UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE ECOLOGÍA

BIOLOGÍA EVOLUTIVA

Caracterización evolutiva de *Pinus pinceana* (Gordon & Glend.) a escalas filogenética, poblacional y funcional con aproximaciones genómicas

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

LAURA ALICIA FIGUEROA CORONA

TUTOR PRINCIPAL DE TESIS: DR. DANIEL PIÑERO DALMAU

INSTITUTO DE ECOLOGÍA, UNAM

COMITÉ TUTOR: DR. LUIS E. EGUIARTE FRUNS

INSTITUTO DE ECOLOGÍA, UNAM

DR. DAVID SEBASTIAN GERNANDT

INSTITUTO DE BIOLOGÍA, UNAM

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

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE ECOLOGÍA

BIOLOGÍA EVOLUTIVA

Caracterización evolutiva de *Pinus pinceana* (Gordon & Glend.) a escalas filogenética, poblacional y funcional con aproximaciones genómicas

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

LAURA ALICIA FIGUEROA CORONA

TUTOR PRINCIPAL DE TESIS: DR. DANIEL PIÑERO DALMAU

INSTITUTO DE ECOLOGÍA, UNAM

COMITÉ TUTOR: DR. LUIS E. EGUIARTE FRUNS

INSTITUTO DE ECOLOGÍA, UNAM

DR. DAVID SEBASTIAN GERNANDT

INSTITUTO DE BIOLOGÍA, UNAM

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

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS

ENTIDAD INSTITUTO DE ECOLOGÍA

OFICIO CPCB/1137/2021

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 el Subcomité de Biología Evolutiva y Ecología, del Posgrado en Ciencias Biológicas, en su **reunión ordinaria del día 16 de agosto de 2021** se aprobó el siguiente jurado para el examen de grado de **DOCTORA EN CIENCIAS** de la estudiante **FIGUEROA CORONA LAURA ALICIA,** con número de cuenta **304289071,** con la tesis titulada**: "CARACTERIZACIÓN EVOLUTIVA DE PINUS PINCEANA (GORDON & GLEND.) A ESCALAS FILOGENÉTICA, POBLACIONAL Y FUNCIONAL CON APROXIMACIONES GENÓMICAS**, Tutor Principal, quedando integrado de la siguiente manera:

Presidente:	DRA. ANA ELENA ESCALANTE HERNÁNDEZ
Vocal:	DRA. NATALIA IVALÚ CACHO GONZÁLEZ
Vocal:	DR. VICENTE DIEGO ORTEGA DEL VECCHYO
Vocal:	DRA. SVETLANA SHISHKOVA
Secretario:	DR. LUIS ENRIQUE EGUIARTE FRUNS

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 25 de noviembre de 2021

COORDINADOR DEL PROGRAMA

DR. ADOLFO GERARDO NAVARRO SIGÜENZA

Agradecimientos institucionales

Al Posgrado en Ciencias Biológicas de la Universidad Nacional Autónoma de México (UNAM), por la formación académica que me ha brindado y por el apoyo dado a lo largo de mis estudios de doctorado. Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca de manutención otorgada durante mis estudios de posgrado (449615), así como al Programa de Apoyo a los Estudios de Posgrado (PAEP) por el apoyo dado para la realización de la estancia de investigación.

Los financiamientos Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica, Dirección General de Asuntos del Personal Académico, UNAM IN209816 a cargo de D. S. Gernandt y los financiamientos otorgados por el Instituto de Ecología, UNAM a cargo de D. Piñero.

Quisiera agradecer sincera y profundamente al Dr. Daniel Piñero, tutor principal y director de esta tesis, por su participación intensiva para mejorar constantemente mis proyectos académicos, por formularme incentivos propositivos que me han hecho mejorar y aprender.

Finalmente, a los miembros de mi comité tutor: el Dr. Luis Eguiarte, el Dr. David Sebastian Gernandt y la Dra. Alejandra Moreno-Letelier, quienes constantemente aportaron su conocimiento para el desarrollo de este trabajo.

Agradecimientos personales

Agradezco a mi familia por el apoyo y la motivación.

A Daniel Piñero, por sus enseñanzas. Por su enorme generosidad, paciencia y sabiduría, por permitirme trabajar con la libertad de explotar nuevos horizontes, y brindarme la confianza para desarrollar planes de la A a la Z.

Agradezco a la Dra. Jill Wegrzyn por involucrarse en este proyecto y siempre a retarme a expandir límites. Al Institute for Systems Genomics Computational Biology Core, University of Connecticut (UCONN) por el acceso al software y hardware.

A todos los que contribuyeron al desarrollo de ésta tesis, especialmente a Patricia Delgado, Jill Wegrzyn, Alejandra Moreno-Letelier, Luis Eguiarte, David Gernandt, Diego Ortega Del Vecchyo, Pablo Peláez, Rubén Montes, y Daniel Piñero, coautores de capítulos de esta tesis, por su apoyo, revisión y esfuerzo por mejorar este trabajo. A los que contribuyeron al trabajo de campo del que este trabajo es producto: Carmen Bautista, Patricia Delgado, Celestino Flores, Erick García, David Gernandt, Socorro González, Azalea Guerra, Ulises Manzanilla, Rubén Montes, Alejandra Moreno, Rodolfo Salas, Alejandra Vázquez-Lobo y Alfredo Villarruel. Y los que me asesoraron en el procesamiento bioinformático: Alejandro Flores, Gustavo Giles, Azalea Guerra, Susan McEvoy, Verónica Reyes, Alex Trouern-Trend, Cynthia Webster y Sumaira Zaman.

A los miembros del Laboratorio de Genética y Evolución y del Plant Computational Genomics Lab; a los técnicos y auxiliares que me facilitaron el material y las herramientas para el desarrollo experimental y el procesamiento de datos. A Erick García por las recomendaciones. A mis amigos, a Fernanda Heredia quien revisó la redacción de este texto a Agustín Ávila, Alicia Gamboa, Violeta García, Yocelyn Gutíerrez, Brisa De Gante, Lev Jardón, Rodolfo Salas, Alejandra Vázquez-Lobo y Sumaira Zaman cuya amistad me ancló a la cordura y calmó las frustraciones.

Finalmente, un agradecimiento especial por los comentarios, correcciones y sugerencias de los revisores de la tesis: Dra. Ivalú Cacho González, Dra. Ana Elena Escalante Hernández, Dr. Luis Eguiarte Fruns, Dr. Diego Ortega Del Vecchyo y Dra. Svetlana Shiskova.

Índice

Abreviaciones

ABA	Ácido abscísico
bp	Base pair
CDS	Coding DNA sequence
ChD	Desierto Chihuahuense
cpDNA	Chloroplast DNA
Gbp	Gigabase pairs; mil millones de nucleótidos
ILS	Incomplete lineage sorting
IUCN	International Union for Conservation of Nature
IR	Repetición invertida
JA	Ácido jasmónico
kya	Thousand years ago
LIG	Last Interglacial
LRR	Leucine-rich repeat
LTR-RT	Long terminal repeat (LTR) retrotransposons
LSC	Large single-copy region
Ma	Millón de años
mRNA	*Messenger RNA*
ROS	Reactive oxygen species
rRNA	ribosomal RNA
SA	Ácido salicílico
SMO	Sierra Madre Oriental
SNP	Single nucleotide polymorphisms
SSC	Small single-copy region
SSRs	Microsatellites o Single Sequence Repeats
tRNA	transfer RNA

Resumen

La respuesta a la aridez constituye un rasgo poligénico que es resultado de interacciones del genotipo - fenotipo con el ambiente; además genera mecanismos ecofisiológicos, y genéticos que están interrelacionados, y que son codependientes y complementarios, para producir una respuesta plástica y/o adaptativa. En este trabajo busco identificar los cambios en la respuesta transcripcional a la aridez en *Pinus pinceana* (Gordon & Glend.) y el efecto de las fuerzas evolutivas que han modificado su arquitectura genómica. En el Primer Capítulo de esta tesis describo los estudios de los niveles de expresión del mRNA, como aproximación de la respuesta fisiológica a los factores bióticos y abióticos en su ambiente. Para esto analicé el transcriptoma, buscando resolver la relación entre la divergencia ambiental y las diferencias en los perfiles de expresión. El estudio del transcriptoma con un ensamble *de novo* caracterizó 45,431 anotaciones del tejido diploide de acículas de siete árboles maduros, y el análisis de expresión diferencial identificó reguladores fitohormonales de ABA y JA y la regulación de ROS posiblemente involucrados en la respuesta al estrés abiótico. Los perfiles de expresión genética resaltaron la expresión diferencial de 26 CDS involucrados en la regulación de respuesta al estrés abiótico, y proteínas con receptores LRR relacionadas con el reconocimiento de patógenos. En el Segundo Capítulo analizo la diversidad genética en la especie, así como su historia demográfica y patrón filogeográfico. Encontré dos grupos genéticos generados por la dinámica interglacial durante los cambios climáticos del Pleistoceno (~627 kya), seguidos de la subfragmentación de ambas regiones (~127.7 - ~539.2 kya), el Desierto Chihuahuense entre su región este y oeste, y la Sierra Madre Oriental en su porción sur y norte, mientras la aridificación aumentaba en el Desierto Chihuahuense. En el Tercer Capítulo estudio los patrones de divergencia dentro y entre los dos clústers genéticos y sus especies hermanas, reconstruyendo las relaciones filogenéticas y la divergencia de los pinos piñoneros de conos grandes a partir de datos del genoma plastídico y la concatenación de SNPs de genes nucleares altamente conservados. Los resultados apoyan la estructuración de *P. pinceana* en dos grupos, descartando para cualquiera de las especies del clado eventos de introgresión o eventos de migración con *P. maximartinezii*.

Abstract

The response to aridity constitutes a polygenic trait that is the result of genotype-phenotype interactions with the environment; it also generates ecophysiological and genetic mechanisms that are interrelated, and that are codependent and complementary, to produce a plastic and/or adaptative response. In this work, I seek to identify the changes in the transcriptional response to aridity in *Pinus pinceana* (Gordon & Glend.) And the effect of evolutionary forces that have modified its genomic architecture. In the First Chapter of this thesis, I describe the studies of mRNA transcription, as an approximation of physiological response to biotic and abiotic factors in their environment. For this, I analyzed the transcriptome, seeking to resolve the relationship between environmental divergence and differences in expression profiles. The study of the transcriptome with a *de novo* assembly characterized 45,431 annotations of diploid tissue of needles from seven mature trees, and the differential expression analysis identified phytohormonal regulators of ABA and JA and the regulation of ROS possibly involved in the response to abiotic stress. The gene expression profiles highlighted the differential expression of 26 CDS involved in the regulation of the response to abiotic stress and proteins with LRR receptors related to the recognition of pathogens. In the Second Chapter, I analyze the genetic diversity in the species, as well as its demographic history and phylogeographic pattern. I found two genetic groups generated by interglacial dynamics during the Pleistocene climate changes (∼627 kya), followed by the subfragmentation of both regions (~ 127.7 - ~ 539.2 kya), the Chihuahuan Desert between its eastern and western regions, and the Sierra Madre Oriental in its southern and northern portions, while aridification increased in the Chihuahuan Desert. In the Third Chapter, I study the patterns of divergence within and between the two genetic clusters and their sister species, reconstructing the phylogenetic relationships and divergence of large cone stone pines from plastidic genome data and the concatenation of nuclear gene SNPs highly preserved. The results support the structuring of *P. pinceana* into two groups, ruling out introgression events or migration events with *P. maximartinezii* for any of the species of the clade.

Introducción

La adaptación, la plasticidad fenotípica y la historia demográfica, y su efecto en la evolución

La respuesta a factores abióticos y bióticos influye en el proceso evolutivo y repercute en los patrones de variación genómica (Hand et al., 2015). El reconocimiento de las rutas de respuesta a estos factores es importante porque permite entender la tolerancia al estrés, la adaptación y la interacción entre el genotipo, el fenotipo y el ambiente. Los mecanismos evolutivos afectan la distribución de la diversidad genética de las poblaciones, y para detectar estos mecanismos se analizan las desviaciones en los modelos neutros (Leigh, 2007). Los estudios de asociación genotipo - fenotipo - ambiente buscan ubicar los caracteres genéticos que determinan la varianza fenotípica y establecen correlaciones con la divergencia del ambiente. Un ejemplo de estos estudios es el análisis de patrones de expresión, que bajo determinadas condiciones, edades o tejidos, en un panorama comparativo y multilocus, describen la respuesta fisiológica (Nosil et al., 2009).

1. La adaptación y la plasticidad en la respuesta a la aridez

Evolutivamente, la regulación de la humedad representa una estrategia clave para el establecimiento de las plantas en ambientes terrestres. La aridez, entendida como el déficit de recursos hídricos en el suelo y la falta de humedad en el aire (Pompa-García, 2017), representa una de las presiones más agresivas para el crecimiento y desarrollo de los organismos sésiles (Le & McQueen-Mason, 2006). Ante fluctuaciones en las condiciones hídricas, los organismos tienen tres escenarios posibles: 1) evadir la exposición a los factores y condiciones estresantes; 2) reducir los niveles de deshidratación con latencia o de reducción de biomasa; o 3) tolerar los cambios ambientales implementando mecanismos de resistencia que contrarresten los daños (Farooq et al., 2012).

Estos escenarios de respuesta al estrés abiótico están interrelacionados, son co-dependientes y complementarios, y es evolutivamente posible que generen respuestas adaptativas de un organismo para establecerse en un ambiente árido. Estas respuestas se clasifican en respuestas proximales, donde la respuesta fisiológica es plástica; son importantes en ambientes heterogéneos pues confieren varianza fenotípica que puede

proporcionar una mayor tolerancia ambiental (Via et al., 1995); y respuestas evolutivas, producidas por cambios y polimorfismos genéticos. Los cambios plásticos en el fenotipo pueden ser adaptativos o no; pero, si existe variabilidad genética apropiada, las respuestas plásticas son potencialmente adaptativas (Via et al., 1995, Schlighting & Levin, 1986).

Los efectos de la aridez repercuten en todas las respuestas fisiológicas y metabólicas de un organismo (Zhu et al., 2002). Hernández et al. (2012) han descrito una conservación de estrategias comunes a la respuesta al congelamiento, la aridez y la salinidad en los linajes de plantas. Los efectos netos en la respuesta a la aridez ocurren a nivel morfológico, fisiológico y molecular, e involucran tres puntos: 1) optimizar el consumo y almacenamiento de agua; 2) reducir la pérdida de agua; y, 3) prevenir el daño celular (De Micco et al., 2012; Huang et al., 2012).

Los cambios morfológicos y de desarrollo involucrados en la tolerancia a la desecación y la regulación del balance del contenido de agua se presentan en todas las plantas terrestres. Estos cambios pueden ser la evolución de la cobertura de esporopolenina en polen y esporas, un politerpeno impermeable, o la aparición de la cutícula; y fisiológicamente a una escala temporal relativamente larga en la vida de una planta ocurren cambios en la arquitectura de la raíz y depósitos de cera en hojas, mientras que a corto plazo tenemos procesos como el cierre de los estomas y el marchitamiento (Vilagrosa et al., 2012; Le & McQueen-Mason, 2006; De Micco et al., 2012). Los mecanismos fisiológicos comprenden las vías de regulación y abastecimiento de factores fitohormonales, osmoprotectores y la regulación de agentes oxidantes (McDowell et al., 2008; Moran et al., 2017).

Tardieu et al. (1998) clasifican la regulación de la evapotranspiración en dos tipos: 1) la regulación isohídrica, en la que se produce el cierre de los estomas implicando la vía de señalización de ácido abscísico (ABA), se presenta en organismos con traqueidas gruesas en coníferas; este tipo de regulación se presenta en la familia Pinaceae; 2) la regulación anisohídrica, que permite la pérdida de recursos hídricos y genera cambios del potencial hídrico, como señal para el cierre de estomas. Este último tipo de regulación es característica de las familias Cupressaceae y Taxaceae, y se ha descrito que el tejido vascular de organismos de regulación anisohídrica tiende a ser más resistente y con menor tendencia a la cavitación (McDowell et al., 2008; Brodribb et al., 2014).

Las estrategias isohídricas son frecuentemente asociadas a plantas con mayor sensibilidad a déficits hídricos. En las coníferas, se presenta la combinación de estrategias para el transporte de agua, y la eficacia de su regulación ha sido asociada al ambiente y la altura del tronco (Brodribb et al., 2014). Hochberg et al. (2017) han señalado que la clasificación de los mecanismos de regulación hídricos no es tajante y es más un continuo, pues 1) la misma especie o especies dentro de la misma familia pueden presentar diferencias en su control hídrico (*p.e*. *Vitis vinifera;* Cupressaceae y Callitroideae; *Populus alba*; Hochberg et al., 2017; Pittermann et al., 2012; Zhang et al., 2020), y 2) al ser una respuesta gradual, los mecanismos no son excluyentes.

La relevancia de la regulación por ABA ha formado parte de los argumentos para explicar la conquista del medio terrestre de las plantas, ya que su participación en la regulación del estrés hídrico y como promotor del desarrollo ha sido documentado en los clados más basales de plantas (Le & McQueen-Mason, 2006; Sun et al., 2020). Pittermann et al. (2012) describen basados en paleoclimas que la diversificación dentro de Cupressaceae partió de ancestros intolerantes a la sequía en hábitats húmedos durante el Mesozoico, y que los cambios climáticos durante los últimos 150 Ma, especialmente el aumento de la aridez desde el Oligoceno, han producido modificaciones en la regulación hídrica, variaciones estructurales del xilema, una menor eficiencia de transporte del xilema y, a nivel de las hojas, una menor capacidad fotosintética entre Cupressaceae y Callitroideae.

1.1 Regulación fitohormonal

Las fitohormonas son los reguladores fundamentales del desarrollo, el crecimiento, la diferenciación y la senescencia de las plantas (Davies, 2010); son necesarias en concentraciones mínimas para dirigir vías de señalización (Farooq et al., 2009a). Las fitohormonas que están involucradas directamente en la respuesta al déficit hídrico son el ácido abscísico (ABA) y el ácido jasmónico (JA); sin embargo, el aumento de estos dos compuestos desencadena la disminución de auxinas, giberelinas, y citoquininas, principales promotores del crecimiento de plantas (Taiz & Zeiger, 2010).

El ABA promueve la latencia del embrión, genera la regulación de la respuesta al estrés osmótico provocado por la sequía, el aumento de salinidad y el congelamiento. Inicialmente induce el cierre de estomas mediante el cambio de turgencia en las células

guardas, activando los canales de transporte de Ca^{2+} y de K^{+} (Corcuera et al., 2012a, b). Sin embargo, se estima que en cambios ambientales repentinos y a medida que la altura de los organismos aumenta la señalización de regulación estomática por ABA se vuelve ineficiente (Moran et al., 2017).

El JA participa en los procesos de maduración de frutos en angiospermas, el desarrollo del polen y el crecimiento de la raíz. Además, actúa como regulador celular de la respuesta de defensa al daño por patógenos, por exposición a ozono y durante el déficit hídrico (Creelman & Mullet, 1995). Durante periodos de estrés, el JA promueve la síntesis de fosfatasas e inhibidores de proteinasas, y enzimas involucradas en la biosíntesis de flavonoides como la chalcona sintasa y la biosíntesis de terpenoides. Estos factores generan la reducción de agentes oxidantes y favorecen la estabilización de la membrana (Creelman & Mullet, 1995).

1.2 Regulación del estrés oxidativo

Debido al suministro limitado de agua y el cierre de estomas, el estrés oxidativo aumenta (Kiani et al., 2007) y produce una disminución de CO_2, que a su vez genera alteraciones en la actividad enzimática y conduce a la producción y acumulación de especies reactivas de oxígeno (ROS) como $^1O_2$, O_2^- , H_2O_2, OH^-, RO^*, ROO^* y $ROOH^*$ (Flexas-Medrano, 2002; Lawlor & Cornic, 2002). Las propiedades oxidantes de las ROS producen daño a lípidos, proteínas y otras macromoléculas (Rout & Shaw, 2001). Al aumentar su concentración, se activan dos estrategias: una involucrala reducción de ROS y la segunda promueve la protección de enzimas, DNA y membranas.

Los intermediarios osmoreguladores y osmoprotectores más importantes son salicilatos y flavonoides (Scandalios et al., 2005; Li et al., 2008; Ozkur et al., 2009). Los flavonoides son compuestos polifenólicos, solubles en agua, como el α-tocoferol, el glutatión, y los β-carotenos, que están involucrados en la filtración UV, y en la fijación simbiótica de nitrógeno. Los flavonoides actúan también como reguladores del ciclo celular, y adicionalmente en la raíz facilitan la formación de nódulos fijadores de Nitrógeno e interacciones micorrícicas.

1.3 Producción de carbohidratos osmoprotectores

Cuando hay déficit de agua, la tasa fotosintética disminuye, se dispara el uso de las reservas de carbono al reducir el almidón en sacarosa y tetralosa, metabolitos que contribuyen al mantenimiento del turgor y la prevención de la desnaturalización de proteínas y la fusión de membranas (Le & McQueen-Mason, 2006). Estos osmoprotectores están relacionados con la protección y reparación del DNA, la estabilización y la reconstrucción de la membrana. Cambios abruptos en salinidad disparan la formación residual de alcoholes polihídricos como manitol, sorbitol y pinitol que regulan la osmolaridad, por lo que se reduce el flujo extracelular y se favorece la retención de agua (Bacelar et al., 2012). Al mismo tiempo que la producción de prolina y compuestos monoméricos de glicina (betaína o trimetilglicina) y las cetonas (chalconas) es promovida por el incremento de aridez, salinidad y temperatura, se logra la estabilización de la estructura cuaternaria de enzimas, complejos proteicos, y la membrana del tilacoide (Xing-Rajashekar, 1999).

