancialmente diferentes para los principales eventos en la diversificación de *Pinus*. Particularmente, las fechas que han estimado los estudios de relojes moleculares también son dispares para el origen y diversificación de *Pinus* subsección *Cembroides* (ver Gernandt et al., 2008; Saladin et al., 2017; Jin et al., 2021), a pesar de que no ha sido el grupo focal de dichos estudios. Nuestras estimaciones del origen y tiempos de divergencia son concordantes con las estimaciones generadas a partir de un pequeño número de loci por Gernandt et al. (2008), quienes reportaron una edad troncal de aproximadamente 27 MA en el Oligoceno y una edad corona de ~11 MA en el Mioceno, pero son más recientes que las fechas estimadas por Jin et al. (2021) con datos genómicos (transcriptomas). Aunque existen diferencias en el número de genes y taxones, los diferentes resultados obtenidos reflejan un impacto de las posiciones filogenéticas de los fósiles que se utilizaron para calibrar los nodos.

En cuanto a las tasas de diversificación, nuestros resultados sugieren que la diversificación en *Cembroides* coincide tanto en condiciones climáticas (enfriamiento) como ecológicas (aridez) que ocurrieron durante el Mioceno (Herbert et al., 2016). Los resultados son congruentes con otros géneros de coníferas adaptados a ambientes

áridos de Norte América como *Juniperus* L. (Mao et al., 2010) y *Ephedra* L. (Loera et al., 2012) donde su diversificación acelerada ocurrió durante el Mioceno al Plioceno. En particular, nuestros resultados sugieren que el mayor número de linajes en *Cembroides* en comparación con las únicas dos subsecciones dentro de la sección *Parrya* (subsecciones *Nelsoniae* y *Balfourianae*) depende de las bajas tasas de extinción, lo que provocó una acumulación de linajes. Este comportamiento sobre las bajas tasas de diversificación también ha sido reportado recientemente para todo el género *Pinus* por Jin et al. (2021). Con estos resultados se cumplen el cuarto y último objetivo específico de la tesis. La información generada a partir de este último capítulo forma parte del tercer artículo de la tesis, que actualmente se encuentra en preparación y se titula "*Timing of diversification in North American pinyon pines using multigene molecular clocks".*

5. CONCLUSIONES GENERALES

Los genes nucleares de baja copia son una fuente rica de marcadores independientes de herencia biparental que proporcionan información filogenética importante. Estos genes parecen ser marcadores moleculares prometedores que han permitido explorar algunas causas de incongruencia filogenética en plantas como sorteo incompleto de linajes y evolución reticulada a través del uso de métodos coalescentes. El empleo de genes nucleares bajo un marco coalescente ha permitido generar hipótesis filogenéticas robustas a distintos rangos taxonómicos, estudiar delimitación de especies en grupos de plantas con historias de vida complejas y explorar preguntas relacionadas con eventos de especiación (como ortología).

El empleo del método Hyb-Seq en este estudio permitió incrementar el número de genes reportados y utilizados para estudiar tanto a *Pinus* subsección *Cembroides* como a las subsecciones *Nelsoniae*, *Balfourianae*, *Gerardianae* y *Krempfianae*. El muestreo de genes permitió extender la longitud de los alineamientos tanto de genes nucleares como de plastomas casi completos en miles de pares de bases, lo que resultó en la inclusión de mayor número de sitios filogenéticamente informativos y, por tanto, precisión de los árboles filogenéticos utilizados para los métodos de coalescencia. Las perspectivas de este estudio sugieren que Hyb-Seq promete ser un método útil para obtener un buen número de genes que permita explorar tanto aspectos filogenéticos como evolutivos en cualquier grupo de plantas.

Nuestros resultados documentaron la discordancia nuclear y citoplasmática en *Pinus* subsección *Cembroides*, principalmente dentro del grupo de los pinos piñoneros de conos pequeños. La discordancia citonuclear parece estar asociada al sorteo incompleto de linajes (ILS) más que a introgresión debido a los resultados arrojados por ASTRAL, que sugirieron un elevado nivel de ILS en el clado de los pinos piñoneros de conos pequeños. Los análisis de redes filogenéticas con Phylonet estimaron eventos de reticulación, pero ninguno fue significativo.

Pinus subsección *Cembroides* es el grupo con mayor número de especies dentro de la sección *Parrya* (subgénero *Strobus*) con 13 especies restringidas a zonas áridas y semiáridas de Norte América. En cuanto a los métodos, encontramos que PTP, GMYC y Tr2 mostraron inconsistencias en el número de especies estimadas. El poder discriminatorio de estos métodos de delimitación se ve sesgado por el resultado del árbol guía utilizado como archivo de entrada. Lo que sugiere que cuando tratamos de delimitar especies que experimentan introgresión por captura de cloroplasto, esto impactará en las relaciones evolutivas en un árbol filogenético y por tanto afectará la estimación del número de especies, particularmente en métodos como PTP y GMYC. Tr2 con genes nucleares parece ser un método con mayor poder discriminatorio, logrando obtener resultados que coincidieron bien con delimitaciones previas basadas en morfología, ADN, geografía y química secundaria.