2. Enfoques transcripcionales aplicados al estudio de la respuesta a la aridez

Actualmente, la respuesta al déficit hídrico se puede estudiar a partir de análisis ecofisiológicos, que conjuntan estudios genómicos, mapeos genéticos (Mackay & Dean, 2011) y la descripción de perfiles transcripcionales que han facilitado el entendimiento de la interacción genotipo-fenotipo con el ambiente, además de favorecer la identificación de genes responsables de la respuesta a la desecación (Cañas et al., 2015; Kramer & Boyer, 2015).

El ensamble de un transcriptoma constituye la medición y caracterización de la actividad transcripcional en una condición determinada en un momento específico (Brown, 2002). Ayuda a la comprensión de las redes de regulación en las estrategias de resistencia y susceptibilidad al estrés abiótico (Mackay et al., 2012). En esta tesis en el Primer Capítulo describo la respuesta al estrés biótico y abiótico en *Pinus pinceana* Gordon & Glend. en su ambiente natural, el cual es heterogéneo en cuanto a la precipitación y humedad.

Estos enfoques genómicos han hecho posible distinguir las respuestas plásticas de los mecanismos evolutivos adaptativos. Para que exista adaptación en términos evolutivos, es necesario que haya selección y que existan bases genéticas; en otras palabras, que sea

heredable. Este proceso selectivo implica cambios y frecuencias alélicas en genes que afectan el fenotipo, relacionado con diferencias en la adecuación de los individuos entre los distintos hábitats. Estas diferencias adaptativas pueden estar dadas por (o relacionadas con) distintas tasas de crecimiento, capacidades reproductivas y de sobrevivencia (Kuparinen et al., 2010). Las variaciones adaptativas pueden aparecer a diferentes escalas, desde una nueva morfología hasta un cambio en los niveles de expresión (Romero et al., 2012). Los estudios que abordan la adaptación local de las poblaciones forestales cubren escalas espaciales amplias entre ambientes heterogéneos y evalúan la varianza genética en rasgos cuantitativos y poligénicos (Morgenstern, 1996; Eckert et al., 2010).

Sin embargo, es importante señalar que los efectos de la selección natural dependen de la arquitectura genética del rasgo fenotípico, su relación con la adecuación, la historia demográfica de las poblaciones, el tipo y fuerza de selección, además del momento del inicio de la selección (Linnen & Hoekstra, 2009). Sin desestimar la detección de genes candidatos bajo selección natural, es importante considerar que el efecto y las interacciones de otras fuerzas evolutivas como la migración y la historia demográfica modifican la diversidad genética de manera similar (Mitton et al., 1989).

Los métodos que buscan identificar adaptación en poblaciones detectan valores atípicos en la frecuencia de alelos entre las poblaciones, comparando la diferenciación poblacional o buscan estructuras de varianza-covarianza entre poblaciones con los factores ambientales (Nielsen, 2005; Gunther & Coop, 2013; Rellstab et al., 2015).

3. Los efectos de la historia demográfica en la diversidad genética

Al analizar la variación genética, es difícil distinguir los efectos selectivos de los efectos histórico-demográficos, debido a que los cambios en el tamaño de la población dejan huella en la distribución de frecuencias de la diversidad genética (Johri et al., 2021). Considerando la variación neutral en un solo locus, las predicciones del modelo de un cuello de botella, y los efectos de un barrido selectivo generan el mismo patrón de variación genética (Tajima 1989; Depaulis et al., 2003, 2009). Por lo tanto, es importante reconocer que, para distinguir estos procesos, se deben analizar polimorfismos en varios loci, ya que los eventos demográficos influyen en todo el genoma. En contraste, la selección afecta solo a regiones específicas del genoma (Rellstab et al., 2015).

Para la identificación del efecto de la demografía histórica en la distribución de las frecuencias alélicas y la diversidad genética se utilizan estadísticos descriptivos como la *D* de Tajima o la *H* de Fay y Wu y el espectro de frecuencias de alelos (Tajima, 1989; Fay & Wu, 2000; Przeworki, 2002; Depaulis et al., 2003). Los valores negativos de *D* indican poblaciones con expansiones recientes, efectos de selección purificadora o un barrido selectivo. Estos valores negativos de *D* indican una distribución de las frecuencias de los alelos que difiere de la expectativa neutral con un exceso de alelos raros o derivados. Valores de *D* positivos evidencian que las poblaciones han pasado por selección balanceadora o cuellos de botella, y han reducido sus niveles de diversidad (Nei et al., 1975, Nielsen & Slatkin, 2013).

En el Segundo Capítulo de esta tesis analizo la demografía histórica de *P. pinceana* y la distribución geográfica de la diversidad genética en un contexto geológico, y de los cambios climáticos que han ocurrido en el Desierto Chihuahuense en el Pleistoceno utilizando polimorfismos del genoma completo.

4. Enfoques genéticos y genómicos a la ecología evolutiva en pinos

Con centro de origen en Asia, el género *Pinus* está actualmente distribuido en el hemisferio norte (Saladin et al., 2017; Wei-Tao et al., 2021). Las especies extendieron el rango de distribución a partir de la migración a través de los sistemas montañosos y se adaptaron a diferentes gradientes de elevación (Perry, 1991; Richardson, 1998; Farjon & Styles, 2010). En México, con una orografía y amplia heterogeneidad ambiental, el género *Pinus* presenta una gran diversidad; de hecho, es el país con mayor diversidad de pinos del mundo, donde se han descrito entre 47 y 51 de las ca. de 111 especies existentes (Richardson et al., 1998; Perry, 1991; Perry et al., 2000).

En *Pinus* se han reportado genomas extremadamente grandes y complejos, que varían de 16 a 35 Gbp (De la Torre et al., 2014; Leitch et al., 2012). A pesar de los costos decrecientes de la secuenciación, la resecuenciación del genoma completo en múltiples individuos aún no es factible. Actualmente, existen dos genomas de referencia para el género: uno de *P. taeda* (22.1 Gbp, con 50,172 genes que codifican a proteínas) y otro de *P. lambertiana* (30 Gbp, 38,518 CDS; Wegrzyn et al., 2014; Zimin et al., 2014; Scott et al., 2020). De la Torre et al., (2014) atribuyen el gran tamaño del genoma en *Pinus* a la

acumulación masiva y eliminación deficiente de elementos transponibles (retrotransposones) con secuencias terminales de repetidos largos (LTR-RT). Los elementos transponibles constituyen el 74% de los genomas de *P. taeda*, y el 79% en *P. lambertiana* (Neale et al., 2014; et al., 2017; Wan et al., 2018; Wegrzyn et al., 2014).

Para el genoma plastídico en las coniferas se hereda paternalmente y, por lo tanto, rastrea la historia evolutiva de la dispersión del polen, independiente del genoma nuclear, y con frecuencia es independiente del genoma mitocondrial (Neale & Sederoff, 1989; Whittall et al., 2010). Se puede obtener fácilmente una historia genealógica única e independiente para la comparación con el genoma nuclear que separen las contribuciones parentales (polen vs semilla). Los plastomas de coníferas también se caracterizan por extensos reordenamientos genómicos y estudios recientes describen cambios principalmente en las dos repeticiones invertidas (IR) de 20 a 30 Kb (IRA e IRB), que generalmente contienen cuatro RNA ribosomales que en Pinaceae aparecen con reordenamientos y delecciones en la región IRB, además de la recombinación de IR cortos (Wu et al., 2011a,b; Hsu et al., 2016).

Tras analizar el genoma de las coníferas, Chen et al. (2013) describen que existe una tendencia por conservar los genes del metabolismo general sin duplicaciones, mientras que las secuencias de proteínas involucradas en el metabolismo especializado y secundario tienden a ser redundantes. Además, se ha reportado un número elevado de genes que codifican para las cinasas con repeticiones ricas en leucina, citocromo P450, y los factores de transcripción MYB (De la Torre et al., 2015; Neale et al., 2014; Pavy et al., 2013; Porth et al., 2011; Warren et al., 2015). Debido al gran tamaño de sus genomas, estrategias de secuenciación de RNA (RNA-seq) y RAD-seq, que se basan en un muestreo fragmentado del genoma representan las mejores alternativas para el genotipado de miles de loci (Schlötterer et al., 2014; Hoban et al., 2016; Jones et al., 2016; Oney-Birol et al., 2018).

La regulación genética de la respuesta a la aridez en los pinos ha sido poco explorada, y, aunque se han propuesto algunas etapas de las vías de señalización de la respuesta al estrés abiótico, para ninguna de ellas se elucidaron intermediarios o sustratos y productos (Balderas-Hernández et al., 2013). En general, el análisis de perfiles transcripcionales en plantas bajo condiciones de estrés sugiere la conservación de respuestas a la aridez, el congelamiento y la salinidad entre angiospermas y gimnospermas (Díaz-Sala, 2013; Moran et al., 2017). Los diseños experimentales que se han explotarado con mayor

detalle para reconocer la respuesta al déficit hídrico se han elaborado en plántulas de *P. taeda* y *P. pinaster* (Lorenz, 2006; Watkinson et al., 2003; Paiva et al., 2008) con el objetivo de reconocer la regulación de las vías de respuesta a la humedad, los patrones de crecimiento y la modificación en los procesos de lignificación.

A nivel genético, los estudios adaptativos en los pinos han reportado niveles altos de d_N / d_S (proporción de sustituciones no sinónimas y sinónimas) en comparación con angiospermas modelo. Esto ha sido interpretado como el efecto de presiones de selección fuertes, probablemente debido a tamaños de población efectivos grandes (De la Torre et al., 2017), mientras que la diferenciación poblacional es alta debido a los niveles de diversidad genética dentro de las poblaciones, y el flujo genético a larga distancia por la dispersión del polen entre poblaciones conduce a un rápido decaimiento del desequilibrio de ligamiento (De la Torre et al., 2014; et al., 2017; Porth & El-Kassaby, 2014).

En general, para las coníferas, los estudios adaptativos se han centrado en preguntas relacionadas con la resistencia al estrés térmico y a la formación de madera (Garnier-Géré & Ades, 2001; Aitken & Hannerz, 2001), elaborados a partir de una selección previa de genes o polimorfismos candidatos identificados por medio de estudios funcionales o de expresión (Neale & Savolainen, 2004) con la hipótesis de que estos loci candidatos presentarán valores de diversidad y diferenciación inusual entre las poblaciones por efecto de la selección natural (Luikart et al., 2003; Kelley et al., 2006). Mosca et al. (2016) describen para *Abies alba, Larix decidua, P. cembra* y *P. mugo*, coníferas de distribución alpina en Europa de uno a siete polimorfismos, dependiendo del género, asociados a la resistencia al congelamiento. En respuesta al estrés térmico por congelamiento o aridez en *Pseudotsuga menziesii* Eckert et al. (2009) reportaron 10 marcadores asociados a la respuesta de resistencia y adaptación al congelamiento. Otros estudios han detectado cambios correlacionados con diferencias fenológicas en *Picea mariana* (Prunier et al., 2011), y cambios ligados a diferencias en el crecimiento en la madera de *Picea sitchensis* (Holliday et al., 2010).

En México, los estudios de adaptación local en coníferas no han encontrado correlaciones entre los mecanismos genéticos de respuesta al estrés abiótico y los cambios ambientales. Pero se ha podido correlacionar la diversidad genética en *Pinus* con los procesos climáticos a nivel histórico para ubicar patrones filogeográficos (Cuenca et al., 2003; Jardón-Barbolla et al., 2011; Rebolledo-Camacho et al., 2018; Moreno-Letelier et al., 2008). Eventos

como los ciclos interglaciares del Pleistoceno y el surgimiento de la Faja Volcánica Transmexicana generaron cambios en los patrones de distribución de la diversidad genética en *Abies* (Giles-Pérez, 2017; Méndez-González et al., 2017), *Juniperus* (Moreno-Letelier et al., 2014; Reyes-Galindo, 2016), *Picea* (Jaramillo-Correa et al., 2006; Quiñones-Pérez et al., 2014), *Taxodium* (Hernández-Leal, 2009) y *Pinus* (Cuenca et al., 2003; Jardón-Barbolla et al., 2011; Rebolledo-Camacho et al., 2018; Moreno-Letelier et al., 2008).

Entre los estudios evolutivos y de historia demográfica en *Pinus* distribuidos en México debemos mencionar el de Cuenca et al. (2003), que describe en *P. nelsonii* una colonización a larga distancia entre 125,000 y 309,000 años y una expansión de 425 veces su tamaño poblacional entre 59,000 y 146,000 años atrás. Jardón-Barbolla et al. (2011) y Rebolledo-Camacho et al. (2018) describen en *P. caribaea* una reducción drástica en el tamaño demográfico, por lo que sugirieren que la endogamia ha jugado un papel importante en la disminución de la variación genética, con una dinámica demográfica de eventos de contracción y expansión generados por cambios climáticos de sequía e inundaciones extremas ocurridos durante los últimos 10,000 años. Moreno-Letelier et al. (2008) encontraron en *P. strobiformis* una diferenciación genética en dos grupos separados por el Desierto Chihuahuense, que han mantenido flujo genético durante las etapas glaciales y ciclos de expansiones poblacionales, favoreciendo la conectividad entre grupos.

5. Los pinos piñoneros (subsección *Cembroides*)

Distribuidas en ambientes áridos de América del Norte, las 13 especies de pinos piñoneros (subsección *Cembroides*) tienen semillas agrandadas con un megagametofito rico en lípidos, con un ala vestigial o sin ella. Características como el número y la rigidez de las acículas, la longitud del cono y el tamaño de la semilla son indicadores de la divergencia morfológica (Malusa, 1992; Farjon, 1996; Montes et al., 2019). Estudios sugieren que las semillas agrandadas sin alas evolucionaron en dos o tres eventos independientes de un ancestro común con una semilla pequeña y alada (Gernandt, 2005).

En especies de conos pequeños de la subsección *Cembroides* se han detectado eventos de introgresión y flujo genético interespecífico. Lanner & Phillips (1992) analizaron la introgresión de *P. edulis* y *P. monophylla* a partir de la variación en caracteres morfológicos de poblaciones simpátricas. Little (1968) sugirió la posibilidad de eventos de hibridación

entre *P. remota*, y *P. edulis*. Por su parte Montes et al., (2019) han descartado la hibridación entre *P. quadrifolia* y *P. monophylla* propuesta por Lanner (1974a,b) para explicar la variación en el número de acículas por fascículo de *P. quadrifolia*. Los eventos en el clado de pinos piñoneros de conos pequeños han sido favorecidos por la superposición de fenología en poblaciones simpátricas, la disperción aerófila del polen y las débiles barreras precigoticas dada su reciente divergencia.

Ortiz-Medrano et al. (2016) describen que, en ambientes donde se incrementa la temperatura y disminuye la humedad se incrementa la cubierta de la semilla, las dimensiones del cono y la altura del árbol, y disminuye el número de acículas por fascículo. Las especies de linajes con divergencias más ancestrales de los pinos piñoneros de México son también las especies que habitan áreas con condiciones ambientales más húmedas, lo que sugiere que los caracteres ancestrales del grupo están relacionados con ambientes húmedos y fríos, a partir de los cuales el resto de las especies han divergido y colonizado ambientes más áridos (Ortiz-Medrano et al., 2016).

En el Tercer Capítulo, analizo las divergencias y cambios en los linajes de pinos piñoneros de conos grandes, para reconstruir las relaciones filogenéticas comparando polimorfismos en regiones de DNA nuclear y el genoma plastídico a fin de resolver la hipótesis propuesta por Ledig et al., (2009), que sugieren que la divergencia de los pinos piñoneros de conos grandes ha presentado eventos de introgresión.

5.1. *Pinus pinceana* Gordon & Glend.

Pinus pinceana es un pino piñonero endémico de México que se distribuye en suelos calcáreos y rocosos (Farjon & Styles, 1997). Se conocen pocas poblaciones, las cuales son pequeñas, dispersas y discontinuas, y están distribuidas a lo largo de la Sierra Madre Oriental y el Desierto Chihuahuense (1480 - 3000 msnm). Esta especie forma parte de las especies de pinos que han sido descritas como más tolerante a condiciones de sequía (Luna-Cavazos et al., 2010; Ortiz-Medrano et al., 2016). La especie está amenazada por su distribución limitada, el aumento en la erosión del suelo, el sobrepastoreo y los incendios inducidos (Martiñón-Martínez et al., 2010; Figueroa-Corona, 2016). Está catalogada como especie en peligro de extinción en la lista de especies amenazadas de la legislación nacional y en el índice internacional de la Unión Internacional para la Conservación de la Naturaleza (IUCN,

2010; NOM 059-SEMARNAT-2010; https://www.biodiversidad.gob.mx/especies/catRiesMexico).

Estudios genéticos a partir de diferentes marcadores en *P. pinceana* han mostrado que existe una alta diferenciación genética entre sus poblaciones (Escalante, 2001; Ledig et al., 2001; Molina-Freaner et al., 2001) en comparación con otras especies de pinos (*e.g.*, Delgado et al., 1999; Cuenca, 2003; Karhu et al., 2006). Escalante (2001) determinó con evidencia de microsatelites plastídicos que existen dos grupos genéticos, los cuales se han mantenido sin flujo genético.

A nivel morfológico, Martiñón-Martínez et al. (2010) han descrito que las poblaciones de *P. pinceana* presentan cambios en la longitud de su raíz, correlacionados con la localidad geográfica. Ramírez-Herrera et al. (2008) encontraron una relación positiva entre la longitud de las raíces y la abundancia de cera en las hojas con la ubicación geográfica de los organismos; por tanto, sugirien una adaptación diferencial al estrés hídrico y las altas temperaturas; siendo que, en las poblaciones del Desierto Chihuahuense hay un mejor desempeño de los individuos en condiciones de estrés hídrico y de temperatura elevada.

El texto está estructurado para reconocer en *Pinus pinceana* la respuesta transcripcional a factores bióticos y abiótico en su particularidad ambiental actual, para ir desarrollando el efecto de la historia demográfica y la filogeografía en la diversidad genética, y, finalmente, reconocer los patrones de divergencia de *P. pinceana* con sus especies hermanas.

Objetivos

- Ensamblar *de novo* el transcriptoma de acículas de árboles maduros de *Pinus pinceana* (Gordon & Glend.).
- Identificar los perfiles de expresión diferencial de *Pinus pinceana* y correlacionar las diferencias a la regulación de respuesta a factores abióticos y bióticos presentes en la distrubución.
- Inferir cómo los procesos geológicos y climáticos durante el Pleistoceno han influido en la variación genética y la historia demográfica de *P. pinceana* utilizando polimorfismos en todo el genoma.
- Reconocer la variación intraespecífica del genoma plastídico y polimorfismos en genes nucleares en los pinos piñoneros de conos grandes.
- Analizar la divergencia entre *P. maximartinezii* (Rzed.) y *P. pinceana* e identificar posibles eventos de reticulación entre los linajes de los pinos piñoneros de conos grandes.

Capítulo 1

Transcriptome of weeping pinyon pine, *Pinus pinceana*, shows differences across heterogeneous habitats

Laura Figueroa-Corona, Patricia Delgado Valerio, Jill Wegrzyn & Daniel Piñero

Trees structure and function.

Doi: https://doi.org/10.1007/s00468-021-02125-8

Publicado el 19 abril de 2021

El siguiente capítulo está constituido por el artículo de requisito, publicado en la revista *Trees structure and function* el 19 abril de 2021. Los fundamentos de la publicación parten de la hipótesis de que la plasticidad fenotípica incluye rasgos fisiológicos afectados por cambios sutiles en los niveles de expresión de mRNA. Con esta consideración evaluamos la respuesta diferencial dentro de la heterogeneidad climática de *Pinus pinceana* (G. Gordon & Glend). Este pino piñonero presenta una distribución geográfica fragmentada que sugiere una adaptación a las condiciones áridas y extremas de las laderas orientales de las montañas del Desierto Chihuahuense, con variaciones significativas en la temperatura y precipitación anual. Evalué si la divergencia fenotípica está respaldada por diferencias en los perfiles de expresión de árboles adultos a lo largo de su distribución natural y si se correlaciona a los cambios en la sequía y otras presiones ambientales.

Los resultados principales incluyen la reconstrucción de un transcriptoma completo con 45,431 anotaciones de genes codificantes a proteínas del tejido de acículas de siete individuos adultos. Para la comparación entre hábitats se encontró que los árboles de las regiones más secas compartieron respuestas activas relacionadas con respuestas metabólicas correlacionadas a factores abióticos. El análisis de expresión diferencial identificó intermediarios y reguladores de la respuesta al estrés abiótico para 26 genes, destacando las familias ortólogas involucradas en la respuesta ambiental al estrés abiótico, como la serina / treonina-proteína quinasa LRR similar al receptor y L -receptores quinasas de lectina de tipo (LecRK), directamente relacionados con la regulación de la respuesta de defensa a patógenos. Cuando se compararon muestras del Desierto Chihuahuense con muestras externas a la distribución natural, en donde encontramos la expresión diferencial de un regulador de la morfogénesis de la raíz, la proteína quinasa tipo receptor rico en cisteína (CRK28), que podría relacionarse con la absorción diferencial en condiciones de sequía.

Trees
https://doi.org/10.1007/s00468-021-02125-8

ORIGINAL ARTICLE

Transcriptome of weeping pinyon pine, *Pinus pinceana*, shows differences across heterogeneous habitats

Laura Figueroa-Corona[1,2] · Patricia Delgado Valerio[3] · Jill Wegrzyn[4] · Daniel Piñero[2]

Received: 8 September 2020 / Accepted: 30 March 2021

Abstract

Key message **We reconstructed the needle tissue Transcriptome of *P. pinceana* for individuals from distinct biogeographic regions across a temperature and precipitation gradient that represents its natural distribution. Gene expression analysis via RNA-Seq identified differential response to biotic stress.**

Abstract Phenotypic plasticity includes physiological traits affected by subtle changes in mRNA expression levels, causing different molecular phenotypes. We assessed the differential response to climate heterogeneity across the geographic range of *Pinus pinceana* G. Gordon & Glendinning. This pinyon pine shows a fragmented geographic distribution suggesting adaptation to the arid and extreme conditions of the eastern mountain slopes of the Chihuahuan desert. The geographic distribution of *P. pinceana* spans regions with significant variation in annual temperature and precipitation. We tested whether phenotypic divergence is supported by differences in expression profiles in *P. pinceana* mature trees along its natural distribution and corresponds to the changes in drought and other environmental stress. The reconstructed Transcriptome included 45,431 high-quality annotations derived from the needles of seven individuals across contrasting biogeographic and climatic localities. Trees from the driest regions shared active responses related to abiotic factors. The differential expression analysis identified intermediates and regulators of abiotic stress response for 26 genes, highlighting families involved in the environmental response to abiotic stress, and proteins linked to up-regulated responses, such as LRR receptor-like serine/threonine-protein kinase and L-type lectin receptor kinases (LecRK), directly related to pathogens. When Chihuahuan desert samples were compared with arboretum samples, we found differential expression of a regulator of root morphogenesis, cysteine-rich receptor-like protein kinase (CRK28), which could relate to differential absorption in drought conditions.

Keywords *Pinus pinceana* · Transcriptome · Drought conifer response · Weeping pinyon

Introduction

Response to changing environments is frequently controlled through modification of mRNA expression and polymorphic variation, which enable phenotypic plasticity and adaptation (Schlichting 1986; Pfennig et al. 2010). Phenotypic plasticity includes physiological traits influenced by changes in mRNA expression levels and causes differential performance in response to environmental conditions (Ackermann et al. 2013; Schlichting 1986). When this plasticity occurs as a consequence of differential expression of the same genotype, directly or indirectly correlated with environmental factors, it could imply selective pressures (Henriques et al. 2018). Over several generations, this phenotype could become fixed in the population and recognized as a local adaptation.

Population studies have been conducted to understand response to drought through the genotype–phenotype

Communicated by J. Hormaza.

✉ Daniel Piñero
pinero@unam.mx

1 Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

2 Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

3 Facultad de Agrobiología, Universidad Michoacana de San Nicolás de Hidalgo, Uruapan, Mexico

4 Department of Ecology and Evolutionary Biology, University of Connecticut, CT 06269 Storrs, USA

Published online: 19 April 2021

environment relationship. This has been approached via ecophysiological studies (Noctor et al. 2014; Álvarez-Yépiz et al. 2014), provenance trials (Correia et al. 2006), and quantitative trait locus mapping (González-Martínez et al. 2006; Aitken et al. 2008; Neale and Kremer 2011). Transcriptional profiles are also frequently studied to examine response to abiotic stress, such as drought. Transcriptomic approaches are much more accessible than genome-scale studies in conifers due to their large genome sizes and significant complexity (repeat content and pseudogenes) (Birol et al. 2013; Neale et al. 2014; Gonzalez-Ibeas et al. 2016; Perera et al. 2018).

Species of conifers that grow along environmental gradients have developed plastic and adaptive mechanisms to respond to fluctuations in water availability to limit exposure and damage. Some conifers distributed in boreal regions, including *Larix decidua*, lose their needles seasonally to avoid freezing, which can result from the storage of air in the vascular tissue (Roden et al. 2015). Others reduce exposure by allowing the maintenance of osmotic balance through stomata closure. This is exhibited by members of the Cupressaceae and Taxaceae (McDowell et al. 2008; Brodribb et al. 2014). Finally, conifers demonstrate drought resistance and restorative mechanisms such as phytohormonal regulation; in the case of *Pinus pinaster*, this is associated with the accumulation of proline, abscisic acid (ABA), and jasmonic acid in the needles (Corcuera et al. 2012).

Distributed across arid environments of North America, the 13 species of pinyon pines (subsection *Cembroides*) have enlarged seeds with lipid-rich endosperms and the absence of a vestigial wing. Their morphological divergence, such as needle number, needle stiffness, length of cone, and seed size (Malusa 1992; Farjon 1996; Montes et al. 2019), is associated with different gradients of aridity (Ortiz-Medrano et al. 2016). Among the pinyon pines of North America, *P. pinceana* can withstand conditions of extreme aridity. It is an endemic species restricted to rocky soils in the Chihuahuan desert (ChD) and the Sierra Madre Oriental (SMO; 1480–3000 masl) (Farjon et al.1997; González-Elizondo et al. 2012; Martínez-Ávalos et al. 2015) and is listed as endangered in the Mexican Official Law (NOM 059-SEMARNAT-2010; https://www.biodiversidad.gob.mx/especies/catRiesMexico).

Morphological differences in *P. pinceana* have been reported across biogeographic provinces. Martiñón-Martínez et al. (2010) detected significant differences in needle wax cover, which was thicker over needles in juvenile individuals from ChD compared to juveniles from SMO. They also reported that thermal increments produced faster growth, likely due to higher water consumption. Córdoba-Rodríguez et al. (2011) described larger primary roots and greater volume on secondary roots in ChD localities. Both studies described distinct physiological differences in contrasting drought conditions. Martinez-Martiñón et al. (2010) reported that in a common garden experiment controlling temperature and humidity, ChD juveniles had less biomass but assigned more resources to root growth than SMO individuals. Population genetic approaches detected that *P. pinceana* exhibited greater diversity and differentiation across its fragmented geographic distribution than would be explained by isolation due to distance (Escalante 2001; Ledig et al. 2001; Molina-Freaner et al. 2001).

Assuming there are associations between environmental variables and expression profiles, we examined mature *P. pinceana* trees and report on three aspects: (1) characterization of the climatic differences along the natural range of *P. pinceana*, (2) a de novo assembled and annotated needle Transcriptome for the species, and (3) differential expression analysis in relation to environmental conditions.

Materials and methods

Climatic characterization

From 57 registered localities recorded in the National Herbarium of Mexico (MEXU), we retained 22 records verified in the previous literature (Escalante 2001; Molina-Freaner et al. 2001), at least 800 m from each other. To construct a potential niche model using the default parameters, we used Maxent v3.3.3e (Phillips and Dudík 2008). Climatic information for each record was extracted from 19 bioclimatic layers (World Clim; Hijmans et al. 2005). To quantify the annual drought changes within the sampled populations, we used four layers from the Global Aridity Index (Trabucco and Zomer 2009). The latter is constructed with the relationship of precipitation and water content in the soil with data from 1970 to 2000 and a 30-arcsecond resolution.

To avoid redundancy in environmental information, we performed pairwise correlations among all variables. Among the highly correlated pairs ($r^2 > 0.75$), we chose variables that best explained the geographic distribution for *P. pinceana* using a threshold of the minimum training presence of 0.5. A restricted niche model was also built in Maxent using annual mean precipitation, isothermality, temperature changes per season, annual mean temperature, and soil water content in the warmest quarter of the year (Fig. S1; Table S1). The information was extracted at each sampled location using the GARP utility v1.0 (Stockwell 1999) on ArcGIS (ESRI Environmental Systems Research 2011). Subsequently, a principal components analysis (PCA) was conducted with *devtools* and *ggbiplot* packages in R (Wickham 2016) to quantify the distribution's environmental variance.

Transcriptome characterization

Plant material, cDNA library construction, and sequencing

Foliar tissue was collected in the dry season (11–21 April 2018) from seven healthy adult individuals selected at random in seven localities of the natural distribution, located in two different biogeographic provinces (Morrone et al. 2017): localities General Cepeda, Sierra de Parras, Palmito, and Mazapil in ChD and localities Nuñez, Maguey Verde, and Tolantongo in SMO. We also sampled two individuals from the arboretum of the Universidad Autónoma de Chapingo in Texcoco City, Mexico, located outside the altitudinal range and the natural distribution (Fig. 1; Table S1).

RNA was extracted using the Spectrum Sigma® protocol (Sigma Aldrich, St. Louis, MO, USA), which uses β-mercaptoethanol to make the lysate, in accordance with the supplier's recommendations. We used the protocol adapted for conifer needles, or leaves with normal content of water and starch storage organs, to bind the RNA using a column. Wash, dry, and elution steps followed standard practice. The cleaned product was visualized on a 1% agarose gel stained with ethidium bromide. To evaluate the occurrence of salts, we quantified all samples with a concentration greater than 100 ng/ml and a 260:280 ratio between 1.8 and 2 samples were further processed in the Genomics Sequencing Laboratory in the Institute of Quantitative Biosciences (QB3) at the University of California, Berkeley. The RNA integrity was first evaluated for values greater than 7.0 in an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). cDNA libraries were constructed for each individual following Illumina TruSeq RNA standard protocol (Illumina, San Diego, CA, USA). Nine libraries were multiplexed and sequenced on a single run of an Illumina HiSeq 4000 (150-bp paired-end reads).

Transcriptome assembly

The adapters were removed, and quality filtering was performed with TRIMMOMATIC v0.36 (Bolger et al. 2014) to retain sequences which were at least 80 bp in length and had a minimum Phred-scaled quality score of 35. The procedure was visualized and confirmed with FASTQC/MULTIQC (Ewels et al. 2016). A de novo assembly for each locality was executed for each paired-end read from each library. This was carried out with Trinity RNA-Seq v2.0.2

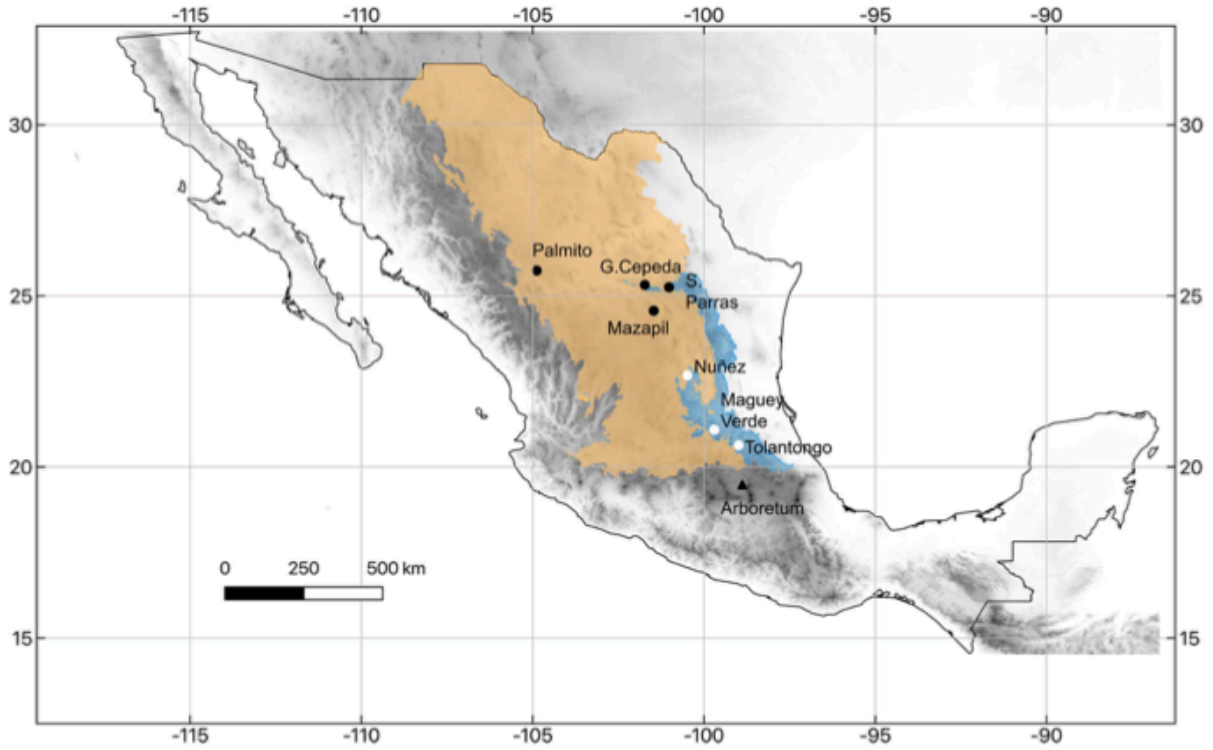

Fig. 1 Localities of *Pinus pinceana* sampled. Black dots correspond to natural distribution, and white dots correspond to pines collected in the arboretum of the Universidad Autónoma de Chapingo, outside its natural distribution. The Chihuahuan desert (ChD) region is shown in orange and Sierra Madre Oriental (SMO) in blue

(Grabherr et al. 2011), requiring paired reads and a minimum contig length of 300 bp, and independently using rnaSPAdes v3.9 CPU following an assembly for paired reads (Bankevich et al. 2012). The assemblies were processed via USEARCH v9.0.2132 to reduce the redundancy of transcripts within each assembly (0.9 similarity index; Edgar 2010). Local clustered assemblies were frame-selected using Transdecoder v3.0.1 (Haas et al. 2013) to identify the longest and most likely open reading frames. To further refine the reference Transcriptome, we built a global reference with the 18 local assemblies provided by Trinity and rnaSPAdes (following clustering and frame-selection) via Evigene-R in EvidentialGene v20190101 (Gilbert 2013). The *trformat* script reduced redundancies on the sequence identifiers, the *tr2aacds* script selected transcripts from the highly similar sequences, and *evgmrna2tsa* formatted the selected sequences.

Ortholog annotation and function

Each reduced local library and the global reference contigs were processed and annotated with EnTAP v 0.9.0-beta (Hart et al. 2020). EnTAP utilizes Diamond v0.9.19 for sequence similarity searches and EggNOG v0.99.1 (Huerta-Cepas et al. 2018) to provide gene family assignments. Functional annotation was carried out against NCBI Refseq Complete-Protein, Refseq Plant-Protein, and Swissprot UniProt using a threshold of 70% similarity target and an e-value of 10^{-6} using DIAMOND (O'Leary et al. 2015; UniProt Consortium 2018). In addition, we selected and separated the matches that corresponded to potential contaminants, such as insects, fungi, and bacteria. The remaining sequences were annotated with Gene Ontology (GO) terms using Blast2GO (Conesa et al. 2005) and InterProScan v5.25 (Jones et al. 2014) for the identification of conserved domains using the Pfam-A v31.0 database. The integrity and quality were reviewed with the single-copy ortholog benchmarking tool BUSCO (v3.0; Waterhouse et al. 2018) and the Embryophyta database (Obd10).

Finally, using the annotated genes from the seven localities as independent proteomes, we composed putative gene families (orthogroups). By considering the classification within orthogroups, we could discriminate a set of genes that descended from a single gene of the considered localities. We did this to minimize the heterogeneity among samples (partial transcripts and isoforms characteristic of de novo assemblies) and maintain the gene families in all localities. Using OrthoFinder v2.4, we compared the translated proteins from the different localities and generated sets of single-copy genes as well as shared families (Emms and Kelly 2015).

Differentially expressed transcripts

The trimmed paired-end reads from each of the nine libraries were mapped to the global reference of *P. pinceana* to calculate transcript abundance via Kallisto v0.44.0 with a pseudo alignment using 100 bootstraps (Bray et al. 2016). To construct and analyze the expression profile, we used the R package for differential expression, DESeq2 v1.26, available in Bioconductor (Love et al. 2014), and all mapped genes were normalized by DESeq2 median ratio (reads/Kb/mapped). Significantly expressed transcripts were defined at a p-adjusted value < 0.1 to report the less differences in samples very heterogeneous among them. Fold change values were reported for samples clustered by region: four libraries from ChD localities and three libraries from SMO localities. We compared samples from the two biogeographic provinces using the localities as biological replicates to understand the transcriptional changes across the natural distribution using the $\log_4$-fold change to obtain a conservative criterion for differences across regions. To control for phenotypic plasticity variation, we conducted a second comparison between each biogeographic province and the two arboretum samples.

Results

Climatic differentiation across the distribution range

The niche-modeled distribution was assigned to zones whose climatic characteristics were at least 50% similar to those of the used localities, defining the specific conditions in which *P. pinceana* grows (Table S2). Variables with the most significant effect on niche modeling were the annual mean temperature, precipitation of the driest quarter, water content index in the warmest quarter of the year, and isothermality (Table S3). PCA showed that the most considerable portion of the variance was explained by PC1 (49.2%) associated with the water content index in the warmest quarter of the year and isothermality. In comparison, PC2 (29.2%) showed the largest loadings of mean annual temperature and precipitation in the driest quarter of the year, demonstrating the environmental heterogeneity between ChD and SMO localities (Fig. 2a). Using locality values, we evaluated the seasonal changes in soil water content. We found that the most arid conditions in ChD occurred during the final quarter of the year, when SMO localities Maguey Verde and Tolantongo had the highest precipitation (Fig. 2b, c).

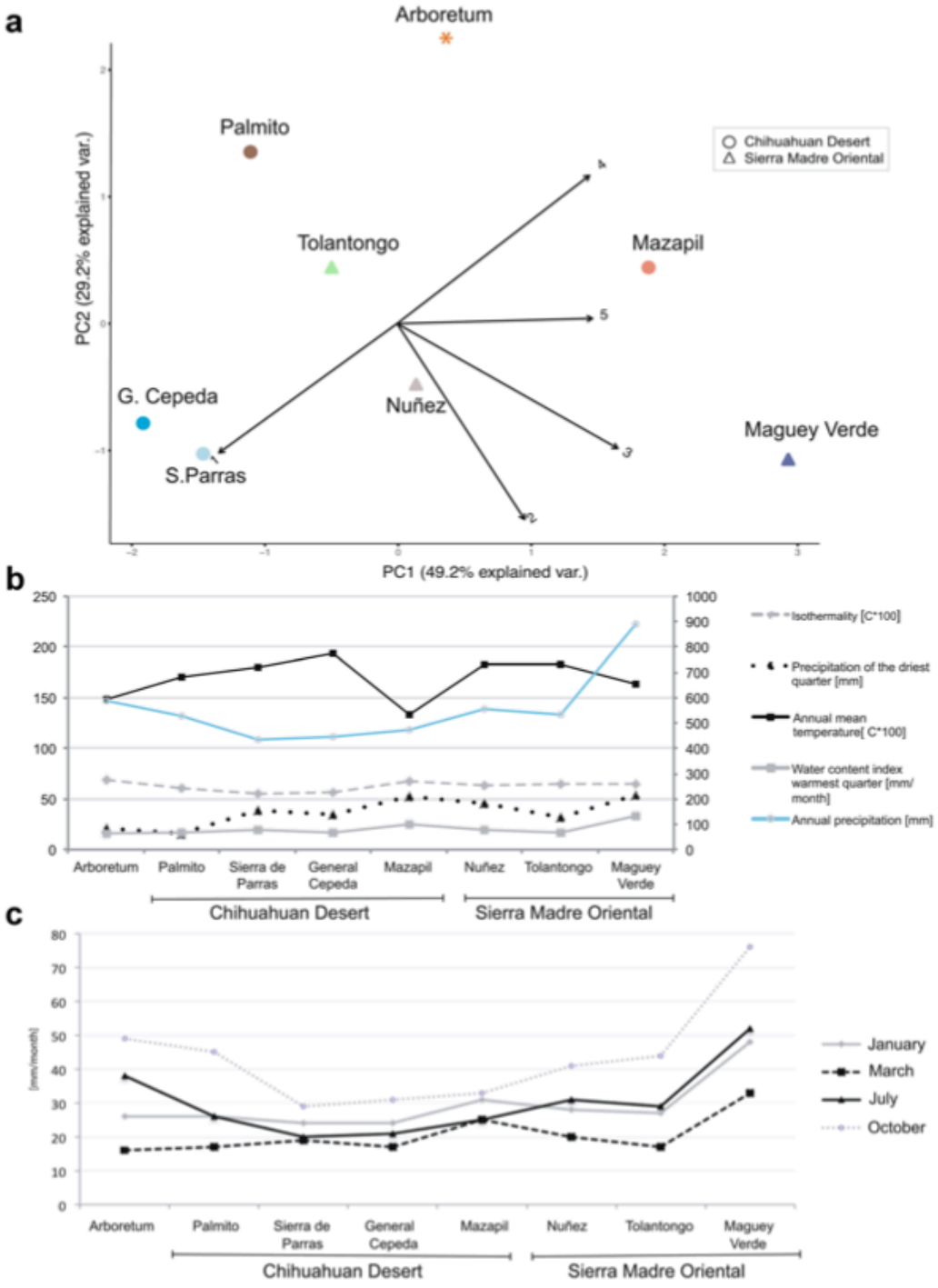

Fig. 2 Climatic characterization of *Pinus pinceana* geographic distribution. **a** Principal components analysis for the sampled localities (triangles show the Sierra Madre Oriental populations; circles, the Chihuahuan desert localities) and the climatic conditions: 1, annual average mean temperature; 2, precipitation of the driest quarter of the year; 3, water content index in the warmest quarter of the year; 4, isothermality; 5, annual precipitation. **b** isothermality, precipitation of the driest quarter of the year, water content index in the warmest quarter, and annual precipitation across localities. **c** Seasonal changes in water content on soil index for each of the studied sampled localities

Transcriptome assembly and annotation

The paired-end RNA sequencing produced 37.3 million reads on average for each library, which was reduced to 30.7 million after quality control. From the de novo assemblies, the average total transcripts for each locality were 109,336 with an average length of 964 bp (N50 1459 bp) and BUSCO completeness of 79% (Table 1). We used USEARCH to reduce the redundancy in each assembly. This produced an average of 90,755 sequences for each library (Table S4). Processing with Transdecoder to determine open reading frames produced an average of 62,219 coding sequences per library.