Re-estimamos una escala de tiempo molecular para la divergencia entre los principales linajes de *Pinus* subsección *Cembroides* a partir de datos multilocus. Reportamos que estimaciones del origen y tiempos de divergencia de los pinos piñoneros de Norte América son concordantes con las estimaciones generadas a partir de un pequeño número de loci. Nuestros resultados apoyan la hipótesis que las especies dentro de *Pinus* subsección *Cembroides* diversificaron recientemente. La escala de tiempo bayesiana sugiere que *Pinus* subsección *Cembroides* se originó en el Oligoceno y que diversificó en el Mioceno. No obstante, nuestros estimados son una hipótesis basada en un registro fósil preservado incompletamente que dificulta comprender bien las relaciones filogenéticas de los fósiles con los linajes de *Pinus* subsección *Cembroides* y su relación con sus grupos hermanos (*Balfourinae* y *Nelsoniae*). Los fósiles que han sido asignados a especies modernas de pinos piñoneros a partir del “emparejamiento" de órganos de pinos existentes podrían proporcionar información no relevante de los caracteres "emparejados". El estudio de caracteres morfológicos y anatómicos en especies tanto

fósiles como vivientes es sumamente relevante en pinos y en cualquier grupo de organismos porque ayuda a identificar el momento de aparición de un estado de caracter ancestral y/o derivado en el registro fósil. Es necesario contar con estudios morfo-anatómicos más completos de los taxones existentes y fósiles con el mayor número de caracteres posibles para mejorar las estimaciones filogenéticas en pinos.

6. REFERENCIAS BIBLIOGRÁFICAS GENERALES

Aguirre-Dugua, X., and D. S. Gernandt. 2017. Complete plastomes of three endemic Mexican pines species (*Pinus* subsection *Australes*). *Mitochondrial DNA* 2: 562–565.

Alanzi, A. A., and J. H. Degnan. 2021. Statistical inconsistency of the unrooted minimize deep coalescence criterion. *PLoS ONE* 16: e0251107.

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

Álvarez, I., and J. F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434.

Asaf, S., A. L. Khan, M. A. Khan, R. Shahzad, Lubna, S. M. Kang, A. Al-Harrasi, A. Al-Rawahi, and I.-J. Lee. 2018. Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* L.) with related species. *PLoS ONE* 13: e0192966.

Axelrod, D. I. 1986. Cenozoic history of some western American pines. *Annals of the Missouri Botanical Garden* 73: 565–641.

Baldwin, B. G., M. J. Sanderson, J. M. Porter, M. F. Wojciechowski, C. S. Campbell, and M. J. Donoghue. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247–77.

Bailey, D. K. 1987. A study of *Pinus* subsection *Cembroides*. I: The single-needle pinyons of the Californias and the Great Basin. *Notes from the Royal Botanic Garden Edinburgh* 44: 275–310.

Bailey, D. K., and F. G. Hawksworth. 1992. Change in status of *Pinus cembroides* subsp. *orizabensis* (Pinaceae) from Central Mexico. *NOVON* 2: 306–307.

Baum, D. A., and S. D. Smith. 2012. *Tree Thinking: An Introduction to Phylogenetic Biology*. Colorado: Roberts and Company Publishers.

Blischak, P. D., J. Chifman, A. D. Wolfe, and L. S. Kubatko, 2018. HyDe: a Python package for genome-scale hybridization detection. *Systematic Biology* 67: 821–829.

Bowers, J., B. Chapman, J. A. Rong, and H. Paterson. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422: 433–438.

Bräutigam, A., and U. Gowik, 2010. What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biology* 12: 831–841.

Brown S. D., R. A. Collins, S. Boyer, M.-C. Lefort, J. Malumbres-Olarte, C. J. Vink, and R. H. Cruickshank. 2012. Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources* 12: 562–565.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and J. A. Drummond. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10: e1003537

Businsky, R., T. Frantík, and P. Vít. 2014. Morphological evaluation of the *Pinus kesiya* complex (Pinaceae). *Plant Systematics and Evolution* 300: 273–285.

Bussell, J. D., M. Waycott, and J. A. Chappill. 2005. Arbitrarily amplified DNA markers as characters for phylogenetic inference. *Perspectives in Plant Ecology, Evolution and Systematics* 7: 3–26.

Cai, L., Z. Xi, E. M. Lemmon, A. R. Lemmon, A. Mast, C. E. Buddenhagen, L. Liu, and C. C. Davis. 2021. The perfect storm: gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, Malpighiales. *Systematic Biology* 70: 491–507.

Carstens, B. C., T. A. Pelletier, N. M. Reid, and J. D. Satler. 2013. How to fail at species delimitation. *Molecular Ecology* 22: 4369–4383.

Castro-Félix, P., J. A. Pérez de la Rosa, G. Vargas-Amado, S. Velásquez-Magaña, A. Santerre, F. López-Dellamary Toral, and A. R. Villalobos-Arámbula. 2008. Genetic relationships among Mexican white pines (*Pinus*, Pinaceae) based on RAPD markers. *Biochemical Systematics and Ecology* 36: 523–530.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, et al. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528–580.

Cole, K. L., J. F. Fisher, S. T. Arundel, J. Cannella, and S. Swift. 2008. Geographical and climatic limits of needle types of one-and two-needled pinyon pines. *Journal of Biogeography* 35: 257–269.

Critchfield, W. B., and E. L. Little. 1966. *Geographic Distribution of the Pines of the World*. Washington, D.C: U.S. Department of Agriculture Miscellaneous Publication 991.

Cronn, R., A. Liston, M. Parks, D. S. Gernandt, R. Shen, and T. Mockler. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research* 36: e122.

Cusimano, N., and S. S. Renner. 2010. Slowdowns in diversification rates from real phylogenies may not be real. *Systematic Biology* 59: 458–64.

Dasarathy, G., R. Nowak, and S. Roch. 2015. Data requirement for phylogenetic inference from multiple loci: a new distance method. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 12: 422–432.

Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24: 332–340.

De La Torre, A. R., Z. Li, Y. Van de Peer, and P. K. Ingvarsson. 2017. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Molecular Biology and Evolution* 34: 1363–1377.

Delgado, P., R. Salas-Lizana, A. Vázquez-Lobo, A. Wegier, M. Anzidei, E. R. Alvarez-Buylla, G. G. Vendramin, and D. Piñero. 2007. Introgressive hybridization in *Pinus montezumae* Lamb and *Pinus pseudostrobus* Lindl. (Pinaceae): morphological and molecular (cpSSR) evidence. *International Journal of Plant Sciences* 168: 861–875.

Després, L., L. Gielly, B. Redoutet, and P. Taberlet. 2003. Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. *Molecular Phylogenetics and Evolution* 27: 185–196.

Dong, S.-S., Y.-L. Wang, N.-H. Xia, Y. Liu, M. Liu, L. Lian, N. Li, L.-F. Li, et al. 2021. Plastid and nuclear phylogenomic incongruences and biogeographic implications of *Magnolia* s.l. (Magnoliaceae). *Journal of Systematics and Evolution* 00: 1–15.

Doyle, J. J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany* 17: 14–163.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Ecology and Evolution* 7:214.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: 0699–0710.

Duminil, J., and M. Di Michele. 2009. Plant species delimitation: a comparison of morphological and molecular markers. *Plant Biosystems* 143: 528–542.

Dvorak, W., A., Jordon, G. Hodge, and J. L. Romero. 2000. Assessing evolutionary relationships of pines in the *Oocarpae* and *Australes* subsections using RAPD markers. *New Forests* 20: 163–192.

Eckenwalder, J. E. 2009. *Conifers of the World: The Complete Reference*. Portland, Oregon: Timber Press.

Edwards S. V. 2009. Is a new and general theory of molecular systematics emerging? *Evolution* 63: 1–19.

Ence, D. D., and B. C. Carstens. 2011. SpedeSTEM: A rapid and accurate method for species delimitation. *Molecular Ecology Resources* 11: 473–480.

Erickson, D. L., J. Spouge, A. Resch, L. A. Weigt, and J. W. Kress. 2008. DNA barcoding in land plants: developing standards to quantify and maximize success. *Taxon* 57: 1304–1316.

Farjon, A., and B. Filer. 2013. *An Atlas of the World's Conifers: An Analysis of their Distribution, Biogeography, Diversity and Conservation Status*. Netherlands: Brill.

Farjon, A. and B. Styles. 1997. *Flora Neotropica Monograph*: Pinus (*Pinaceae*). New York: The New York Botanical Garden.

Flouri, T., X. Jiao, B. Rannala, and Z. Yang. 2020 A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Molecular Biology and Evolution* 37: 1211–1223.

Folk, R. A., J. R. Mandel, and J. V. Freudenstein. 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systematic Biology* 66: 320–337.

Fujisawa, T., A. Aswad, and T. G. Barraclough. 2016. A rapid and scalable method for multilocus species delimitation using Bayesian model comparison and rooted triplets. *Systematic Biology* 65: 759–771.

Fujita, M. K., A. D. Leaché, F. T. Burbrink, J. A. McGuire, and C. Moritz. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution* 27: 480–8.

Gernandt, D. S., X. Aguirre Dugua, A. Vázquez-Lobo, A. Willyard, A. Moreno Letelier, J. A. Pérez de la Rosa, D. Piñero, and A. Liston. 2018. Multi-locus phylogenetics,

lineage sorting, and reticulation in *Pinus* subsection *Australes*. *American Journal of Botany* 105: 711–725.

Gernandt, D. S., G. Gaeda López, S. Ortiz García, and A. Liston. 2005. Phylogeny and classification of *Pinus*. *Taxon* 54: 29–42.

Gernandt, D. S., G. Holman, C. Campbell, M. Parks, S. Mathews, L. A. Raubeson, A. Liston, R. A. Stockey, and G. W. Rothwell. 2016. Phylogenetics of extant and fossil Pinaceae: methods for increasing topological stability. *Botany* 94: 863–884.

Gernandt, D. S., A. Liston, and D. Piñero. 2001. Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: implications for molecular systematic studies of pine species complexes. *Molecular Phylogenetics and Evolution* 21: 449–467.

Gernandt, D. S., A. Liston, and D. Piñero. 2003. Phylogenetic of *Pinus* subsections *Cembroides* and *Nelsoniae* inferred from cpDNA sequences. *Systematic Botany* 28: 657–673.

Gernandt, D. S., S. A. Magallón, G. Geada López, O. Zerón Flores, A. Willyard, and A. Liston. 2008. Use of simultaneous analyses to guide fossil-based calibrations of Pinaceae phylogeny. *International Journal of Plant Sciences* 169: 1086–1099.

Gernandt, D. S., C. Reséndiz Arias, T. Terrazas, X. Aguirre-Dugua, and A. Willyard. 2018b. Incorporating fossils into the Pinaceae tree of life. *American Journal of Botany* 105: 1329–1344.

Gernandt, D. S., O. Zerón, and I. Goyenechea. 2007. Inferencia filogenética mediante secuencias de ADN: un ejemplo con los pinos piñoneros. *En La Sistemática Base del Conocimiento de la Biodiversidad*, UAEH.

Gitlin, A. R., C. M. Sthultz, M. A. Bowker, S. Stumpf, K. L. Paxton, K. Kennedy, A. Muñoz, and T. G. Whitham. 2006. Mortality gradients within and among dominant plant populations as barometers of ecosystem change during extreme drought. *Conservation Biology* 20: 1477–1486.

Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, et al. 2009. Solution hybrid selection with ultra-long

oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27: 182-189.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kirchernick, N. Patterson, H. Li, et al. 2010 A draft sequence of the Neandertal genome. *Science* 328: 710-722.

Hanover, J. W. 1992. Applications of terpene analysis in forest genetics. *New Forests*. 6: 159–178.

Heath, T. A., S. M. Hedtke, and D. M. Hillis. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46: 239–257.

Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences of United States of America* 111: E2957-E2966.