Through Evigene, we built the global reference of 149,408 sequences from 18 original assemblies. A total of 27,952 sequences aligned to the databases selected in EnTAP. Of these, 16,276 were considered informative (based on the description provided by the target sequence). The other 11,676 were defined as uninformative (with descriptions focused on terms such as hypothetical or unknown). Alignments to sequences of the model moss, *Physcomitrella patens* (461; 1.66%) were prevalent, as well as more basal angiosperms, including *Amborella trichopoda* (2,838; 10.15%) and *Nelumbo nucifera* (631; 2.26%). In addition, other angiosperms, *Arabidopsis thaliana* (1,510; 5.4%), *Cucurbita pepo* (1,156; 4.14%), *Quercus suber* (2,303; 8.24%), and *Elaeis guineensis* (536; 1.9%), were prevalent. This phylogenetic bias can be understood as an artifact consequence of the limited genomic resources available for gymnosperms, where among genera, only 349 sequences were aligned (1.25%). Gene family assignment via EggNOG identified 45,135 transcripts that could be assigned. In addition, a subset of 34,232 sequences was assigned to at least one GO term. In total, 45,431 (54.2%) transcripts were assigned to either an EggNOG gene family or aligned (with confidence) to a sequence in our selected databases.

At this stage, we removed 1753 sequences that aligned to contaminant categories, corresponding to bacteria (25; 1.43%), insects (348; 19.85%), and fungi (1380; 78.76%) in the global reference Transcriptome. The description and abundance of the contaminants in the local and the global assemblies are shown in Table S5. Overall, a rich representation of parasitic yeasts was present, including *Saccharomyces cerevisiae*; *S. pombe*; yeasts associated with wood-ingesting insects, including *Aureobasidium namibiae*, *Candida albicans*, and *Sugiyamaella lignohabitans*; and pathogens of termites, like *Cryptotermes secundus*. The identified contaminants were consistent across all assemblies, including the arboretum samples. We detected an abundant presence of pathogenic fungi in SMO localities (Maguey Verde and Nuñez), including *Diplodia corticola, Bipolaris maydis, B. oryzae, Sclerotinia sclerotiorum, Cercospora beticola*, and the endophytic fungus *Xylona*

Table 1 Sequencing summary statistics sequencing and de novo assembly

Sample	Condition	Total number of reads sequenced (M)	Total number of reads trimmed (M)	Trinity				rnaSPAdes			
				Total transcripts	Transcript N50	Median length	BUSCO completeness	Total transcripts	Transcript N50	Median length	BUSCO completeness
Arboretum		36.9	29.1	85,286	1721	708	82.9%	15,645	2085	473	76.8%
Arboretum		46.7	39.4	117,534	1565	602	87.0%	21,284	1355	591	87.8%
Tolantongo	SMO	26.5	21.7	86,111	1569	607	80.2%	17,023	300	546	80.4%
Maguey Verde	SMO	45.5	40.1	148,791	1396	559	83.1%	23,887	294	570	83.7%
Nuñez	SMO	40.6	33.1	82,559	1156	501	61.5%	13,327	722	502	64.8%
Mazapil	ChD	37.7	30.5	106,542	1500	589	84.6%	19,311	247	582	85.5%
Sierra de Parras	ChD	28.4	23.8	110,243	1416	565	79.5%	19,259	235	567	80.7%
General Cepeda	ChD	41.6	34.2	118,247	1524	598	84.6%	20,632	259	577	84.9%
Palmito	ChD	32.2	25	128,715	1286	541	77.6%	21,037	2763	523	77.8%

heveae. In ChD localities, we detected more abundance and diversity of pathogenic insects, including *Dendroctonus ponderosae* in Palmito and *Cryptotermes secundus* in Mazapil and General Cepeda. We also found across localities ant species such as *Vollenhovia emeryi, Trachymyrmex septentrionalis*, and *T. zeteki*. The natural growth conditions can likely explain their presence as many of the species listed as contaminants are known to be pathogenic, commensal, or mutualistic to *P. pinceana*.

Following removal of the contaminants, the final Transcriptome represented 147,531 sequences. The estimated completeness of the 45,431 annotated genes (a subset of the full reference) was 89.3%, as estimated by the BUSCO Embryophyta database. Sequences retained in each annotation stage are described for the local assemblies in Table 2.

Gene family analysis

The final location-specific transcriptomes revealed high-quality annotations: 29,731 in Tolantongo (SMO), 42,590 in Maguey Verde (SMO), 27,217 in Nuñez (SMO), 32,842 in Mazapil (ChD), 36,279 in General Cepeda (ChD), 33,012 in Sierra de Parras (ChD), and 33,556 in Palmito (ChD; Table 2). Using each locality as an identity unit in an OrthoFinder analysis, 58,757 orthogroups representing 403,614 transcripts were generated (Table S6). The expressed core present in all localities of *P. pinceana* was composed of 19,890 orthogroups representing 223,273 genes (33.8% of the total) (Fig. 3). These orthogroups contained at least 2 and as many as 115 genes. The locality-specific orthogroups ranged from 773 families represented by 1,827 genes (1.31% of the total) identified in Maguey Verde to 240 families represented by 514 genes (0.4%) in Tolantongo. Between biogeographic provinces, more shared gene families were found between Maguey Verde (SMO) and Palmito (ChD; 933) than between Maguey Verde and General Cepeda (ChD; 908; Table S6).

Present exclusively in ChD localities, we detected 1,992 genes representing 402 (0.68%) of the orthogroups. We identified gene families associated with abiotic stress response, like RNA polymerase II promoter in response to salt stress (GO:0036251); cell wall organization; and biogenesis of cell wax, such as MAP kinase kinase activity (GO: 0038068). Restricted to SMO localities, we found 885 function enriched genes grouped in 227 orthogroups (0.38%), including deacetoxycephalosporin-C hydroxylase activity (GO:0045442). Finally, orthogroups GO:0045414, GO:0045430, GO:0045442, GO:0045493, GO:0045522, GO:0045525, and GO:0045526 were related to interleukin production and regulation responsible for plant disease resistance and response to pathogens (Meyers et al. 2002).

Table 2 Sequences processed at sequential annotation stages

	Arboretum 1	Arboretum 2	Tolantongo	Maguey Verde	Nuñez	Mazapil	General Cepeda	Sierra de Parras	Palmito	Global reference
1. Similarity search	24,603	27,017	22,670	30,406	19,211	24,905	27,124	24,417	24,911	27,952
2. EggNOG	32,301	35,335	29,590	42,367	27,094	32,692	36,124	32,852	33,384	45,135
3. Gene ontology	26,746	29,042	24,490	35,158	22,751	26,863	29,751	27,079	27,689	34,232
4. HQ annotations	32,429	35,522	29,731	42,590	27,217	32,842	36,279	33,012	33,556	45,431

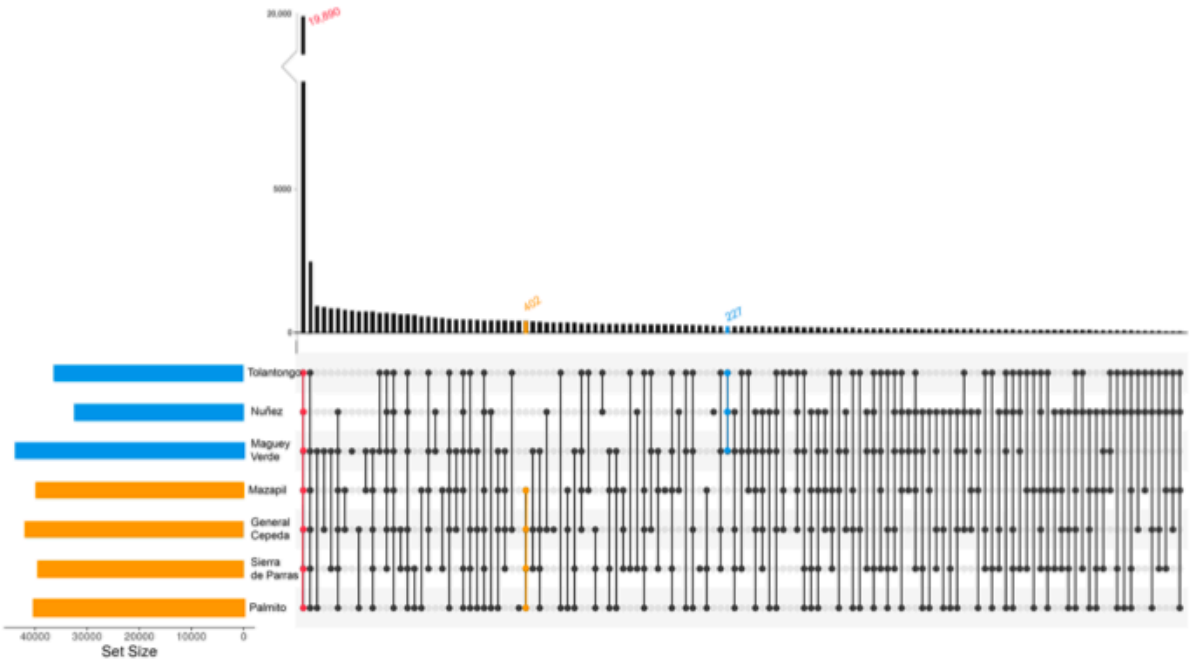

Fig. 3 Orthogroups shared among localities. Tiling beneath the histogram indicates the membership of genes to each orthogroup in the set. Set size is displayed as height on histogram. Horizontal histogram indicates the number of orthogroups found in each locality. Orange indicates data from Chihuahuan desert; blue, Sierra Madre Oriental localities

Differential expression

An average of 58% of the reads, per sample, aligned to the global Transcriptome reference; this shows the high diversity of transcriptional expression across localities (Table S7). The expression profiles were constructed by dividing the localities according to their biogeographic provenance. This was done to compare the climatic differences across a gradient from more humid SMO localities to drier ChD localities.

Using the Transcriptome reference of 147,531 genes, and after normalization, we removed 94,153 transcripts with low read counts and 1,216 outliers. The reference Transcriptome of 45,431 sequences was indexed for differential expression analysis. Three pairwise comparisons were conducted: SMO vs. ChD, arboretum vs. ChD, and arboretum vs. SMO.

After applying a threshold of fourfold change, we found 26 differentially expressed genes, 10 down-regulated and 16 up-regulated (Fig. 4; Fig. S2; Table S8). These genes were primarily associated with abiotic stress response and pathogen interaction. Those expressed in SMO localities included L-type lectin-domain containing receptor kinase IV.2 (LRK42) and membrane-spanning receptor-like kinases. These have putative roles in biotic and abiotic stress response and in plant development. A study in *A. thaliana* by Wang et al. (2014) reported overexpression of LRK42 upon infection of *Pseudomonas syringae* and even suppressed pathogen resistance. In addition, involved in the response to biotic stress, we detected increased expression of finger protein ATL3 (RING-H2) related to disease resistance against tobacco mosaic virus (TMV) and *P. syringae*. This up-regulation has also been related to increased oxidative stress tolerance and drought tolerance in seed germination in rice (Liu et al. 2008). Disease-resistance proteins, such as DRL29, were significantly up-regulated in ChD individuals (samples from Sierra de Parras and General Cepeda) compared to SMO individuals. Finally, the disease-resistance protein At4g27220, with a LRR domain linked to plant immunity, was also up-regulated in ChD individuals.

The comparison between the two arboretum samples and ChD localities identified 44 differentially expressed genes, 17 up-regulated and 27 down-regulated (at least fourfold change) (Fig. S3a). In the arboretum samples, we identified up-regulated genes that are associated with pathogen recognition and the subsequent activation of plant defense mechanisms, including LRR receptor-like serine/threonine-protein kinase (At1g67720), universal stress family protein, and oxalate-CoA ligases, and related to plant defense against fungal pathogens, like phenylcoumaran benzylic ether reductase PT1 (PCBER), an intermediate in lignan and isoflavonoid production responsible for durability, longevity, and resistance of heartwoods in *Pinus taeda* and *Populus trichocarpa* (Gang et al. 1999). A pathogenesis-related

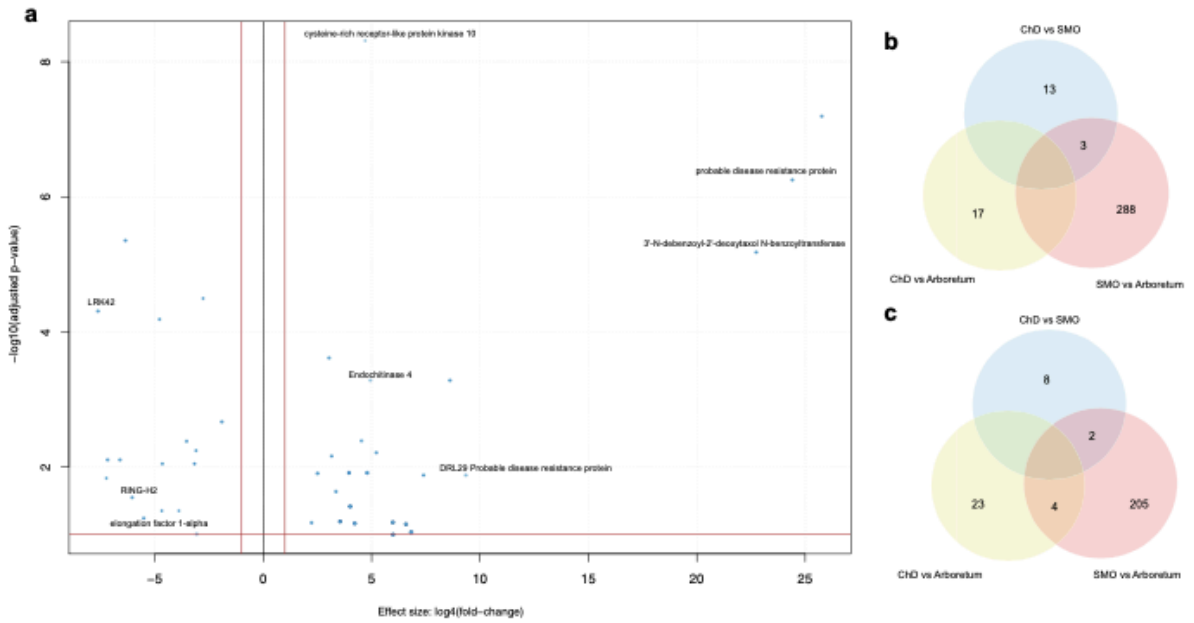

Fig. 4 Comparison of levels of gene expression. **a** Volcano plot with the dispersion at $\log_4$-fold change on the comparison between Sierra Madre Oriental and Chihuahuan desert. **b** Genes up-regulated in different comparisons between localities and the two arboretum individuals. **c** Genes down-regulated across different comparisons between localities and the two arboretum individuals

thaumatin family protein, linked to early antifungal activity in various plant species, was significantly up-regulated in the arboretum samples (Zhang et al. 2018). The pathogenesis-related protein 1 A (PR1) is known to play an important role in stress-related infection of *Fusarium oxysporum* but also is an indicator of tolerance in tomato varieties under drought stress (Akbudak et al. 2020). In contrast, in ChD localities, we identified a threefold increase in the expression of disease-resistance protein (TAO1) compared to our arboretum samples. Gene ontology links probable disease-resistance protein (At4g33300) to *P. syringae* infection resistance in sweet pepper (Cheng et al. 2020). We also found overexpression of shikimate O-hydroxycinnamoyl transferase-like that has been identified as related to heat stress (Sun et al. 2018). Remarkably, we found overexpression of cysteine-rich receptor-like protein kinase (CRK28), regulated by ABA, which has been shown to slow down root growth and lateral root formation in *A. thaliana* (Pelagio-Flores et al. 2020).

The comparison between the two arboretum samples and SMO localities identified 502 differentially expressed genes, 291 up-regulated and 211 down-regulated (fourfold change; Fig. S3b). In SMO localities, we identified genes related to biotic response, including LRR receptor-like serine/threonine-protein kinases (At1g56140, At3g47570) associated with innate immunity in plants. In addition, proteins related to disease resistance (At5g63020) and monooxygenase 2 (MO2) overexpressed in our samples from SMO have been linked to LRR proteins as a response to oomycete infections (Hok et al. 2011). In addition, we found overexpressed in SMO two different TMV-resistance proteins N that, with LRR proteins, have been shown to activate innate immunity regulating infections mediated by pathogen recognition in *Zea mays* and *A. thaliana* (Afzal et al. 2008). Linked to abiotic stress, we detected overexpression in SMO of UDP-glycosyltransferase 72B1 related to monolignol biosynthesis, lignin polymerization, and cell wall-related transcription, factors that produce secondarily thickened cells, making them rigid, which was show by Lin et al. (2016). In addition, we found sodium/hydrogen exchanger 7, which has been linked to improving salt tolerance (Shi et al. 2003). The SMO samples show an overexpression of MYB4 transcription repressor, which encodes the control of flavonoid and shikimate biosynthesis that, under drought conditions, increases hormone signals such as ABA, jasmonate, and brassinosteroids (Wang et al. 2019). SMO samples also showed overexpression of aryl-alcohol dehydrogenase (AAD4), which participates in the degradation of lignin when carbon or nitrogen is limited (Reiser et al. 1994). In addition, we found overexpression of class II heat shock protein-like, which becomes highly abundant during heat stress but is also implicated in response to stomata regulation and response to viral or fungal disease (Lin

et al. 2018). SMO localities show overexpressed transcripts related to abiotic and biotic stress, like peroxidase 12-like, involved in removing H_2O_2, oxidation of toxic reductants, defense against insect attack, symbiosis, and normal cell growth (Cosio and Dunand 2009). Linked to abiotic and biotic response, we detected three different heat shock protein-like and trans-cinnamate 4-monooxygenase involved in lignification and protection from UV-induced DNA damage, predators, and pathogens (Schilmiller et al. 2009). Proteins that show an up-regulation in the arboretum samples when compared to SMO localities were related to secondary metabolite production involved in the defense response, like taxadien-5-alpha-ol O-acetyltransferase (T5AT) involved in taxol and other alkaloid biosynthesis, and dihydroflavonol 4-reductase, linked to the biosynthesis of anthocyanins and condensed tannins involved in reproduction, growth, and protection against pathogens (Petit et al. 2007). In the arboretum samples, we also found overexpression of proteins linked to the immune response to invasive biotrophic pathogens and defense mechanisms, like four different LRR receptor-like serine/threonine-protein kinase (At3g47570, At5g59680, At5g04720, and At1g56140). Furthermore, we found overexpressed proteins that participate in the activity of two LRR-like genes with TMV protein resistance N-like activity, one gene with toll/interleukin-1 receptor-like protein function, three genes with lipase-like (PAD4) activity that induce the biosynthesis of salicylic acid (Jirage et al. 1999), and one gene with rust resistance kinase (Lr10) activity (Feuillet 1997). Mainly related to the response against pathogens, we detected that overexpression of patatin-like protein 2 promotes cell death and affects resistance to *Botrytis cinerea* and *P. syringae* in *Arabidopsis* sp. This overexpression contributes to resistance to cucumber mosaic virus (La Camera et al. 2009). We also found antifungal protein ginkbilobin-like (GNKL) linked to the resistance to infections of *Fusarium oxysporum, Trichoderma reesei*, and *C. albicans*. In addition, we detected nine sequences up-regulated in the SMO localities that are related to both stress (GO:0006950) and defense response (GO:0006952). Five sequences were down-regulated in the SMO samples characterized by orthologous proteins related to response to stress (GO:0006950), defense response (GO:0006952), response to reactive oxygen species (GO:0000302), response to osmotic stress (GO:0006970), or response to oxidative stress (GO:0006979).

In summary, these comparisons show that only two genes are up-regulated in ChD but down-regulated in SMO localities when compared to arboretum samples; one lacks an annotation and the other, 7-deoxyloganetin glucosyltransferase, is involved in the biosynthesis of alkaloids. We were not able to detect up-regulated genes in the arboretum samples compared to the ChD localities. In contrast, for down-regulated genes, ChD and the arboretum samples shared four genes, but only one was annotated, a catechol O-methyltransferase. This gene is likely involved in flavonoid biosynthesis. Finally, two genes showed overexpression in SMO samples when compared with ChD and the arboretum samples, one associated with the response to ABA (GO:0009737; Fig. 4b, c) and the other not yet characterized.

Discussion

Transcriptome characterization

Phenotypic plasticity allows organisms to react to changes in their environment by adjusting their metabolism through the regulation of gene expression (Ackermann et al. 2013). The broad ranges of transcriptional phenotypes in the natural localities of *P. pinceana* in different environments are controlled by complex regulatory networks in response to heterogeneous environmental stimuli (abiotic and biotic).

Precipitation and temperature are highly divergent between ChD and SMO; ChD localities are typically much drier than SMO (Fig. 2; Ruiz-Sanchez et al. 2012). The climatic limits used in niche modeling and the low abundance in present day localities have contributed to *P. pinceana's* geographic restriction. Overall, the climate characterization indicates that the index of water content at ChD localities shows less soil water accumulation during the year (Fig. 2b). Bioclimatic and physiological differences could produce ecological differentiation, reproductive isolation, and genetic divergence, as reported in events of parapatric speciation in other pines, including *P. contorta* (Eckert et al. 2012) and *P. hwangshanensis* (Zhou et al. 2017).

Between 27,217 and 42,590 genes were characterized in each of seven localities sampled for *P. pinceana* (Table 2). These numbers are largely in agreement with previous studies of conifer needle transcriptomes, as well as the estimated number of genes in the conifer genomes annotated to date. For example, 34,389 unique genes were reported in *P. lambertiana* (Gonzalez-Ibeas et al. 2016), 50,172 for *P. taeda* (Perera et al. 2018), and 23,932 in the needle Transcriptome of *P. monticola* (Baker et al. 2018).

Transcript heterogeneity among samples could be explained by variation between regions but is also related to variation in weather and time of day for the collections. Samples collected from the landscape are not subject to the controls of a greenhouse and are likely responding to both abiotic and biotic stimuli that could not be detected (Cho et al. 2018). The sampling design used in this study with only one individual for each locality has the advantage that several localities were sampled in two different biogeographic regions but also has the limitation that we cannot make inferences within a given locality.

The presence of 1753 contaminant sequences can be explained by the nature of the samples during collection and their biological interactions rather than subsequent contamination. We found insect and fungal pathogens in all localities, including the arboretum, suggesting that trees are responding to different levels of biotic stresses. In particular, the gene family term for thigmotropism (GO:0009652) is exclusively expressed in Maguey Verde and corresponds to the invasive growth of hyphae of *C. albicans* and/or *Sugiyamaella lignohabitans* (Davies et al. 1999). A phytopathological analysis in Maguey Verde is necessary since we found a relatively abundant representation of parasitic fungus in this locality. The most well-represented organisms reflect infections of *S. cerevisiae* in different localities. This organism has been reported to produce a delay in growth and even cause plantlets to die (Gognies et al. 2001). We also recognized direct pathogenic infections of wood-ingesting insects and indirect evidence from the yeast associated with them. An important pathogenic species is the mountain pine beetle (*Dendroctonus ponderosae*) since larvae feed on the phloem tissue. This does not appreciably harm the living trees, until pupae eventually mature and adults emerge from, and kill, the host tree (Amman 1982). Infections are rampant across North America, where around 8.5 million ha of forests are perturbed (Hicke et al. 2006). Recognizing the rate of insect infections in Palmito, which has also been impacted by new human settlements in the last 15 years, it will be important to evaluate and mitigate the potential damage.

The characterization of differential water absorption performance or morphological differentiation has been described in juvenile ChD individuals (Martinez-Martiñón et al. 2010; Córdoba-Rodríguez et al. 2011). This study identified candidate genes (Table S8) that can be further characterized in larger populations during both dry and rainy seasons. This can expand our understanding of local adaptation in particular localities in the same biogeographical provinces. The regional similarities that we found describe a plastic variation that corresponds to climatic heterogeneity (Fig. 2).

Climatic differences related to changes in gene families

The 19,890 core gene families shared among sampled regions (Fig. 3) are likely related to constitutive metabolic components, such as photosynthesis elements, transcription factors, and regulatory factors of secondary metabolites. The differential abundance of exclusive orthogroups corresponds directly to the number of genes specific to different localities. The drought response-related gene family (GO:0009819, drought recovery) is present in all localities, with more representation in General Cepeda. On the other hand, genes associated with water deprivation (GO:0042631) are present in all ChD localities (Palmito, General Cepeda, and Mazapil). The primary differences between regions included sequences related to response to stress such as herbivory. This produces short-term physiological changes and decreases nutrient assimilation, modifying the photosynthetic machinery. We found 227 orthogroups with functions associated with disease resistance and pathogen response restricted to SMO localities. In particular, we noted differential expression of genes supporting the regulation of interleukins (Bayless et al. 2020; Bréda et al. 2006).

Differential expression in different regions reveals variations in stress regulators

Expression differences included genes related to biotic stress response. Several of these are involved in innate immunity through LRR proteins that are known to mediate pathogen recognition (Meyers et al. 2002; Eitias et al. 2008; Afzal et al. 2008). In addition, we identified pathogen response genes such as L-type lectin receptor kinases (LecRK), overexpressed in SMO in comparison to ChD localities, and disease-resistance protein (TAO1), overexpressed in ChD localities compared to arboretum samples. These proteins, detected in abundance in Nuñez, have been reported by Wang et al. (2014) and Eitas et al. (2008) as responsible for resistance to *Pseudomonas* infections of wood-ingesting nematodes and insects. We also identified patatin protein 2 (PLP2) linked to *Botrytis cinerea* and the antifungal protein ginkbilobin-like (GNKL), reported in *C. albicans* infections (La Camera et al. 2009; Gao et al. 2016).

We used samples from different localities in two biogeographic regions as biological replicates assuming environmental homogeneity across regions. In spite of this, neither climate (Fig. 2) nor gene expression (Fig. S4) showed a clear pattern among localities within each region. Likely, environmental conditions are more important for *P. pinceana* individuals than are the abiotic conditions measured here. Nonetheless, we found intermediates in the synthesis and regulation of salicylic acid, jasmonic acid, brassinosteroid-like *pr-1* genes, and lipase-like protein (PAD4) which are implicated in abiotic stress responses, such as freezing, salinity, and osmotic stress (Akbudak et al. 2020; Jirage et al. 1999).

Furthermore, the occurrence of genes that regulate oxidative stress tolerance and increase drought tolerance was more evident in the independent comparisons between arboretum samples and the natural localities. These include MYB4 transcription factors and several heat shock proteins. Both are important for heat stress, leaf water balance, and response to viral or fungal pathogens (Wang et al. 2019).

In the comparisons between the arboretum and ChD localities, the overexpression in ChD samples includes

cysteine-rich receptor-like protein kinase (CRK28), previously characterized as a regulator of abiotic stress response in relation to ABA signaling mechanisms (Lopez-Molina et al. 2001). Pelagio-Flores et al. (2020) have related it to the development of roots in drought conditions in *Arabidopsis*. This could contribute to the phenotypic differences described in *P. pinceana* by Córdoba-Rodríguez et al. (2011). The absence of this expression pattern in the comparison between the SMO remains unclear. The overlaps between arboretum and ChD, and between SMO and arboretum, are largely related to secondary metabolite production, a key aspect of conifer plant defense (Shalev et al. 2018; Tang et al. 2005).

Conclusions

This study represents a comparative analysis of a de novo Transcriptome assembly in a pinyon pine that grows in contrasting climate localities and in a geographically fragmented distribution. We characterized transcripts across their natural distribution and examined the response of metabolic and resistance pathways impacted by abiotic stress. The differences in transcript abundance in the sampled localities were consistent with the aridity gradient. Determining whether differences among the regions are due to local adaptation or phenotypic plasticity based on RNA sequencing expression profiles will rely on future studies that include more tissues, regional replicates, and seasonal time points.

Author contribution statement All the authors contributed to the study conception and design. Material preparation, data collection, and analyses were performed by LFC. The first draft of the manuscript was written by LFC, JW, and DP. All the authors commented on previous versions of the manuscript and read and approved the final manuscript.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00468-021-02125-8.

Acknowledgements We thank L. Eguiarte and Jessica Gunoskey for valuable suggestions on data analysis and comments on earlier versions of the manuscript; the staff of the Herbarium of the UACh, especially Dr. E. Guizar Nolazco for permission to collect samples in the Pinetum Maximino Martínez; C. Bautista, E. García-Morales, D. Gernandt, S. Gonzalez, A. Moreno-Letelier, R. Salas-Lizana, A. Vázquez-Lobo, and A. Villarruel for fieldwork assistance; A. Gamboa de Buen, M. Garciadiego, M. López, and S. Shishkova for the RNA extraction training; P. Peláez, A. Trouern-Trend, and S. Zaman for bioinformatic assistance; and the Institute for Systems Genomics, Computational Biology Core, UCONN for access to both software and hardware. This work constitutes a partial fulfillment of the requirements for the Ph.D. degree at the Posgrado de Ciencias Biológicas at the Universidad Nacional Autónoma de México (UNAM) for LFC.

Funding Fieldwork was supported by PAPIIT-DGAPA UNAM grant IN209816 to David Gernandt. Lab processing and sequencing was financed by Instituto de Ecología (DP) and CONACYT financial scholarship number 449615 to LFC.

Availability of data and material The RNA reads and the de novo Transcriptome assemblies are available in NCBI via Bioproject PRJNA63711. The code used is available at https://gitlab.com/lcorona/transcriptome_of_weeping_pinyon_pine.git.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Ackermann M, Sikora-Wohlfeld W, Beyer A (2013) Impact of natural genetic variation on gene expression dynamics. PLOS Genetics 9(6):1003514

Afzal AJ, Wood AJ, Lightfoot DA (2008) Plant receptor-like serine threonine kinases: roles in signaling and plant defense. Mol Plant Microbe Interact 21(5):507–517

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol Appl 1:95–111

Akbudak MA, Yildiz S, Filiz E (2020) Pathogenesis related protein-1 (PR-1) genes in tomato (*Solanum lycopersicum* L.) bioinformatics analyses and expression profiles in response to drought stress. Genomics 112(6):4089–4099

Álvarez-Yépiz JC, Cueva A, Dov M, Teece M, Yepez EA (2014) Ontogenetic resource-use strategies in a rare long-lived cycad along environmental gradients. Conserv Physiol 2:1–12

Amman GD (1982) Characteristics of mountain pine beetles reared in four pine hosts. Environ Entomol 11 3(1):590–593

Baker EA, Wegrzyn JL, Sezen UU et al (2018) Comparative transcriptomics among four white pine species. G3 8(5):1461–1474

Bayless AM, Nishimura MT (2020) Enzymatic functions for *Toll/interleukin-1* receptor domain proteins in the plant immune system. Front Genet 11:539

Birol I, Raymond A, Jackman SD et al (2013) Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. Bioinformatics 29(12):1492–1497

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2212

Bankevich A, Nurk S, Antipov D, Gurevich A et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–527

Bréda N, Huc R, Granier A, Dreyer E (2006) Temperate forest trees and stands under severe drought: a review of ecophysiological responses, adaptation processes and long-term consequences. Ann For Sci 63(6):625–644

Brodribb TJ, McAdam SA, Jordan GJ, Martins S (2014) Conifer species adapt to low-rainfall climates by following one of two divergent pathways. PNAS 111(40):14489–14493

Cheng Y, Jiao Y, Miao R, Tian R, Liang Y, Qiao N (2020) Exploring differentially expressed genes of microspore embryogenesis under heat stress in sweet pepper. Afr J Biotech 19(9):661–674

Cho SM, Lee H, Jo H, Kang Y, Park H, Lee J (2018) Comparative transcriptome analysis of field- and chamber-grown samples of *Colobanthus quitensis* (Kunth) Bartl, an Antarctic flowering plant. Sci Rep 8(1):11049

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676

Corcuera L, Gil-Pelegrin E, Notivol E (2012) Aridity promotes differences in proline and phytohormone levels in *Pinus pinaster* populations from contrasting environments. Trees 26:799–808

Córdoba-Rodríguez D, Vargas-Hernández JJ, López-Upton J, Muñoz-Orozco A (2011) Crecimiento de la raíz en plantas jóvenes de *Pinus pinceana* Gordon en respuesta a la humedad del suelo. Agrociencia 45:493–506

Correia MJ, Osório ML, Osório J, Barrote I, Martins M, David MM (2006) Influence of transient shade periods on the effects of drought on photosynthesis, carbohydrate accumulation and lipid peroxidation in sunflower leaves. Environ Exp Bot 58(1):75–84

Cosio C, Dunand C (2009) Specific functions of individual class III peroxidase genes. J Exp Bot 60(2):391–408

Davies JM, Stacey AJ, Gilligan CA (1999) *Candida albicans* hyphal invasion: thigmotropism or chemotropism? FEMS Microbiol Lett 171(2):245–249

Eckert AJ, Shahi H, Datwyler SL, Neale DB (2012) Spatially variable natural selection and the divergence between parapatric subspecies of lodgepole pine (*Pinus contorta*, Pinaceae). Am J Bot 99:1323–1334

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19):2460–2461

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16(1):157

ESRI Environmental Systems Research (2011) ArcGIS Desktop: Release 10. Redlands, CA

Eitas TK, Zachary L, Nimchuk J, Dangl L (2008) *Arabidopsis* TAO1 is a *TIR-NB-LRR* protein that contributes to disease resistance induced by the *Pseudomonas syringae* effector *AvrB*. PNAS 105(17):6475–6480

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32(19):3047–3048

Farjon A (1996) Biodiversity of *Pinus* (Pinaceae) in Mexico: Speciation and Palaeo-Endemism. Bot J Linn Soc 121(4):365–384

Farjon A, Perez de la Rosa JA, Styles BT (1997) A Field Guide to the Pines of Mexico and Central America. Royal Botanic Gardens. Kew Richmon

Feuillet C, Schachermayr G, Keller B (1997) Molecular cloning of a new receptor-like kinase gene encoded at the Lr10 disease resistance locus of wheat. Plant J 11(1):45–52

Gao N, Wadhwani P, Mühlhäuser P, Liu Q, Riemann M, Ulrich AS, Nick P (2016) An antifungal protein from *Ginkgo biloba* binds actin and can trigger cell death. Protoplasma 253(4):1159–1174

Gang DR, Kasahara H, Xia ZQ, Vander Mijnsbrugge K, Bauw G, Boerjan W, Van Montagu M, Davin LB, Lewis NG (1999) Evolution of plant defense mechanisms Relationships of *phenylcoumaran benzylic ether reductases* to *pinoresinol-lariciresinol* and *isoflavone reductases*. J Biol Chem 274(11):7516–7527

Gilbert D (2013) Gene-omes built from mRNA seq not genome DNA. 7th annual arthropod genomics symposium. Notre Dame

Gognies S, Belarbi A, Barka E (2001) *Saccharomyces cerevisiae*, a potential pathogen towards grapevine, *Vitis vinifera*. FEMS Microbiol Ecol 37(2):143–150

González-Elizondo MS, González-Elizondo M, Tena-Flores JA, Ruacho-González L, López-Enríquez LI (2012) Vegetación de La Sierra Madre Occidental, México: Una Síntesis. Acta Botánica Mexicana 100:351–403

Gonzalez-Ibeas D, Martinez-Garcia PJ, Famula RA, Delfino A, Stevens KA, Loopstra CA, Langley CH, Neale BD, Wegrzyn JL (2016) Assessing the gene content of the megagenome: sugar pine (*Pinus lambertiana*). G3 6(12):3787–3802

Gonzalez-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag SNPs at candidate genes for drought-stress response in *Pinus taeda* L. Genetics 172:1915–1926

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29(7):644–652

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M (2013) De novo transcript sequence reconstruction from RNA-Seq using the trinity platform for reference generation and analysis. Nat Protoc 8(8):1494–1512

Hart AJ, Ginzburg S, Xu M, Fisher CR, Rahmatpour N, Mitton JB, Paul R, Wegrzyn JL (2020) EnTAP: bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Mol Ecol Res 20:591–604

Henriques D, Wallberg A, Chávez-Galarza J et al (2018) Whole genome SNP-associated signatures of local adaptation in honeybees of the Iberian Peninsula. Sci Rep 8:11145

Hicke JA, Logan JA, Powell J, Ojima DS (2006) Changing temperatures influence suitability for modeled mountain pine beetle (*Dendroctonus ponderosae*) out breaks in the western United States. J Geophys Res 111:G02019

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25(15):1965–1978

Hok S, Danchin EG, Allasia V, Panabières F, Attard A, Keller H (2011) An *Arabidopsis* (*malectin like*) leucine rich repeat receptor like kinase contributes to downy mildew disease. Plant Cell Environ 34:1944–1957

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P (2018) EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47(D1):D309–D314

Jirage D, Tootle TL, Reuber TL, Frost LN, Feys BJ, Parker JE, Ausubel FM, Glazebrook J (1999) *Arabidopsis thaliana* PAD4 encodes a lipase-like gene that is important for salicylic acid signaling. PNAS 96(23):13583–13588

Jones P, Binns D, Chang HY et al (2014) InterProScan 5: Genome-Scale Protein Function Classification. Bioinformatics 30(9):1236–1240

La Camera S, Balagué C, Göbel C, Geoffroy P, Legrand M, Feussner I, Roby D, Heitz T (2009) The *Arabidopsis patatin-like protein 2* (PLP2) plays an essential role in cell death execution and differentially affects biosynthesis of oxylipins and resistance to pathogens. Mol Plant Microbe Interact 22(4):469–481

Ledig FT, Capo-Arteaga MA, Hodgskiss PD, Sbay H, Flores-Lopez CM. Conkle T, Bermejo-Velazquez B (2001) Genetic Diversity and the Mating System of a Rare Mexican Pinon, *Pinus pinceana*, and a Comparison with *Pinus maximartinezii* (Pinaceae). Am J Botany 88(11):1977–1987

Lin JS, Huang XX, Li Q, Cao Y, Bao Y, Meng XF, Li YJ, Fu C, Hou BK (2016) UDP *glycosyltransferase 72B1* catalyzes the glucose conjugation of monolignols and is essential for the normal cell wall lignification in *Arabidopsis thaliana*. Plant J 88:26–42

Lin JS, Kuo CCh, Yang IC et al (2018) MicroRNA160 Modulates Plant Development and Heat Shock Protein Gene Expression to Mediate Heat Tolerance in *Arabidopsis*. Front Plant Sci 9:68

Liu H, Zhang H, Yang Y et al (2008) Functional analysis reveals pleiotropic effects of rice *RING-H2* finger protein gene *OsBIRF1* on regulation of growth and defense responses against abiotic and biotic stresses. Plant Mol Biol 68:17–30

Lopez-Molina L, Mongrand S, Chua NH (2001) A postgermination developmental arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in Arabidopsis. PNAS 98(8):4782–7

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol 15(12):550

Malusa J (1992) Phylogeny and biogeography of the Pinyon Pines (Pinus Subsect. Cembroides) . Syst Bot 17(1):42–66

Martínez-Ávalos JG, Sánchez-Castillo C, Martínez-Gallegos R, Sánchez-Ramos G, Trejo I, Mora-Olivo A, Alanis E (2015) Primer Registro de *Pinus pinceana* (Pinaceae) Para Tamaulipas: Aspectos Ecológicos y Estructurales. Bot Sci 93(2):357–360

Martiñón-Martínez RJ, Vargas-Hernández J, López-Upton J, Gómez-Guerrero A, Vaquera-Huerta H (2010) Respuesta de *Pinus pinceana* Gordon a estrés por sequía y altas temperaturas. Rev fitotec Mex 33(3):239–248

McDowell N, Pockman WT, Allen CD, Breshears DD, Cobb N, Kolb T, Plaut J, Sperry J, West A, Williams DG, Yepez EA (2008) Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? New Phytol 178:719–739

Meyers BC, Morgante M, Michelmore RW (2002) TIR-X and TIR-NBS Proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. Plant J 32(1):77–92

Molina-Freaner F, Delgado P, Piñero D, Perez-Nasser N, Alvarez-Buylla E (2001) Do rare pines need different conservation strategies? Evidence from three Mexican species. Can J Bot 79(2):131–138

Montes JR, Peláez P, Willyard A, Moreno-Letelier A, Piñero D, Gernandt DS (2019) Phylogenetics of *Pinus* subsection *Cembroides* Engelm. (Pinaceae) inferred from low-copy nuclear gene sequences. Syst Bot 44(3):501–518

Morrone JJ, Escalante T, Rodríguez-Tapia G (2017) Mexican biogeographic provinces: Map and shapefiles. Zootaxa 4277(2):277–279

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. Nat Rev Genet 12:111–122

Neale DB, Wegrzyn JL, Stevens KA et al (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol 15(3):R59

Noctor G, Mhamdi A, Foyer CH (2014) The roles of reactive oxygen metabolism in drought: not so cut and dried. Plant Physiol 164(4):1636–1648

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D et al (2015) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44(D1):D733–D745

Ortiz-Medrano A, Scantlebury DP, Vázquez-Lobo A, Mastretta-Yanes A, Piñero D (2016) Morphological and niche divergence of pinyon pines. Ecol Evol 6:2886–2896

Pelagio-Flores R, Muñoz-Parra E, Barrera-Ortiz S et al (2020) The cysteine-rich receptor-like protein kinase CRK28 modulates *Arabidopsis* growth and development and influences abscisic acid responses. Planta 251:2

Perera D, Magbanua ZV, Thummasuwan S, Mukherjee D, Arick M, Chouvarine P, Nairn CJ, Schmutz J, Grimwood J, Dean J, Peterson DG (2018) Exploring the loblolly pine (*Pinus taeda* L.) genome by BAC sequencing and Cot analysis. Gene 663:165–177

Petit P, Granier T, d'Estaintot BL, Manigand C, Bathany K, Schmitter JM, Lauvergeat V, Hamdi S, Gallois B (2007) Crystal structure of grape *Dihydroflavonol 4-Reductase*, a key enzyme in flavonoid biosynthesis. J Mol Biol 368:1345–1357

Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek PA (2010) Phenotypic plasticity's impacts on diversification and speciation. Trends Ecol Evol 25(8):459–467

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31(2):161–175

Reiser J, Muheim A, Hardegger M, Frank G, Fiechter A (1994) *Aryl-alcohol dehydrogenase* from the white-rot fungus *Phanerochaete chrysosporium*. Gene cloning, sequence analysis, expression, and purification of the recombinant enzyme. J Biol Chem 269(45):28152–28159

Roden J, Kahmen A, Buchmann N, Siegwolf R (2015) The enigma of effective path length for 18O enrichment in leaf water of conifers. Plant Cell Environ 38(12):2551–2565

Ruiz-Sanchez E, Rodriguez-Gomez F, Sosa V (2012) Refugia and geographic barriers of populations of the desert poppy, *Hunnemannia fumariifolia* (Papaveraceae). Org Divers Evol 12(2):133–143

Schlichting CD (1986) The evolution of phenotypic plasticity in plants. Ann Rev Ecol Syst 17:667–693

Shalev TJ, Yuen MMS, Gesell A, Yuen A, Russell JH, Bohlmann J (2018) An annotated transcriptome of highly inbred *Thuja plicata* (Cupressaceae) and its utility for gene discovery of terpenoid biosynthesis and conifer defense. Tree Genet Genomes 14(3):35

Shi H, Lee BH, Wu SJ, Zhu JK (2003) Over expression of a plasma membrane Na^+/H^+ antiporter gene improves salt tolerance in *Arabidopsis thaliana*. Nat Biotech 21(1):81–85

Schilmiller AL, Stout J, Weng JK, Humphreys J, Ruegger MO, Chapple C (2009) Mutations in the *cinnamate 4 hydroxylase* gene impact metabolism, growth and development in *Arabidopsis*. Plant J 60:771–782

Stockwell D (1999) The GARP modelling system: problems and solutions to automated spatial prediction. Int J G I Sc 13(2):143–158

Sun CH, Yang CY, Jason TC (2018) Molecular identification and characterization of hydroxycinnamoyl transferase in tea plants (*Camellia sinensis* L.). Int J Mol Sci 19(12):3938

Tang W, Peng X, Newton RJ (2005) Enhanced tolerance to salt stress in transgenic loblolly pine simultaneously expressing two genes encoding mannitol-1-phosphate dehydrogenase and glucitol-6-phosphate dehydrogenase. Plant Physiol Biochem 43(2):139–146

Trabucco A, Zomer RJ (2009) Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database. CGIAR Consortium for Spatial Information. Published Online. Accessed on 24 Mar 2020

UniProt Consortium (2018) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47(D1):D506–D515

Wang Y, Bouwmeester K, Beseh P, Shan W, Govers F (2014) Phenotypic analyses of *Arabidopsis* T-DNA insertion lines and expression profiling reveal that multiple *L-type lectin receptor kinases* are involved in plant immunity. Mol Plant Microbe Interact 27(12):1390–1402

Wang J, Hu M, Wang J, Qi J, Han Z, Wang G, Qi Y, Wang HW, Zhou JM, Chai J (2019) Reconstitution and structure of a plant NLR resistosome conferring immunity. Science 364:(6435)

Wang XC, Wu J, Guan ML, Zhao CH, Geng P, Zhao Q (2020) *Arabidopsis* MYB4 plays dual roles in flavonoid biosynthesis. Plant J 101:637–652

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 35(3):543–548

Wickham H (2016) Ggplot2: elegant graphics for data analysis. Springer, New York

Zhang J, Wang F, Liang F et al (2018) Functional analysis of a pathogenesis-related thaumatin-like protein gene *TaLr35PR5* from wheat induced by leaf rust fungus. BMC Plant Biol 18:76

Zhou Y, Duvaux L, Ren G, Zhang L, Savolainen O, Liu J (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(3):211–220

Supplementary materials

Table S1. Geography and biogeography of sampled *P. pinceana* locations.