Herbert, T., K. Lawrence, A. Tzanova, L. C. Peterson, R. Caballero-Gill, and C. S. Kelly. 2016. Late Miocene global cooling and the rise of modern ecosystems. *Nature Geoscience* 9: 843–847.

Hernández-León, S., D. S. Gernandt, J. A. Pérez de la Rosa, and L. Jardón-Barbolla. 2013. Phylogenetic relationships and species delimitation in *Pinus* section *Trifoliae* inferred from plastid DNA. *PLoS ONE* 8: e70501.

Ho, S. Y., and S. Duchêne. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology* 23: 5947–5965.

Hollingsworth, P. M. 2011. Choosing and using a plant DNA barcode. *PLoS ONE* 6: e19254.

Hong, Y.-P., A. B. Krupkin, and S. H. Strauss. 1993. Chloroplast DNA transgresses species boundaries and evolves at variable rates in the California closed-cone pines (*Pinus radiata*, *P. muricata*, and *P. attenuata*). *Molecular Phylogenetics and Evolution* 2: 322–329.

Hudson, R. R. 1983 Testing the constant-rate neutral allele model with protein sequence data. *Evolution* 37: 203–217.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* 7: 44.

Huelsenbeck, J. P., P. Andolfatto, and E. T. Huelsenbeck. 2011. Structurama: Bayesian inference of population structure. *Evolutionary Bioinformatics* 7: 55–59.

Huelsenbeck, J. P., B. Rannala, and J. P. Masly. 2000. Accommodating phylogenetics uncertainty in evolutionary studies. *Science* 288: 2349–50.

IUCN, 2021. The IUCN Red List of Threatened Species. Version 2021-1.

Jin, W. T., D. S. Gernandt, C. Wehenkel, X. M. Xia, X. X. Wei, and X. Q. Wang. 2021. Phylogenomic and ecological analyses reveal the spatiotemporal evolution of global pines. *Proceedings of the National Academy of Sciences of the United States of America* 118.

Jones, G., Z. Aydin, and B. Oxelman. 2015. DISSECT: An assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics* 31: 991–998.

Kapli, P., S. Lutteropp, J. Zhang, K. Kobert, P. Pavlidis, A. Stamatakis, and T. Flouri. 2017. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* 33: 1630–1638.

Kingman, J. F. C. 1982. The coalescent. *Stochastic Processes and their Applications* 13: 235–248.

Knowles, L. L., H. C. Lanier, P. B. Klimov, and Q. He. 2012. Full modeling versus summarizing gene-tree uncertainty: method choice and species-tree accuracy. *Molecular Phylogenetics and Evolution* 65: 501–509.

Kubatko, L. S., and J. Chifman. 2019. An invariants-based method for efficient identification of hybrid species from large-scale genomic data. *BMC Evolutionary Biology* 19: 112.

Kubatko, L. S., B. C. Carstens, and L. L. Knowles. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25: 971–973.

Lahaye, R., V. Savolainen, S. Duthoit, O. Maurin, and M. van der Bank. 2008. A test of *psbK-psbI* and *atpF-atpH* as potential plant DNA barcodes using the flora of the Kruger National Park (South Africa) as a model system. *Nature Precedings*. https://doi.org/10.1038/npre.2008.1896.1.

LaHood. E. 1995. A chloroplast DNA phylogeny of nine taxa in *Pinus* subsection *Cembroides*. MSc. dissertation, Northern Arizona University, Flagstaff, Arizona, USA.

Lanner, R. M. 1974. A new pine from Baja California and the hybrid origin of *Pinus quadrifolia*. *The Southwestern Naturalist* 19: 75–95.

Liston, A., M. Parker-Defeniks, J. V. Syring, A. Willyard, and R. Cronn. 2007. Interspecific phylogenetic analysis enhances intraspecific phylogeographical inference: a case study in *Pinus lambertiana*. *Molecular Ecology* 16: 3926–3937.

Lanner, R. M. 1981. *The Piñon Pine: A Natural and Cultural History*. Nevada: University of Nevada Press.

Larget, B. R., S. K. Kotha, C. N. Dewey, and C. Ané. 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* 26: 2910-2911.

Leaché, A. D., and B. Rannala. 2011. The accuracy of species tree estimation under simulation: a comparison of methods. *Systematic Biology* 60: 126–137.

Ledig, F. T., M. A. Capó-Arteaga, Paul D. Hodgskiss, H. Sbay, C. Flores-López, M. T. Conkle, and B. Bermejo-Velázquez. 2001. Genetic diversity and the mating system of a rare Mexican piñon, *P. pinceana*, and a comparison with *P. maximartinezii* (Pinaceae). *American Journal of Botany* 88:1977–1987.

Lee, M. S. Y., and S. Y. W. Ho. 2016. Molecular clocks. *Current Biology* 26: R399–R402.

Li, D.-Z., L.-M. Gao, H.-T. Li, H. Wang, X.-J. Ge, J.-Q. Liu, Z.-D. Chen, S.-L. Zhou, et al. 2011. Comparative analysis of a large dataset indicates that Internal Transcribed Spacer (ITS) should be incorporated into the core barcode for seed plants.

Proceedings of the National Academy of Sciences of the United States of America 108: 19641–19646.

Li, X., Y. Yang, R. J. Henry, M. Rossetto, Y. Wang, and S. Chen. 2015. Plant DNA barcoding: from gene to genome. *Biological Reviews* 90: 157–166.

Lidholm, J., and P. Gustafsson. 1991. The chloroplast genome of the gymnosperm *Pinus contorta*: a physical map and a complete collection of overlapping clones. *Current Genetics* 20: 161–166.

Ligon, J. D. 1978. Reproductive interdependence of pinon jays and pinon pines. *Ecological Monographs* 48: 111–126.