Location	Condition	State	Latitude (N)	Longitude (W)	Elevation (masl)
Arboretum		Edo. México	19.493	98.848	2318
Tolantongo	SMO	Hidalgo	20.635	98.990	1900
Maguey Verde	SMO	Querétaro	21.090	99.695	2300
Nuñez	SMO	San Luis Potosí	22.671	100.483	1539
Mazapil	ChD	Zacatecas	24.599	101.472	2823
Sierra de Parras	ChD	Coahuila	25.255	101.020	2450
General Cepeda	ChD	Coahuila	25.301	102.607	1789
Palmito	ChD	Durango	25.741	104.881	2000

Table S2. Niche modeling predictions

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate
1.000	0.088	Fixed cumulative value 1	0.807	0.000
5.000	0.192	Fixed cumulative value 5	0.664	0.000
10.000	0.256	Fixed cumulative value 10	0.560	0.000
48.463	0.501	Minimum training presence	0.179	0.000
48.463	0.501	10 percentile training presence	0.179	0.000
53.664	0.526	Equal training sensitivity and specificity	0.149	0.111
48.463	0.501	Maximum training sensitivity plus specificity	0.179	0.000
5.194	0.195	Balance training omission, predicted area and threshold value	0.659	0.000
7.507	0.230	Equate entropy of thresholded and original distributions	0.607	0.000

Table S3. Climatic characteristics by location

Location	Isothermality [°C]	Annual Precipitation [mm]	Precipitation of the driest quarter of the year [mm]	Annual mean temperature [°C]	Water content index in the warmest quarter of the year [mm/month]
Arboretum	0.69	589	21	14.9	17
Maguey Verde	0.65	891	54	16.3	33
Tolantongo	0.65	532	32	18.3	17
Nuñez	0.64	557	46	18.3	20
Mazapil	0.68	471	52	13.3	25
Sierra de Parras	0.55	432	39	18.0	19
General Cepeda	0.56	446	34	19.4	17
Palmito	0.61	526	16	17.0	17

Table S4. Quality statistics and sequences processed

	USEARCH	BUSCO completeness	Frame selected Transdecoder	BUSCO completeness	Genes annotated	BUSCO completeness
Arboretum	70446	81.3%	56340	80.8%	32429	80.8%
Arboretum	96923	85.4%	66398	85%	35522	85%
Tolantongo	71792	79.1%	53449	78.8%	29731	78.8%
Maguey Verde	123042	80.5%	81495	80.1%	42590	80.1%
Nuñez	69189	60.4%	47864	60%	27217	60%
Mazapil	88813	82.9%	60729	82.4%	32842	82.4%
Sierra de Parras	92212	78.4%	61408	77.6%	33012	77.6%
General Cepeda	98173	82.9%	67174	82.4%	36279	82.4%
Palmito	106205	77.3%	65117	75.8%	33556	75.8%

Table S5. Most abundant contaminants

Specie	Tolantongo	Maguey Verde	Nuñez	Mazapil	General Cepeda	Sierra de Parras	Palmito	Global reference
Acyrthosiphon pisum				24				
Alternaria alternata:			8					
Anoplophora glabripenni							4	
Arthrobotrys oligospora			3					
Aureobasidium namibiae	14	172	25		17	38	6	334
Aquifex aeolicus								
Bacillus subtilis	46	65	33	55	57	49	51	50
Baudoinia panamericana		73			8	9		145
Bipolaris maydis		46	4					61
Bipolaris oryzae		39	6					89
Botrytis cinerea		24				11		
Candida albicans	5	3	4	5				
Cercospora beticola		38						
Clavispora lusitaniae	9	29	13	5	19	13		
Coniosporium apollinis		140						163
Cryptotermes secundus	6			7	7			
Cyphomyrmex costatus						7		
Dendroctonus ponderosae							4	
Diplodia corticola		47						68
Drosophila melanogaster	9			21		12	8	
Emericella nidulans	12	63	8	14	14	20	10	101
Escherichia coli (strain k12):	9		6	10	14	12	12	
Gibberella zeae		21						
Glarea lozoyensis		73			13	40		143
Halobacterium salinarum					12	9		
Halyomorpha halys							4	
Kluyveromyces marxianus	3							

Metarhizium robertsii				8	6			
Mycobacterium tuberculosis	8			11			9	
Myzus persicae				24				
Neosartorya fumigata		30						50
Neurospora crassa	9	155	22		22	50	8	335
Nostoc sp.			6		8			
Onthophagus taurus	10	24	4	9	15	12	16	
Papilio polytes							5	
Phaeosphaeria nodorum		17						
Phialocephala scopiformis						15		
Protochlamydia amoebophila	9				8			
Pseudocercospora fijiensis						7		
Pseudomonas aeruginosa			7					
Rasamsonia emersonii		23						
Saccharomyces cerevisiae	45	201	33	38	53	61	43	291
Schizosaccharomyces pombe	80	331	70	89	105	113	90	418
Sclerotinia sclerotiorum		39			7	30		89
Sugiyamaella lignohabitans	12	25	11	12	15	18	16	
Synechococcus elongatus				8			8	
Synechocystis sp.	39	38	28	37	33	34	37	
Trachymyrmex septentrionalis	7	6			7		8	
Trachymyrmex zeteki	2						4	
Vollenhovia emeryi	7	16	13	12		20	29	60
Xanthobacter autotrophicus			6					
Xylona heveae		44						58
Yarrowia lipolytica		20				8		
Zootermopsis nevadensis			4	6				
Overall	126	1012	153	157	225	315	178	1,753

Table S7. Statistics number of reads per sample and aligned to the reference transcriptoma for each sample

	TOTAL READS	# R1 Reads	# R2 Reads	MAPPED READS	%
Arboretum 1	58,224,292	29,112,146	29,112,146	50.4%	
Arboretum 2	78,869,060	39,434,530	39,434,530	62.36%	
Nuñez	66,220,808	33,110,404	33,110,404	51.16%	
Mazapil	60,991,584	30,495,792	30,495,792	61.42%	
Maguey Verde	80,176,240	40,088,120	40,088,120	60.6%	
Sierra de Parras	47,661,070	23,830,535	23,830,535	59.95%	
General Cepeda	68,413,772	34,206,886	34,206,886	59.44%	
Tolantongo	43,366,814	21,683,407	21,683,407	59.98%	
Palmito	50,070,730	25,035,365	25,035,365	56.67%	

To consult Table S6 and S8 please see: https://doi.org/10.1007/s00468-021-02125-8.

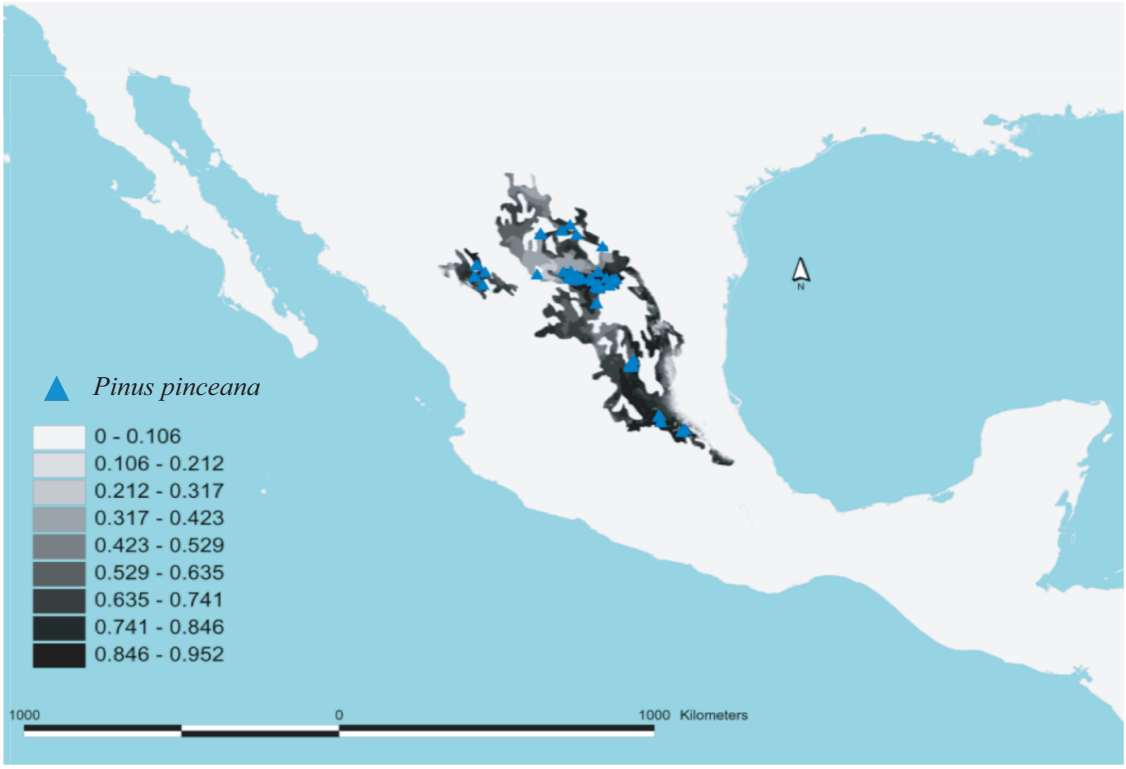

Figure S1. Potential distribution based on climatic conditions. In blue are shown the registered samples from the National Herbarium (MEXU) Black and greys show the probability of *P. pinceana* presence with respect to climatic conditions

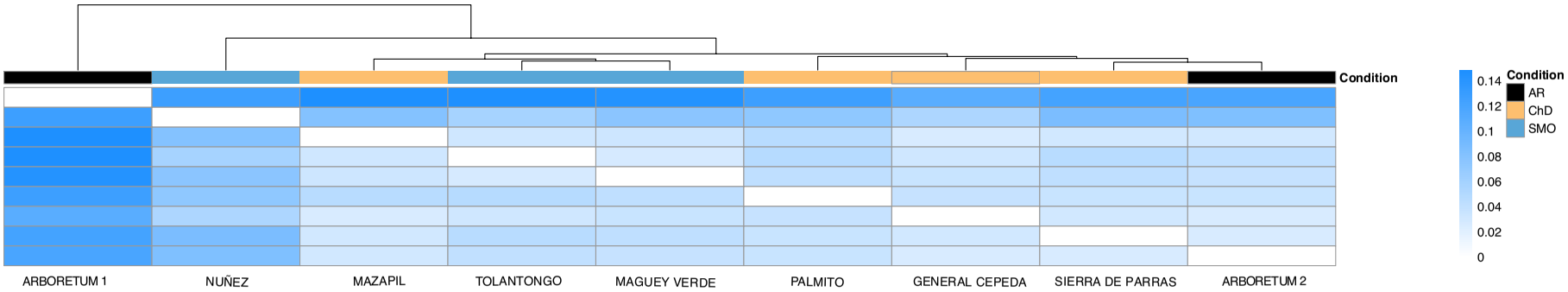

Figura S2. Sample heatmap with the comparison among levels of gene expression. between SMO and ChD.

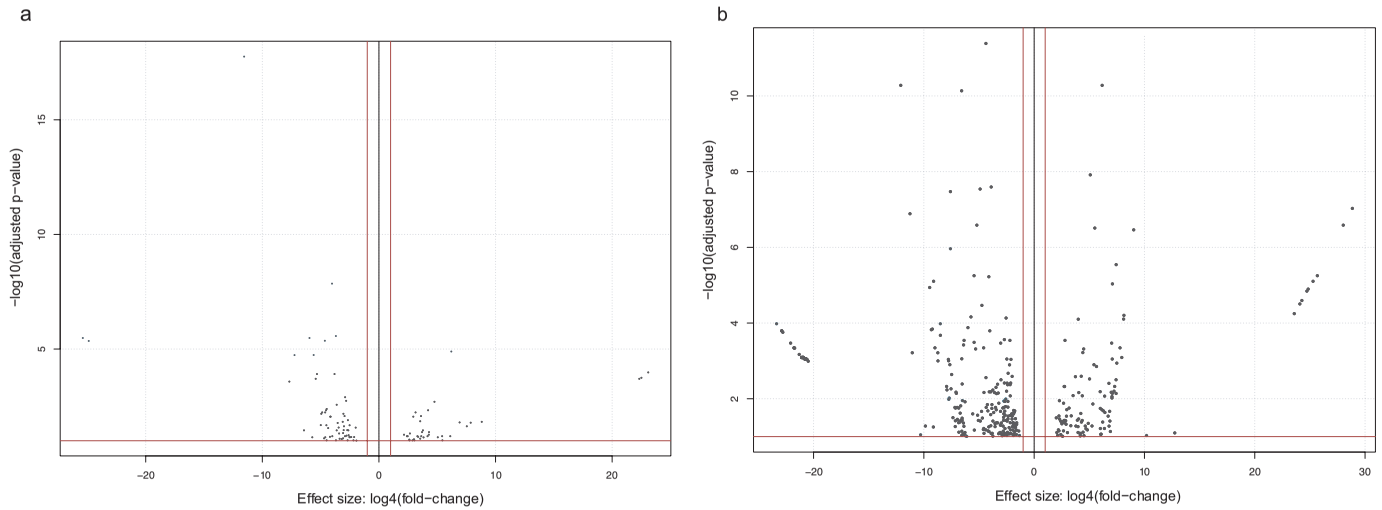

Figure S3 Volcano plot with the dispesion on the expression pattern agains natural locations and arboretum samples.
a. Chihuahuan Desert vs Arboretum. b. Sierra Madre Oriental vs Arboretum

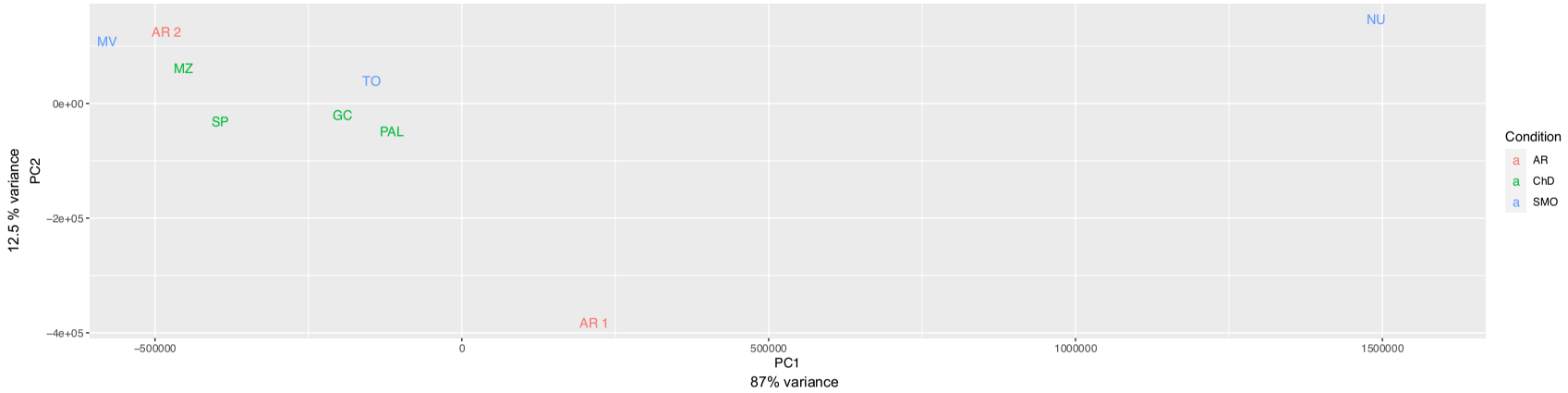

Figure S4. PCA analysis clustering of samples according the expression pattern.

Capítulo 2

Changes in demography and geographic distribution in the weeping pinyon pine (*Pinus pinceana*) during the Pleistocene

Laura Figueroa-Corona, Alejandra Moreno-Letelier, Diego Ortega Del-Vecchyo, Pablo Peláez, David S. Gernandt, Luis E. Eguiarte, Jill Wegrzyn & Daniel Piñero

El siguiente capítulo está constituido por el artículo: Changes in demography and geographic distribution in the weeping pinyon pine (*Pinus pinceana*) during the Pleistocene. Los objetivos buscan identificar el efecto de cambios geológicos y climáticos durante el Pleistoceno a la historia demográfica y la diversidad genética de *Pinus pinceana* (Gordon & Glend.). Utilizando 4,141 SNPs obtenidos por Genotyping by sequencing (GBS) para 90 individuos adultos de diez poblaciones naturales en la distribución geográfica para inferir la demografía histórica usando un método basado en el espectro de frecuencias por sitio (fastsimcoal v2.6; fsc26) con el objetivo de contrastar posibles escenarios filogeográficos y reconstruir la divergencia de linajes. Además se recontruyó la distribución histórica de *P. pinceana* utilizando reconstrucciones climáticas para el Holoceno medio (MH; 6 kya), Último Máximo Glaciar (LGM, 22 kya) y Último Interglaciar (LIG; 120-140 kya). Los resultados principales señalan que *Pinus pinceana* divergió en dos linajes ~ 627 kya (95% CI: 584.52 - 633.37), colonizando dos regiones: la Sierra Madre Oriental (SMO) y el Desierto Chihuahuense (ChD). En consecuencia, el aislamiento tuvo un impacto en cómo se distribuye la diversidad genética en estos dos grupos genéticos, mientras los ciclos interglaciares ocurrieron y la aridificación aumentó en el Desierto Chihuahuense con la redución de los paleolagos. Los modelos de nicho ecológicos mostraron una mayor estabilidad del hábitat en Mazapil, Zacatecas (Central ChD), una de las poblaciones con mayor diversidad genética sugiriendo que los ciclos interglaciares, mantuvieron esta región como un refugio seguido de la progresiva aridificación del Desierto Chihuahuense.

Changes in demography and geographic distribution in the weeping pinyon pine (*Pinus pinceana*) during the Pleistocene

Laura Figueroa-Corona[✉1 2], Alejandra Moreno-Letelier[3], Diego Ortega-Del Vecchyo[4], Pablo Peláez [5], David S. Gernandt[6], Luis E. Eguiarte[2], Jill Wegrzyn[7], Daniel Piñero[2]

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

[2]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

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

[4] Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Juriquilla, Mexico

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

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

[7]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA

L.F.C. laurafc25@gmail.com

A.M.L. amletelier@ib.unam.mx

D.O.D. dortega@liigh.unam.mx

P.P. ppelaez@ccg.unam.mx

D.S.G. dgernandt@ib.unam.mx

L.E.F. fruns@unam.mx

J.W. jill.wegrzyn@uconn.edu

D.P. pinero@unam.mx

Acknowledgments

To the memory of Dr. Miguel Angel Soto Arenas, in recognition for the discovery of El Palmito locality. We particularly thank Velia and Miriam Soto Arenas for showing us the population in 2009. We also thank Carmen Bautista, Patricia Delgado, Celestino Flores, Socorro González, Azalea Guerra, Ulises Manzanilla, Rosa Maricel Portilla, Rodolfo Salas-Lizana, Alejandra Vázquez-Lobo, Alfredo Villarruel and Ana Wegier for fieldwork assistance, and Sumaira Zaman for bioinformatic assistance in variant calling. We also acknowledge the Institute for Systems Genomics Computational Biology Core, UCONN for providing access to software and hardware. Fieldwork funding was provided by PAPIIT-DGAPA, UNAM Grant IN209816 to D.S.G. and sequencing funding was provided by IE (UNAM) to D.P. This work was supported by Consejo Nacional de Ciencia y Tecnología through PhD Scholarship 449615 and constitutes a partial fulfillment of the Ph.D. degree (Posgrado de Ciencias Biológicas, Universidad Nacional Autónoma de México, UNAM) to L.F.C.