Little, E. L. 1968. Two new pinyon varieties from Arizona. *Phytologia* 17:329–342.

Liu, L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* 2421: 2542–2543.

Liu, L., and S. V. Edwards. 2009. Phylogenetic analysis in the anomaly zone. *Systematic Biology* 58: 452–460.

Liu, L., and D. K. Pearl. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic Biology* 56: 504–514.

Liu, L, and L. Yu. 2011. Estimating species trees from unrooted gene trees. *Systematic Biology* 60: 661–667.

Liu L., L. Yu, and S. V. Edwards. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology* 10: 302.

Liu, L., D. K. Pearl, R. T. Brumfield, and S. V. Edwards. 2008. Estimating species trees using multiple-allele DNA sequence data. *Evolution* 62: 2080–2091.

Liu L., L. Yu, D. K. Pearl, and S. V. Edwards. 2009. Estimating species phylogenies using coalescence times among sequences. *Systematic Biology* 58: 468–477.

Loera, I., V. Sosa, and S. M. Ickert-Bond. 2012. Diversification in North American arid lands: Niche conservatism, divergence and expansion of habitat explain speciation in the genus *Ephedra*. *Molecular Phylogenetics and Evolution* 65: 437–450.

López-Reyes, A., J. A. Pérez de la Rosa, E. Ortiz, and D. S. Gernandt. 2015. Morphological, molecular, and ecological divergence in *Pinus douglasiana* and *P. maximinoi*. *Systematic Botany* 40: 658-670.

Luo, A., C. Ling, S. Y. W. Ho, and C-D. Zhu. 2018. Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Systematic Biology* 67: 830–846.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.

Madrigal, S. X. and D. M. Caballero. 1969. Una nueva especie mexicana de *Pinus*. Boletín Técnico del Instituto Nacional de Investigaciones Forestales 26: 1–11.

Magallón, S. A., and M. J. Sanderson. 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. *Evolution* 59: 1653–70.

Malusa, J. 1992. Phylogeny and biogeography of the pinyon pines (*Pinus* subsect. *Cembroides*). *Systematic Botany* 17: 42–66.

Mao, K., G. Hao, J. Liu, R. P. Adams, and R. I. Milne. 2010. Diversification and biogeography of *Juniperus* (Cupressaceae): Variable diversification rates and multiple intercontinental dispersals. *New Phytologist* 188: 254–272.

Meng, C., and L. S. Kubatko. 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. *Theoretical Population Biology* 75: 35–45.

Millar, C. I. 1993. Impact of the Eocene on the evolution of *Pinus* L. *Annals of the Missouri Botanical Garden* 80: 471–498.

Millar, C. I., S. H. Strauss, M. T. Conkle, and R. D. Westfall. 1988. Allozyme differentiation and biosystematics of the Californian closed-cone pines (*Pinus* subsect. *Oocarpae*). *Systematic Botany* 13: 351–370.

Miller, C. N. 1992. Structurally preserved cones of *Pinus* from the Neogene of Idaho and Oregon. *International Journal of Plant Sciences*. 153: 147–154.

Mirarab, S., and T. Warnow. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 12: i44–i52.

Mirarab, S., L. Nakhleh, and T. Warnow. 2021. Multispecies coalescent: theory and applications in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 52: 1.

Mirarab, S., R. Reaz, M. Bayzid, T. Zimmermann, M. Swenson, and T. Warnow. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30: i541–i548.

Mirarab, S., Md. Shamsuzzoha Bayzid, and T. Warnow. 2016. Multilocus species tree estimation in the presence of incomplete lineage sorting. *Systematic Biology* 65: 366–380.

Mirov, N. T. 1967. *The Genus Pinus*. New York: Ronald Press.

Mitic, Z. S., B. M. Nikolic, M. S. Ristic, V. V. Tesevic, S. R. Bojovic, and P. D. Marin. 2017. Terpenes as useful markers in differentiation of natural populations of relict pines *Pinus heldreichii*, *P. nigra*, and *P. peuce*. *Chemistry and Biodiversity* 14: e1700093.

Montes, J. R., P. Peláez, A. Willyard, A. Moreno-Letelier, D. Piñero, and D. S. Gernandt. 2019. Phylogenetics of *Pinus* subsection *Cembroides* Engelm. (Pinaceae) inferred from low-copy nuclear gene sequences. *Systematic Botany* 44: 501–518.

Morales-Briones, D. F., A. Liston, and D. C. Tank. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytologist* 218: 1668–1684.

Morales-Saldaña, S., S. Valencia-Ávalos, K. Oyama. E. Tovar-Sánchez, A. L. Hipp, and A. González-Rodríguez. 2021. Even more oak species in Mexico? Genetic structure and morphological differentiation support the presence of at least two specific entities within *Quercus laeta*. *Journal of Systematics and Evolution* 0: 1-16.

Moreno-Letelier, A., A. Ortiz-Medrano, and D. Piñero. 2013. Niche divergence versus neutral processes: combined environmental and genetic analyses identify contrasting patterns of differentiation in recently diverged pine species. *PLoS ONE* 8: e78228.

Mort, M. E., J. K. Archibald, C. P. Randle, N. D. Levsen, T. R. O'leary, K. Topalov, C. M. Wiegand, and D. J. Crawford. 2007. Inferring phylogeny at low taxonomic levels: utility of rapidly evolving cpDNA and nuclear ITS loci. *American Journal of Botany* 94: 173–183.

Mossel, E., and S. Roch. 2010. Incomplete lineage sorting: consistent phylogeny estimation from multiple loci. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 7: 166–171.

O'Hara, R. J. 1992. Telling the tree: narrative representation and the study of evolutionary history. *Biology and Philosophy* 7: 135–160.

O'Meara, B. C. 2010. New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology* 59: 59–73.

O' Meara, B., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60: 922–933.

Ortiz-Medrano, A., D. P. Scantlebury, A. Vázquez-Lobo, A. Mastretta-Yanes, and D. Piñero. 2016. Morphological and niche divergence of pinyon pines. *Ecology and Evolution* 6: 2886–2896.

Palmé, A. E., T. Pyhäjärvi, W. Wachowiak, and O. Savolainen. 2009. Selection on nuclear genes in a *Pinus* phylogeny. *Molecular Biology and Evolution* 26: 893–905.

Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* 5: 568–583.

Pang, X., H., Luo, and C. Sun. 2012. Assessing the potential of candidate DNA barcodes for identifying non-flowering seed plants. *Plant Biology (Stuttgart)* 14: 839–44.

Parks, M., R. Cronn, and A. Liston. 2012. Separating the wheat from the chaff: mitigating the effects of noise in plastome phylogenomic data set from *Pinus* L. (Pinaceae). *BMC Evolutionary Biology* 12: 100.

Passini, M.-F. 1987. The endemic pinyon of Lower California *Pinus lagunae* M.-F. Passini. *Phytologia* 63: 337–338.

Passini, M.-F. 1994. Synonymie entre *Pinus discolor* Bailey & Hawksworth et *Pinus johannis* M.-F. Robert. *Acta Botanica Gallica* 141: 387–388.

Perry, J. P. 1991. *The Pines of Mexico and Central America*. Portland: Timber Press.

Peterson, A. T. 2006. Application of molecular clocks in ornithology revisited. *Journal of Avian Biology* 37: 541–544.

Petit, R. J., and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution and Systematics* 37: 187–214.

Plessas, M. E., and S. H. Strauss. 1986. Allozyme differentiation among populations, stands, and cohorts in Monterey Pine. *Canadian Journal of Forest Research* 16: 1155–1164.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, et al. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* 55: 595–609.

Poulos, H. M., and G. P. Berlyn. 2007. Variability in needle morphology and water status of *Pinus cembroides* across an elevational gradient in the Davis Mountains of west Texas, USA. *Journal of the Torrey Botanical Society* 134: 281–288.

Price, R. A., A. Liston, and S. H. Strauss. 1998. Phylogeny and systematics of *Pinus*. *In Ecology and biogeography of Pinus*. D. M. Richardson [Ed.], pp 49–68. Cambridge: Cambridge University Press.

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

Puillandre, N., A. Lambert, S. Brouillet, and G. Achaz. 2011. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* 21: 1864–1877.

Rannala, B., S. V. Edwards, A. Leaché, and Z. Yang. 2020. The multi-species coalescent model and species tree inference. *In Phylogenetics in the Genomic Era*. C. Scornavacca, F. Delsuc and N. Galtier [Eds.], No. 3.3. pp 1–21. https://hal.inria.fr/PGE.

Rieseberg, L. H., and D. E. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5: 65–84.

Roberts, D. R. 1970. Within-tree variation of monoterpene hydrocarbon composition of slash pine oleoresin. *Phytochemistry* 9: 809–815.

Roch, S., and M. Steel. 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. *Theoretical Population Biology* 100: 56–62.

Rokas, A., and S. B. Carroll. 2005. More genes or more taxa: the relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution* 22: 1337–44.

Rokas, A., B. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798–804.

Rose, J. P., C. A. P. Toledo, E. Moriarty Lemmon, A. R Lemmon, and K. J. Sytsma. 2021. Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant genus *Polemonium* (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Systematic Biology* 70: 162–180.

Rosenberg, N. A. 2003. The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly and polyphyly in a coalescent model. *Evolution* 57: 1465–1477.

Saladin, B., A. B. Leslie, R. O. Wüest, G. Litsios, E. Conti, N. Salamin, and N. E. Zimmermann. 2017. Fossils matter: Improved estimates of divergence time in *Pinus* reveal older diversification. *BMC Evolutionary Biology* 17: 95.

Sang, T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37: 121–147.

Sarac, Z., S. Bojovic, B. Nikolic, V. Tesevic, I. Dordevic, and P. D. Marin. 2013. Chemotaxonomic significance of the terpene composition in natural populations of *Pinus nigra* J. F. Arnold from Serbia. *Chemistry & Biodiversity* 10: 1507–520.

Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian, and R. H. Crozier. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology* 55: 421-438.

Scotland, R. W., R. G. Olmstead, and J. R. Bennett. 2003. Phylogeny reconstruction: the role of morphology. *Systematic Biology* 52: 539–548.

SEMARNAT. 2020. Norma Oficial Mexicana NOM-059-ECOL-2010. Protección ambiental. Especies nativas de México de flora y fauna silvestres. Categorías de riesgo y especificaciones para su inclusión, exclusión o cambio. Lista de especies en riesgo. Diario Oficial de la Federación, 30 de diciembre de 2010.

Shao, Y.-Z., and Q.-P. Xiang. 2015. Species delimitation and phylogeography of the *Abies chensiensis* complex inferred from morphological and molecular data. *Botanical Journal of the Linnean Society* 177: 175–188.

Show, G. R. 1909. *The Pines of Mexico*. Publications of the Arnold Arboretum No. 1. Boston: J. R. Ruiter & Co.

Silba, J. 1990. A supplement to the international census of the coniferae, II. *Phytologia* 68: 7–78.

Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals conflict, concordance and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15: 150.

Smith S. A., N. Walker-Hale, J. F. Walker, and J. W. Brown. 2020. Phylogenetic conflicts, combinability, and deep phylogenomics in plants. *Systematic Biology* 69: 579–592.

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

Stegemann, S., M. Keuthe, S. Greiner, and R. Bock. 2012. Horizontal transfer of chloroplast genomes between plant species. *Proceedings of the National Academy of United States of America 109*: 2434–2438.