Conflict of Interest Statement

The authors declare no conflict of interest

Abstract

Aim

To identify the processes underlying the demographic history and the genetic diversity of *Pinus pinceana* (Gordon & Glend.), distributed in the Chihuahuan Desert and the Sierra Madre Oriental, in the context of geological and climatic changes during the Pleistocene.

Location

The Chihuahuan Desert and Sierra Madre Oriental, Mexico.

Methods

We generated 4 141 SNPs from genotyping by sequencing on 90 adult individuals from ten natural populations covering the entire geographic distribution of *P. pincean*. This resource was used to infer its population history using a site frequency spectrum-based method (fastsimcoal v2.6; fsc26) to contrast possible phylogeographic scenarios of lineage divergence. The past distribution for *P. pinceana* was studied using climate reconstructions

for the Mid-Holocene (MH; 6 kya), Last Glacial Maximum (LGM, 22 kya), and Last Interglacial (LIG; 120-140 kya).

Results

Pinus pinceana diverged into two lineages ~627 kya (95% CI: 584.52 - 633.37), colonizing two regions: the Sierra Madre Oriental (SMO) and the Chihuahuan Desert (ChD). Consequently, isolation had an impact on how genetic diversity is distributed in these two lineages, while the interglacial cycles occurred and aridification increased in the Chihuahuan Desert while paleolakes were reduced in the area. Ecological niche models showed major habitat stability in Mazapil, Zacatecas (Central ChD), one of the most genetically diverse populations.

Main conclusion

The population genomic analysis supports the divergence of two separate *P. pinceana* lineages during the climatic changes of the Pleistocene. These changes were related to the glacial-interglacial cycles, followed by the gradual aridification of the Chihuahuan desert, resulting in population isolation.

Keywords

Chihuahuan Desert, Pleistocene refugia, *Pinus pinceana*, demographic expansion.

Introduction

Chihuahuan Desert (ChD) communities have a long history of high variability in their spatial dimensions driven primarily by water input, which is scarce and has fluctuated over time (Zavala-Hurtado & Jiménez, 2020; Noy-Meir, 1973). Thus, the geological and climatic events of the Miocene–Pliocene, as well as the climatic cyclical changes of the Pleistocene, affected the diversity patterns of Chihuahuan Desert communities (Scheinvar et al., 2020). Indeed, several phylogeographic studies have reported dynamic scenarios that populations migrated and expanded over the Central Mexican Plateau from north to south and during the interglacial (15-20 glacials cycles in the Pleistocene) recolonized the northern Chihuahuan Desert (Van Devender & Burgess, 1985). See, for instance, the cases of *Ephedra compacta* or of *Agave lechuguilla* (Loera et al., 2017; Scheinvar et al., 2017), and others like *Pinus remota* which reduced to Pleistocene refugia in the Bolsón de Mapimí (Lanner & Van Devender, 1981).

The ChD is a high-elevation desert ranging from 600 to 1 675 masl and receives more rain than other deserts (235 mm of mean annual precipitation; National Park Service, https://www.nps.gov/im/chdn/ecoregion.htm). ChD is the third-largest desert in the American continent, and it is considered the most diverse in the Western Hemisphere, and one of the most diverse arid regions in the world (Beck & Gibbens, 1999). It also has both a high species diversity and richness of endemism (Rzedowski & Calderón de Rzedowski, 1993). ChD includes different endorheic basins, including those of the Nazas, Aguanaval, and Casas Grandes Rivers. It also possesses bolsons or endorheic desertic valleys, like the Bolsón de Mapimí and El Salado. The ChD landscape maintains different vegetation types, such as grass steppes, xeric shrubs in the intermountain plains and valleys, and open woodlands at higher elevations (Van Devender & Burgess, 1985; Czaja et al., 2014).

The Sierra Madre Oriental (SMO) is a high mountain range (1 480 - 3 720 masl) formed by a mountain chain dated to the Late Cretaceous and early Tertiary Laramide (80 – 40 Ma) formed by the deformation of Mesozoic rocks (Eguiluz de Antuñano et al., 2000). With a high climatic diversity due to its complex physiographic heterogeneity and meteorological phenomena (Hernández-Cerda & Carrasco-Anaya, 2004), SMO includes the Sabinas and Pánuco Basins, and receives moisture from the Gulf of Mexico; while the eastern slopes have tropical and temperate forests, the western slope is much drier with xeric scrubs and pine-oak-juniper woodlands (Rzedowski, 1978; Morrone et al., 2017).

The ChD biota has experienced climate fluctuations throughout geologic time. Some of these fluctuations occurred during the Late Quaternary. Paleoecological reconstructions suggest that a woodland corridor covered the area between the Sierra Madre Oriental and the Sierra Madre Occidental in Mexico (Lanner & Van Devender, 1981; Metcalfe, 2006). The Late Wisconsin vegetation that covered most of the ChD during the Pleistocene was composed of pinyon pines, junipers, and oaks, and started to decrease 11 kya (Lanner & Van Devender, 1981; Van Devender & Burgess, 1985; Elias & Van Devender, 1990). One of several reductions in pinyon-juniper and oak woodland occurred during the last 2.5M years (Pleistocene), which included intense environmental and climatic changes. During the Pleistocene, 11 climatic cycles of growth and reduction of the polar ice cap occurred in North America, affecting the global climatic conditions and consequently the composition of species associated with this woodland corridor (Lanner & Van Devender, 1981). A drastic

climate change during the Holocene turned this corridor into a warmer and drier area, starting with the reduction of the paleolakes formed during the last glaciation (~115-117 kya) and culminating with the present-day aridification, resulting in a shift to xeric vegetation (Ortega-Ramirez et al., 2004; Castiglia & Fawcett, 2006; Czaja et al., 2014).

Pinus pinceana is an endangered conifer (SEMARNAT 059-2010; https://www.profepa.gob.mx/innovaportal/file/435/1/nom_059_semarnat_2010.pdf) that inhabits rocky soils with conditions of extreme aridity. It is locally restricted to slopes of the ChD and the western slopes of the SMO (1 480 - 3 000 masl; Farjon et al., 1997). Previous population genetic studies have shown that *P. pinceana* has a high diversity within populations and high differentiation between populations (Escalante, 2001; Ledig et al., 2001; Molina-Freaner et al., 2001).

This study focuses on how phylogeography, historical demography, and historical migration have influenced the current genetic variation and its geographic distribution of *P. pinceana*. We used genome-wide single nucleotide polymorphisms (SNPs) data to apply a coalescent-based simulation framework to infer its past demographic history during the expansion over the ChD and the SMO. We also tested the hypothesis that intraspecific gene flow may have occurred due to increased geographical contact between the two lineages during the last 2 Ma.

Methods

Sampling and genetic diversity characterization

Needles from a total of 90 healthy adult trees were arbitrarily chosen from ten populations distributed across the entire native range of *P. pinceana*. Figure 1 shows the regionalization in biogeographic provinces following Morrone et al. (2017). DNA was extracted using a modified CTAB protocol (Doyle & Doyle, 1987). Samples with a DNA concentration over 100 ng/ml were double-digested with the restriction enzymes *Pstl* / *Mspl* and fragments were single-end sequenced in 26 cycles with an Illumina Hi-Seq 2500 at the Wisconsin Biotechnology Center, Illinois.

The read adapters were removed using TRIMMOMATIC (v0.36; Bolger, 2014), and sequences with a minimum length of 80 bp in and a minimum Phred quality score of 30, were retained. The procedure was visualized with FASTQC (v0.11.7;

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MULTIQC (v1.9; Ewels, 2016).

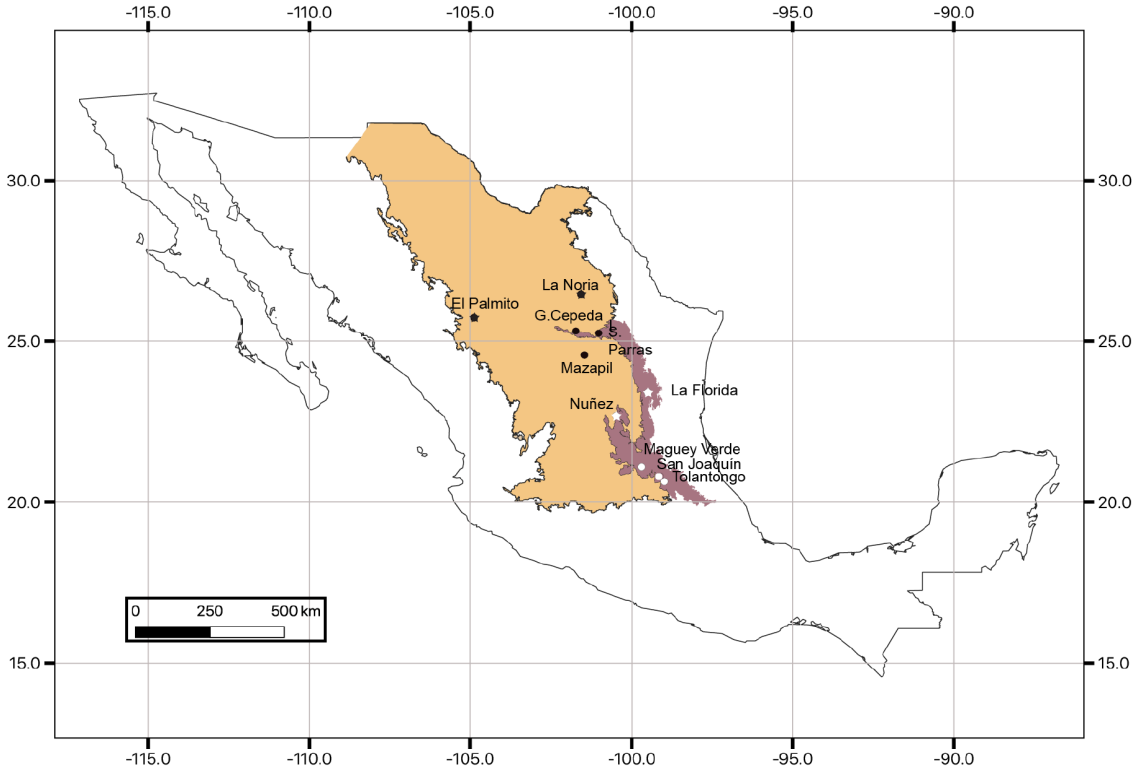

Figure 1. Geographic locations of 10 populations of *Pinus pinceana* involved in this study. The biogeographic province that encompasses the Chihuahuan Desert is located inside the sand color area, and the sampled populations from this province are indicated by a black dot, black stars correspond to Chw. The Sierra Madre Oriental is located inside the purple area and the sampled populations from this area are show with white dots, white stars correspond to SMOn.

High-quality reads were mapped to the *Pinus lambertiana* genome assembly (Stevens et al., 2016; v1.5) using the BWA-MEM algorithm (v0.7.10; Li et al., 2009). Unmapped reads were discarded. The MarkDuplicates feature from the Picard package (http://broadinstitute.github.io/picard/) was used to correct for artifacts of PCR duplication. Finally, we used HaplotypeCaller and VariantFiltration tools from the Genome Analysis Toolkit (GATK v3.7; McKenna et al., 2010) to filter SNPs by quality-by-depth (QD 2.0), mapping quality (MQ 40.0), Fisher strand bias test (to determine differences in the number

of sequences that support the reference and alternate alleles on each strand; FS > 60.0), rank sum test for evaluation of the mapping qualities supporting the reference or alternate alleles (MQRankSum < -12.5), rank sum test for assessment of bias in the position of alleles within the sequences (ReadPosRankSum < -8.0), and strand bias between forward and reverse strands by the symmetric odds ratio test (SOR > 4.0).

We obtained a set of biallelic SNPs that were filtered prior to analysis using VCFtools (v0.1.16; Danecek et al., 2011). The filtering parameters to retain biallelic SNPs had a minimum read depth of 15, and a maximum missing data of 0.75. We expected to get a better representation of singletons and common alleles controlling the proportion of missing data to reduce the stochastic variation (mutation) incorporated during the library construction and the sequencing process to obtain more confidence in the accuracy of the variant call. Finally, to avoid putative paralogs in both arrays, we used HDplot (v0.5-7; McKinney et al., 2017) to filter polymorphisms with *H* (heterozygosity) larger than 0.6 and *D* (deviation from even read-ratios in heterozygous) outside the range of -10 and +10. These parameters were chosen according to the proportion of heterozygous individuals within a population and allelic ratios within heterozygous individuals. We retained a total of 4 141 SNPs.

We estimated the genetic diversity for the ten populations using observed heterozygosity (H_O), gene diversity (H_E), and the inbreeding coefficient (F_{IS}) using the *R* package *Hierfstat* (Goudet, 2005). Additionally, DNAsp (v6; Rozas et al., 2019) was used to calculate allelic richness, per site nucleotide diversity (π), and pairwise population differentiation (F_{ST}; Weir & Cockerham, 1984).

Phylogeographic and population structure

The Admixture (v1.3.0; Alexander & Lange, 2011) was used to estimate the population structure among individuals for *K* values from 2 to 10 admixture proportions. Genotypes were converted to ordinary PLINK files (.ped) using PLINK (v1.9; Purcell et al., 2007).

To estimate the clustering among samples without any prior for geographic provenance, we ran a discriminant analysis of principal components (DAPC) with the *r* package *Adegenet* (Jombart & Ahmed, 2011).

Hypotheses of demographic history

Demographic history of *P. pinceana* was inferred by testing six different scenarios via the composite likelihood approach of fastsimcoal (v2.6; Excoffier & Foll, 2011; Excoffier et al., 2013) using the site (allele) frequency spectrum (SFS) of 4 141 SNPs filtered without bias of frequency computed by *easySFS.py* (Figure S1; https://github.com/isaacovercast/easySFS) to infer the divergence times between and within biogeographic regions (Figure 2).

Following the clustering of DAPC our six scenarios modeled were: Divergence 1, to estimate divergence times between ChD and SMO populations (T_{DIV}), and effective population sizes (N_e). Divergence 2, to estimate the divergence times between the ChD (T_1) in the northwest El Palmito and La Noria populations (ChDw), followed by the divergence between ChD and SMO populations (T_{DIV}). Finally, Divergence 3 followed the clustering of DAPC to estimate the divergence time between and within regions as colonization or fragmentation events, within the ChD (T_1) in the northwest El Palmito and La Noria populations (ChDw) and within SMO (T_2) in the northern Nuñez and La Florida populations (SMOn), after the divergence event between ChD and SMO populations (T_{DIV}). These three scenarios were redrawn with the same parameters ranges, adding asymmetrical migration rates among groups to construct the migration versions 1M, 2M, and 3M.

For each scenario, we estimated divergence times converted from generations to years using a generation time of 100 years. This generation time represents an approximation of the average age of reproductive individuals in a population, or the generation time for overlapping generations for this species (Franco & Silvertown, 2004). The mutation rate was set at 7.28×10^{-10} per base pair per generation as reported for *Pinus* (De la Torre et al., 2015). Thus, settings considered that individuals of *P. pinceana* have extended lifespans and a delayed reproduction (Pavek, 1994; Villanueva-Díaz et al., 2010).

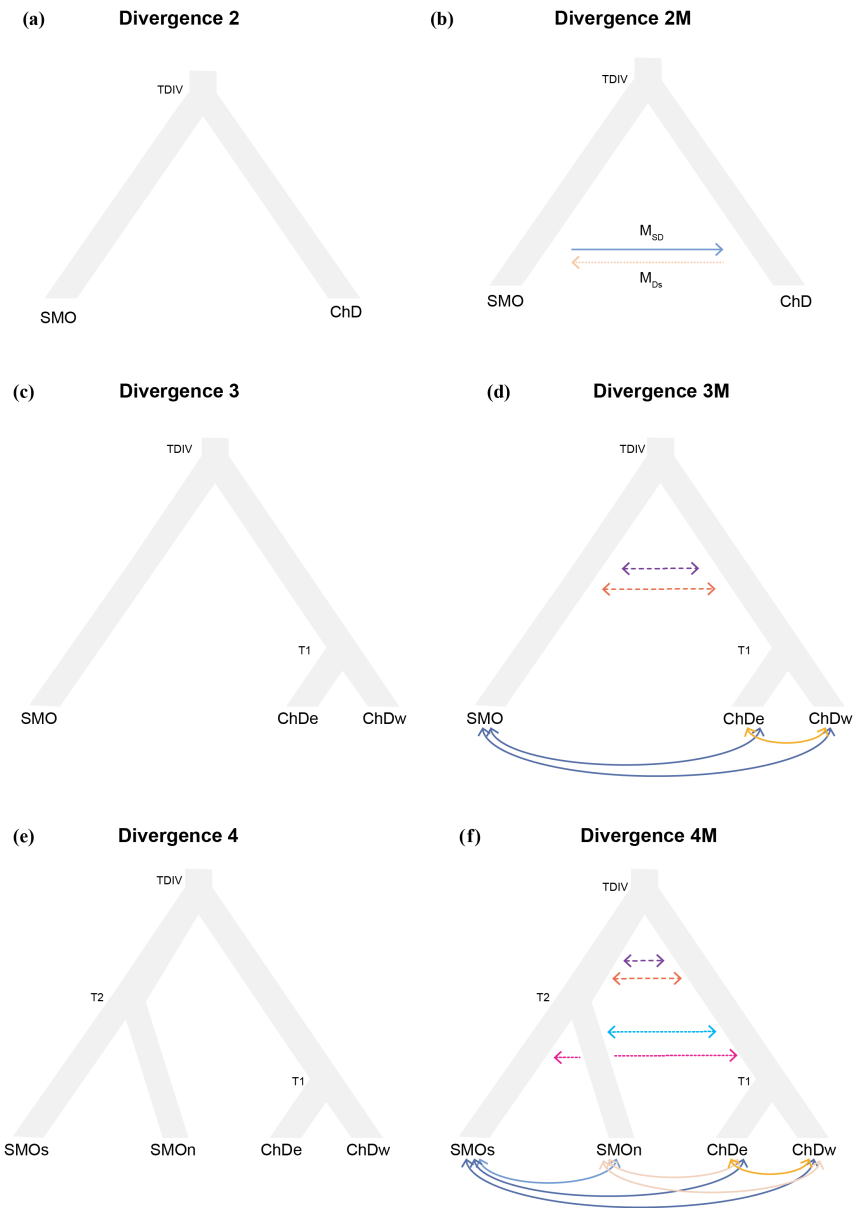

Figure 2. Modeled demographic histories for *Pinus pinceana.* (a) Divergence 1, model draw for the divergence between SMO and ChD. (b) Divergence 1M including asymmetrical and bidirectional migration among regions. (c) Divergence 2, model draw for the divergence within the ChD followed by divergence between SMO and ChD. (d) Divergence 2M including asymmetrical and bidirectional migration among regions. (e) Divergence 3 model draw including fragmentation events in ChD to the NW and SMO to the SMO_n. (f) Divergence 3M including asymmetrical and bidirectional migration among every region.

Divergence 1 with no-migration model was set using search ranges for the effective population sizes (N_e) from 100 to 10 000 individuals in the SMO and ChD lineages, and a T_{DIV} between 5 00 and 9 000 generations ago. Divergence 2 was set using search ranges for the effective population sizes (N_e) the SMO and ChD from 100 to 10 000 individuals, and from 70 to 1 000 individuals for ChDw. T_1 was set to date the divergence event within the ChD at between 100 and 1 000 generations ago and T_2 was set to date the divergence event between ChD and SMO with the range of 5 000 and 10 000 generations ago. Divergence 3 no-migration N_e of the northwestern Chihuahuan Desert (NW) and the SMOn region was set from 10 to 900 individuals and between 1 000 and 10 000 individuals for the ChDe and SMOs. For date T_1 the divergence event within the ChD was set from 50 to 1 000 generations, for T_2 the divergence event within the SMO from 1 000 to 2 000 generations ago, and for T_3 date the divergence event between ChD and SMO from 5 000 to 10 000 generations ago. Finally, for migration models, we used the same range adding between 1 $x10^{-9}$ to 1 $x10^{-4}$, per base per generation, drawn from a log uniform prior distribution.

For each of the six models, we simulated 100 replicates, 150 000 iterations, and 30 ECM cycles. To select the best scenario, we estimated the Akaike Information Criterion AIC value (Akaike, 1974) using the maximum composite likelihood estimated from the best replicate and compared the models using the weighted Akaike information criterion (wAIC). Finally, we calculated confidence intervals of parameter estimates from 100 parametric bootstrap replicates by simulating SFS from the maximum composite likelihood estimates and re-estimating the parameters each time (Excoffier et al., 2013).

Ecological niche modeling and refugia hypotheses

We compiled 37 records verified from previously published results and from specimen records in the National Herbarium of Mexico (MEXU). Each presence point was separated by at least 5 km from other presence points, to avoid spatial autocorrelation. We extracted the values for 19 bioclimatic variables with a resolution of 30 arc-seconds from WorldClim (Fick et al., 2017) using QGIS (v3.10; https://qgis.org/es/site/) and performed an outlier analysis using Mahalanobis distances using the package JMP (v9; SAS Institute Inc., Cary, NC, 2010).