Stubbs R. L., R. A. Folk, C.-L. Xiang, S. Chen, D. E. Soltis, and N. Cellinese. 2020. A phylogenomic perspective on evolution and discordance in the alpine-arctic plant clade *Micranthes* (Saxifragaceae). *Frontiers in Plant Science* 10: 1773.

Sukumaran, J., and A. Holder. 2010. Dendropy: a Python library for phylogenetic computing. *Bioinformatics*. 26: 1569–1571.

Suzán-Azpiri, H., G. Sánchez-Rámos, J. G. Martínez-Avalos, S. Villa-Melgarejo, and M. Franco. 2002. Population structure of *Pinus nelsoni* Shaw, an endemic pinyon pine in Tamaulipas, Mexico. *Forest Ecology and Management* 165: 193–203.

Syring, J., K. Farrell, R. Businský, R. Cronn, and A. Liston. 2007. Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. *Systematic Biology* 56: 163–181.

Syring, J., A. Willyard, R. Cronn, and A. Liston. 2005. Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. *American Journal of Botany* 92: 2086–2100.

Szmidt, A., X. R. Wang, and M. Z. Lu. 1996. Empirical assessment of allozyme and RAPD variation in *Pinus sylvestris* (L.) using haploid tissue analysis. *Heredity* 76: 412–420.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.

Tang, C. Q., A. M. Humphreys, D. Fontaneto, and T. G. Barraclough. 2014. Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods in Ecology and Evolution*. 5: 1086–1094.

Than, C., and L. Nakhleh. 2009. Species tree inference by minimizing deep coalescences. *PLOS Computational Biology* 5: e1000501.

Than, C., D. Ruths, and L. Nakhleh. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9: 322.

Uckele, K. A., R. P. Adams, A. E. Schwarzbach, and T. L. Parchman. 2021. Genome-wide RAD sequencing resolves the evolutionary history of serrate leaf *Juniperus* and reveals discordance with chloroplast phylogeny. *Molecular Phylogenetics and Evolution* 156: 107022.

Vachaspati, P., and T. Warnow. 2015. ASTRID: Accurate Species TRees from Internode Distances. *BMC Genomics* 16: S.

Vidakovic, M. 1991. *Conifers Morphology and Variation*. Graficki Zavod Hrvatske, Zagreb.

Wang, H.-X., D. F. Morales-Briones, M. J. Moore, J. Wen, and H.-F. Wang. 2021. A phylogenomic perspective on gene tree conflict and character evolution in Caprifoliaceae using target enrichment data, with Zabelioideae recognized as a new subfamily. *Journal of Systematics and Evolution* 59: 897–914.

Wang, X.-Q., D, Tank, and T. Sang. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular Biology and Evolution* 17: 773–781.

Weeden, N. F., and J. F. Wendel. 1989. Genetics of plant isozymes. *In Isozymes in Plant Biology.* D. E Soltis [Ed.], pp. 46–72. Portland: Dioscorides Press.

Weitemier, K., S. C. K. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: 1400042

Wen, D., Y. Yu, and L. Nakhleh. 2016. Bayesian inference of reticulate phylogenies under the multispecies network coalescent*. PLoS Genetics* 12: e1006006.

Wendel, J. F., and J. Doyle. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. *In Molecular Systematics of Plants II: DNA Sequencing*. D. Soltis, P. Soltis, and J. J. Doyle [Eds.], pp. 265–296. Dordrecht: Kluwer Academic Publishers.

Whittall, J. B., J. Syring, M. Parks, J. Buenrostro, C. Dick, A. Liston, and R. Cronn. 2010. Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology* 19: 100–114.

Willyard, A., R. Cronn, and A. Liston. 2009. Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Molecular Phylogenetics and Evolution* 52: 498–511.

Willyard, A., D. S. Gernandt, B. Cooper, C. Douglas, K. Finch, H. Karemera, E. Lindberg, S. K. Langer, et al. 2021. Phylogenomics in the hard pines (*Pinus* subsection *Ponderosae*; Pinaceae) confirms paraphyly in *Pinus ponderosa*, and places *Pinus jeffreyi* with the California big cone pines. *Systematic Botany* 46: 538–561.

Willyard, A., J. Syring, D. S. Gernandt, A. Liston, and R. Cronn. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution* 24: 90–101.

Wolfe, J. A., and H. Schorn. 1990. *Taxonomic Revision of the Spermatopsida of the Oligocene Creede Flora, Southern Colorado*. U.S. Geological Survey Bulletin 1923. Denver, Colorado.

Wu, J., K. V. Krutovskii, and S. H. Strauss. 1999. Nuclear DNA diversity, population differentiation, and phylogenetic relationships in the California closed-cone pines based on RAPD and allozyme markers. *Genome* 42: 893–908.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586–1591.

Yang, Z. 2015. The BPP program for species tree estimation and species delimitation. *Current Zoology* 61: 854–865.

Yang, Z., and B. Rannala. 2012. Molecular phylogenetics: principles and practice. *Nature Reviews Genetics* 13: 304-14.

Yin, J., C. Zhang, and S. Mirarab. 2019. ASTRAL-MP: scaling ASTRAL to very large datasets using randomization and parallelization. *Bioinformatics* 35: 3961–3969.

Yu, Y., J. Dong, K. J. Liu, and L. Nakhleh. 2014. Maximum likelihood inference of reticulate evolutionary histories. *Proceedings of the National Academy of Sciences of the United States of America* 111: 16448–16453.

Zavarin, E., and K. Snajberk. 1986. Monoterpenoid differentiation in relation to the morphology of *Pinus discolor* and *Pinus johannis*. *Biochemical Systematics and Ecology* 14: 1–11.

Zhang, J., P. Kapli, P. Pavlidis, and A. Stamatakis. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 22: 2869–2876.

Zhang, C., H. A. Ogilvie, A. J. Drummond, and T. Stadler. 2018. Bayesian inference of species networks from multilocus sequence data. *Molecular Biology and Evolution* 35: 504–517.

Zhang, D., T. Xia, M. Yan, X. Dai, J. Xu, S. Li, and T. Yin. 2014. Genetic introgression and species boundary of two geographically overlapping pine species revealed by molecular markers. *PLoS ONE* 9: e101106.

Zhou, Y. F., R. J. Abbott, Z. Y. Jiang, F. K. Du, R. I. Milne, and J. Q. Liu. 2010. Gene flow and species delimitation: a case study of the two pines species with overlapping distributions in southeast China. *Evolution* 64: 2342–2352.

Zhou, B. F., S. Yuan, A. Crowl, Y. Y. Liang, Y. Shi, X. Y. Chen, Q-Q. An, M. Kang, P. Manos, and B. Wang. 2021. Evolutionary dynamics driving continental radiations of Fagaceae forests across the Northern Hemisphere. *Research Square*. DOI: https://doi.org/10.21203/rs.3.rs-968321/v1

Zimmer, E. A., and J. Wen. 2015. Using nuclear gene data for plant phylogenetics: progress and prospects. *Molecular Phylogenetics and Evolution* 65: 774–785.

Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. *In Evolving Genes and Proteins*. V. Bryson and H. J Vogel [Eds.], pp 97–166. New York: Academy Press.

7. APÉNDICES

7.1. Population structure, linkage disequilibrium, diversifying selection and local adaptation in *Pinus patula*.

Peláez, P., Ortíz Martínez, A., Figueroa Corona, L., **Montes, J.R.,** and Gernandt, D. S. (2020)

American Journal of Botany 107(11):1555–1566.
URL: https://doi.org/10.1002/ajb2.1566

American Journal of Botany

RESEARCH ARTICL

Population structure, diversifying selection, and local adaptation in *Pinus patula*

Pablo Peláez[1], Alfredo Ortiz-Martínez[2], Laura Figueroa-Corona[3], José Rubén Montes[2], and David S. Gernandt[1,4]

Manuscript received 13 February 2020; revision accepted 20 July 2020.

[1] Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México 04510, México

[2] Posgrado en Ciencias Biológicas, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México 04510, México

[3] Posgrado en Ciencias Biológicas, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México 04510, México

[4] Author for correspondence (e-mail: dgernandt@ib.unam.mx)

Citation: Peláez, P., A. Ortiz-Martínez, L. Figueroa-Corona, J. R. Montes, and D. S. Gernandt. 2020. Population structure, diversifying selection, and local adaptation in *Pinus patula*. *American Journal of Botany*. 107(11): 1–12.

doi:10.1002/ajb2.1566

PREMISE: Climate change is predicted to affect natural and plantation forests. The responses of conifers to overcome changing environments will depend on their adaptation to local conditions; however, intraspecific adaptive genetic variation is unknown for most gymnosperms. Studying genetic diversity associated with phenotypic variability along environmental gradients will enhance our understanding of adaptation and may reveal genetic pools important for conservation and management.

METHODS: We used target enrichment and genome skimming to obtain single nucleotide polymorphisms (SNPs) from 61 individuals of *Pinus patula*, a pine tree native to Mexico widely used in plantation forestry. We investigated the adaptive genetic variation of two varieties with morphological and distributional differences potentially related to genetic and adaptive divergence.

RESULTS: Population structure and haplotype network analyses revealed that genetic diversity between *P. patula* var. *patula* and *P. patula* var. *longipedunculata* was structured, even within populations of *P. patula* var. *longipedunculata*. We observed high genetic diversity, low inbreeding rate, and rapid linkage disequilibrium (LD) decay in the varieties. Based on outlier tests, loci showing signatures of natural selection were detected in geographically distant *P. patula* var. *longipedunculata* populations. For both varieties, we found significant correlations between climate-related environmental variation and SNP diversity at loci involved in abiotic stress, cell transport, defense, and cell wall biogenesis, pointing to local adaptation.

CONCLUSIONS: Overall, significant intraspecific adaptive genetic variation in *P. patula* was detected, highlighting the presence of different genetic pools and signs of local adaptation that should be considered in forestry and conservation.

KEY WORDS adaptation; climate change; forest trees; Hyb-Seq; natural selection; Pinaceae; single nucleotide polymorphisms.

Forests cover approximately one-third of the world's land surface and constitute a habitat for most terrestrial species (Duraiappah et al., 2005) and are important storehouses of carbon, with a very important contribution to climate regulation (Ellison et al., 2017). Predictions indicate that populations of tree species adapted to temperate and cold regions will change their composition and distribution with the arrival of warmer temperatures coupled with severe droughts (Allen et al., 2010). Changes in the tree species composition of forests generated by modifications in species distributions are predicted to alter the functionality of forests ecosystems (Morin et al., 2018). Therefore, understanding the responses of plants to climate change is crucial for the conservation of ecosystems (Duraiappah et al., 2005). Nowadays, it is recognized that intraspecific variation in functional traits resulting from dissimilar evolutionary histories of the populations of a species will produce different responses toward climate change (Pauls et al., 2013). Although intraspecific variation has been understudied in conifers, studies of the adaptation of conifers at the genotypic level have increased in number recently with the arrival of next-generation sequencing technologies (Prunier et al., 2016). Testing the extent of genetic adaptation to climate in conifers will help define future management of plantation and natural forests and accelerate breeding for complex traits (Prunier et al., 2016).

Identification of adaptive traits in pines is challenging due to their long generation times and their capacity for long-distance gene flow (Kremer et al., 2012). Conifer genome sizes range from ~8