To reduce model overfitting, we selected independent bioclimatic variables before constructing the ENMs with a Principal Component Analysis using JMP (v9; SAS Institute Inc., Cary, NC, 2010). We retained Annual Mean Temperature (Bio1), Isothermality (Bio3), Temperature Seasonality (Bio4), Annual Precipitation (Bio12), and Precipitation in the Driest Month (Bio14) from PC1; Precipitation Seasonality (Bio15), and Precipitation of Warmest Quarter (Bio18) from PC2; and Minimum Temperature of Coldest Month (Bio6) from PC3.

To construct a potential niche model, we used Maxent (v3.3.3e; Phillips & Dudík, 2008) with linear and quadratic features, a regularization multiplier of 1, 30% of random test percentage, clamping, and 10 replicates. We performed projections to three points in the past: Mid-Holocene (MH; 6 kya), Last Glacial Maximum (LGM, 22 kya), and Last Interglacial (120-140 kya; Otto-Bliesner et al., 2008). For the Mid-Holocene (MH) and LGM, we used CCSM4 and MIROC-ESM World climate models. We processed the resulting average models by removing the lower 10^{th} percentile of probability values to eliminate areas of over prediction. All raster processing was performed with the package *raster* in RStudio (Hijmans & van Etten, 2012). Areas were estimated by counting pixels with the package *raster,* where each pixel corresponds to approximately 1 km^2.

Results

Sequencing and genotype characterization

The single-end sequences produced an average of 2.8 M for each individual, which was reduced to 2.54 M after quality control. The datasets obtained with GATK produced 613 651 variants. VCFtools was used to obtain the final processed data with 4 141 biallelic SNPs, with a moderate amount of missing data (mean=9.89%), and a mean depth per site of 61.7×.

Genetic diversity was estimated for the ten populations (Table 1). Globally, the observed heterozygosity (H_O) was 0.19, the expected heterozygosity (H_E) was 0.1231, and the inbreeding coefficient (F_{IS}) was -0.5653. Levels of diversity (H_E) were similar among populations and ranged from 0.103 (Maguey Verde, SMO) to 0.1354 (General Cepeda; ChD). Levels of H_O were highest for the ChD location, General Cepeda (0.2166), and the lowest for Maguey Verde (0.16) in SMO. Estimates for the inbreeding coefficient F_{IS} were negative for all populations, ranging from -0.5412 (Maguey Verde; SMO) to -0.6058 (La

Florida; SMO), while the highest nucleotide diversity was found in General Cepeda (0.0116), and the lowest in Maguey Verde (0.0083).

Table 1. Measures of genetic diversity for 90 individuals of *P. pinceana* from ten populations calculated from 4 141 single nucleotide polymorphism loci.

	Population	H_E	H_o	F_{IS}	π	Tajima D	θ_W
SMO	Tolantongo	0.1127	0.1790	-0.5456	0.0104	-0.0550 $_{NS}$	0.1333
	San Joaquín	0.1111	0.1811	-0.5865	0.0108	0.0782 $_{NS}$	0.1400
	Maguey Verde	0.1033	0.1640	-0.5412	0.0083	-0.0345 $_{NS}$	0.1140
	Nuñez	0.1166	0.1907	-0.5947	0.0113	0.0923 $_{NS}$	0.1448
	La Florida	0.1130	0.1843	-0.6058	0.0099	0.1307 $_{NS}$	0.1445
ChD	El Palmito	0.1166	0.1849	-0.5927	0.0097	0.0247 $_{NS}$	0.1408
	Mazapil	0.1269	0.2024	-0.5475	0.0101	0.0303 $_{NS}$	0.1566
	La Noria	0.1291	0.2069	-0.5551	0.0109	0.0861 $_{NS}$	0.1615
	G. Cepeda	0.1354	0.2166	-0.5522	0.0116	0.1444 $_{NS}$	0.1686
	Sierra Parras	0.1324	0.2144	-0.5759	0.0115	0.1746 $_{NS}$	0.1702
Global		0.1231	0.1900	-0.5653	0.0074	-0.5203 $_{NS}$	0.1535

H_O, observed heterozygosity; H_E, expected heterozygosity within populations; F_{IS}, inbreeding coefficient; π, nucleotide diversity

The average F_{IS} by locus was -0.195 while the median was -0.03. This difference is due to a biased average due to high frequency of loci with a high proportion of heterozygous SNPs. Table S1 shows the calculation for the mean and median values on F_{IS} for differents sets after deleting the extreme values (F_{IS} < -0.799). This changes the average F_{IS} estimates (using the remaining 3 762 SNPs) to an observed heterozygosity (H_O) of 0.093, an expected heterozygosity (H_E) equal to 0.116, and an average inbreeding coefficient (F_{IS}) of -0.35. This effect was observed in all populations (Table S2).

Genetic structure and phylogeography pattern

The global F_{ST} value was 0.0612 (Table 2). Within SMO, the populations had a mean F_{ST} = 0.012, compared to 0.010 within the ChD populations. Between biogeographic regions the mean value was 0.055.

The discriminant analysis of principal components (DAPC) showed two clear population clusters divided by the PC1 describing 14.9% of the total variance, and sub-clustering within the biogeographically regions showed by PC 2 that describe 4.09% of the total variance, where La Noria and El Palmito had the greatest divergence within the ChD populations and Nuñez and La Florida show the largest divergence within the SMO (Figure 3A). Admixture analysis also showed that *K*= 2 supports the clustering in two biogeographic areas. For *K*= 2, 42 of the 44 samples from SMO were classified 100% in one cluster (Figure 3B) and 43 samples from ChD in the second cluster. Three samples were found with a level of admixture between both groups composed of individuals from Maguey Verde and northwest populations El Palmito and La Noria. The second-best supported clustering was found for *K*= 3 denoting a longitudinal gradient of admixture in the ChD samples from the northwest to the center populations (Figure 3C). Finally, for *K*= 4 (Figure 3D) the two biogeographical regions showed substructure between the NW localities El Palmito and La Noria and the rest of the ChD and between the SMO with Nuñez and La Florida and the south of the SMO.

Table 2. Pairwise differentiation (F_{ST}) among the ten populations sampled of *P. pinceana* with 4 141 SNPs

SMO	Tolantongo									
	San Joaquín	-0.0012								
	Maguey Verde	-0.0004	-0.0006							
	Nuñez	0.0022	0.0084	0.0101						
	La Florida	0.0018	0.0093	0.0104	0.0008					
ChD	El Palmito	0.0331	0.0373	0.0362	0.0351	0.0358				
	Mazapil	0.0377	0.0388	0.0445	0.0424	0.0418	0.0074			
	La Noria	0.0402	0.0418	0.0469	0.0440	0.0445	-0.0007	0.0007		
	G. Cepeda	0.0410	0.0451	0.0496	0.0477	0.0459	0.0087	-0.0012	-0.0040	
	Sierra Parras	0.0469	0.0458	0.0527	0.0516	0.0501	0.0101	-0.0025	0.0016	-0.0057

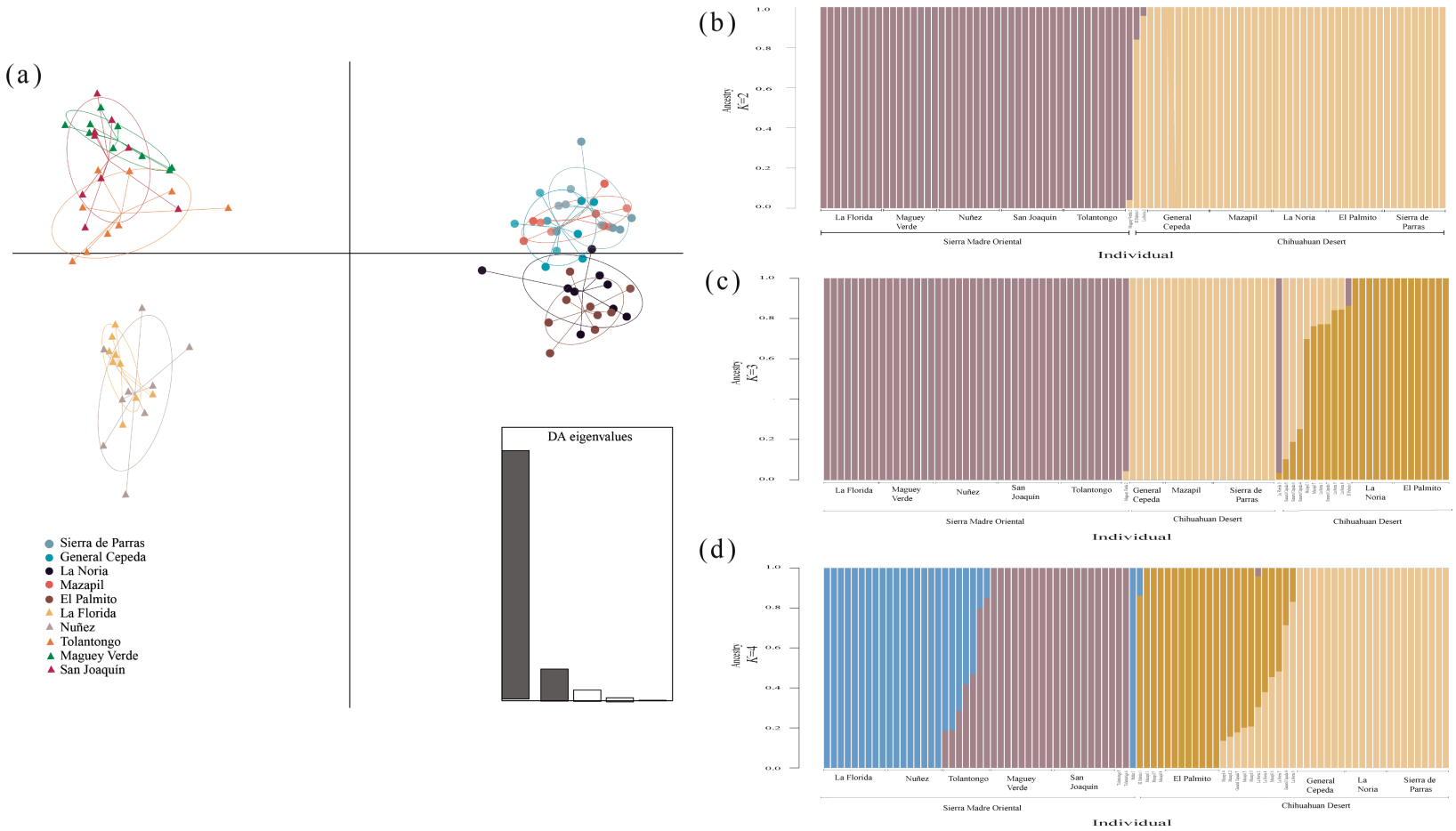

Figure 3. Population structure and nucleotide diversity based on 4 141 genome-wide SNPs in *Pinus pinceana.* (a) Scatterplots resulting from discriminant analysis of principal components (DAPC) for all 90 samples, and the eigenvalues of discriminant functions portrayed in the bottom right. (b) Admixture for *K*= 2, (c) for *K*= 3 and (d) for *K*= 4.

Demographic history

We used the entire 4 141 SNP dataset although it could include paralogs since there was no filter by frequency for heterozygous alleles. The population divergence time hypotheses from the demographic simulations for the six scenarios resulted in divergence time estimated during the Middle Pleistocene (Table 3). The best-supported model was the Divergence 1M according to the ΔAIC and considering the split of the two genetic clusters. The dated divergence (T_{DIV}) between them was estimated in the Middle Pleistocene (627 kya; 95% CI: 584.52 - 633.37; Table 4). Effective population sizes (*N*e) for groups were similar between groups, SMO with 6 711 individuals (95% CI: 6 477.24 – 6 763.9), and ChD with 6 639 individuals (95% CI: 6 422 – 6 689). Migration rates per generation (*m*) were estimated to be significantly higher 1.03 $x10^{-5}$ (95% CI: 9.09 $x10^{-6}$ - 1.07 $x10^{-5}$) from SMO to ChD, than in the opposite direction (1.61 $x10^{-6}$; 95% CI: 1.04 $x10^{-6}$ - 1.83 $x10^{-6}$; Table 5). The sub-fragmentation according to the Divergence 3 and 3M occurred during the interglacial cycles and the Middle Pleistocene (~127.7 - ~539.2 kya).

Table 3. Comparison of demographic models analyzed with fastsimcoal

Model	*k*	AIC	ΔAIC
Divergence 1	3	40 347.47	1074.42
Divergence 1M	5	39 273.05	0
Divergence 2	5	45 847.82	6 574.77
Divergence 2M	13	39 768.40	495.35
Divergence 3	7	41 405.14	2 132.09
Divergence 3M	20	39 983.50	710.44

Table 4. Parameters inferred for time divergence and effective population sizes from coalescent simulations with fastsimcoal under the demographic models

	SMO	***ChD***	T_{DIV}
Divergence 1	6 739	6 640	627 300
Divergence 1M	6 711	6 639	627 000

	SMO	ChD_e	ChD_w	T_1	T_{DIV}
Divergence 2	57 186	57 227	5 870	1 840 800	3 483 100
Divergence 2M	44 836	22 880	3 784	1 208 500	2 667 900

	SMO_s	SMO_n	ChD_e	ChD_w	T_1	T_2	T_{DIV}
Divergence 3	7 831	268	25 322	2 191	223 900	127 700	572 000
Divergence 3M	22 872	472	7 619	1 564	539 700	539 200	541 600

Population names refers to the N_e number of individuals in the different groups; T_1, T_2, and T_{DIV} are in years; SMO: Sierra Madre Oriental, SMOs: south SMO Tolantongo, San Joaquín and Maguey Verde populations, SMOn: northern SMO Nuñez and La Florida populations, ChD: Chihuahuan Desert, ChDe: east ChD Mazapil, General Cepeda and Sierra de Parras, ChDw: northwest ChD El Palmito and La Noria populations.

Table 5. Migration rates inferred from coalescent simulations with fastsimcoal under the demographic models

Divergence 1M			SMO		SMO	ChD
			ChD		1.03×10^{-5}	1.61×10^{-6}
Divergence 2M				SMO	ChDe	ChDw
	T_1	SMO			3.44×10^{-7}	1.69×10^{-8}
		ChDe		7.72×10^{-5}		5.19×10^{-8}
		ChDw		1.58×10^{-5}	1.91×10^{-4}	
				SMO	ChD	
	T_2	SMO			7.35×10^{-8}	
		ChD		1.45×10^{-8}		
Divergence 3M			SMO_s	SMO_n	ChD_w	ChD_e
		SMO_s		1.80×10^{-6}	1.03×10^{-8}	5.46×10^{-9}
	T_1	SMO_n	1.3×10^{-8}		2.52×10^{-7}	2.90×10^{-7}
		ChD_w	7.74×10^{-9}	6.58×10^{-5}		5.08×10^{-6}
		ChD_e	4.01×10^{-9}	1.24×10^{-4}	3.47×10^{-8}	
				SMO	ChD_w	ChD_e
	T_2	SMO			1.93×10^{-8}	2.83×10^{-9}
		ChD_w		3.50×10^{-7}		2.97×10^{-8}
		ChD_e		9.22×10^{-8}	2.72×10^{-9}	
				SMO	ChD	
	T_3	SMO			8.39×10^{-6}	
		ChD		3×10^{-7}		

T_1, T_2, and T_{DIV} are in years; SMO: Sierra Madre Oriental, SMOs: south SMO Tolantongo, San Joaquín and Maguey Verde populations, SMOn: northern SMO Nuñez and La Florida populations, ChD: Chihuahuan Desert, ChDe: east ChD Mazapil, General Cepeda and Sierra de Parras, ChDw: northwest ChD El Palmito and La Noria populations.

Historical distribution projections

The present and past projections cover a broader area than what is observed in the current populations (Figure 4). There was a significant latitudinal range shift, as well as an area increase from the Last Interglacial (LIG) to the present. The LIG distribution shifted to the southern limit of the ChD, bordering with the Rio Pánuco biogeographical province (Scheinvar et al., 2020), and decreased by ca. 70% from the present area (116 k pixels), with 34 k pixels. The Last Glacial Maximum (LGM; 89 k pixels) and mid-Holocene (MH; 114k pixels) predicted ranges occupy a smaller area (23% and 7% smaller than the current distribution, respectively), but close to areas in the current range. Flatlands and basins seem to be a barrier in the distribution of *P. pinceana*, as the areas identified as current or past water bodies have not been occupied by the species. The areas that remained stable, that is, with suitable conditions for *P. pinceana* since the LIG were quite small (2 k pixels) and mostly found in the southern ChD.

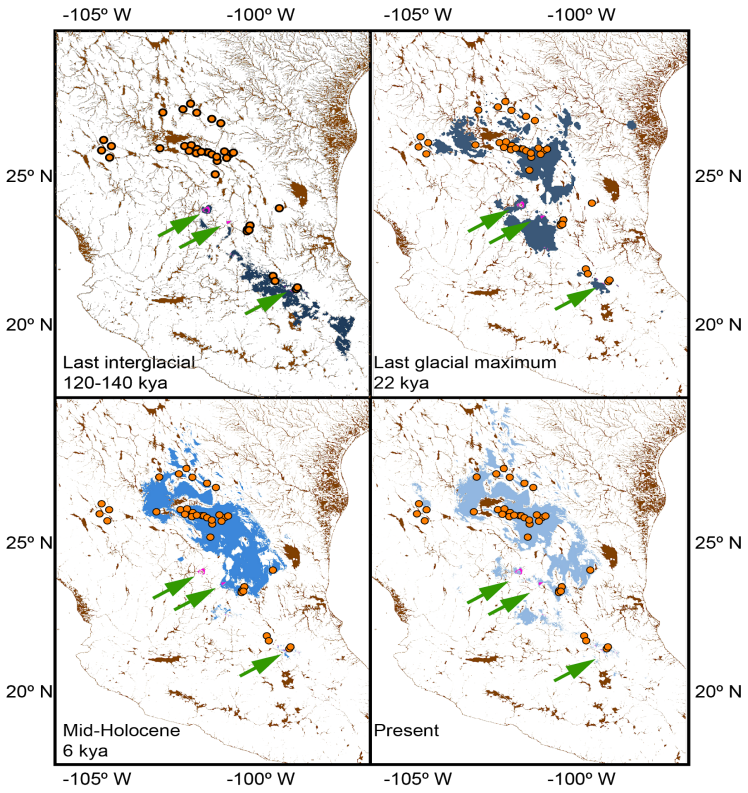

Figure 4. Ecological niche modeling predictions of *Pinus pinceana* current climate change scenarios (Last interglacial (LIG), Last glacial maximum (LGM), Mid-Holocene (MH), and present). Orange circles represent present localities, and blue indicates areas predicted for the species. Green arrows indicate areas that remained stable from the LIG to present.

Discussion

Phylogeography

We detected a clear geographic structure of genetic diversity of *P. pinceana* in the Chihuahuan Desert (Ortega-Ramirez et al., 2004; Castiglia & Fawcett, 2006; Czaja et al., 2014). This structure is related to climate dynamics during the Quaternary. These changes highlight the importance of two distinct processes: first, genetic drift due to reductions in effective population sizes in different areas; and second, fragmentation due to the aridification of the ChD.

The genetic diversity patterns were similar to the genetic diversity (H_O) values described in conifer species with broad geographic ranges in Mexico like *Pinus patula*, *Pinus strobiformis,* and *Picea chihuahuana* (Peláez et al., 2020; Moreno-Letelier & Piñero, 2009; Jaramillo-Correa et al*.,* 2006). The values previously reported for microsatellite (cpDNA) SSR loci in *P. pinceana* (Escalante, 2001; Figueroa-Corona, 2012) was substantially higher than the values found here. This is expected because microsatellites have higher mutation rates than biallelic SNPs located in functional genes (Hamblin et al., 2007). We found that four of the populations: La Noria in the northwest and Sierra de Parras, General Cepeda, and Mazapil in the ChD showed the highest levels of diversity (Table 1).

In accordance with previous reports in *P. pinceana* (Ledig et al., 2001; Molina-Freaner et al., 2001; Ledig et al., 1986) we found high observed heterozygous values. On the other hand, we surprisingly found negative F_{IS} averages. Similar situations with high heterozygosities and negative F_{IS} averages has been found in some conifers like *Criptomeria japonica* var. *sinensis, Pinus patula* and *P. albicaulis* (Cai et al., 2020; Peláez et al., 2020; Liu et al., 2016).

We attribute these F_{IS} values to two artifacts, first the sub-estimation of linked loci between the SNPs giving that analyzing the most extreme values of F_{IS} (-0.8 to -1) in the 4 141 SNPs dataset we found 377 fixed heterozygous SNPs, establishing a distance based on Pavy (et al., 2012) of 100 bp at maximum between polymorphisms we obtained 276 of them linked loci in 96 linkage groups conformed by two to eleven SNPs. The second possibility is the presence of some paralogous SNPs that were not filtered with the approach of McKinney

(et al., 2017) as has been suggested in view of the complexities of conifer genomes (De la Torre et al., 2014; Leitch et al., 2019).

The differentiation of trees from the SMO and ChD was attributed to isolation by distance in a previous study based on cpDNA SSR loci (Escalante, 2001). Our results of genetic structure clearly identified two lineages, corresponding to the same two biogeographic regions. Furthermore, these lineages present relatively high levels of genetic differentiation between them ($F_{ST} = 0.0612$), like that typically reported for conifer taxa with wide distribution ranges (Pelaez et al., 2020; Cobo-Simon et al., 2020; Jaramillo-Correa et al., 2020). Nevertheless, the discriminant components analysis shows that SMO and ChD groups have significant levels of admixture between them.

The structured biogeographic pattern between ChD and SMO along the Central Mexican Plateau has been reported for several desert scrub plant species from the ChD (*e.g.,* Sosa et al., 2009; Vasquez-Cruz & Sosa, 2016; Scheinvar et al., 2017). In all cases, a strong phylogeographic structure found within the SMO was correlated with geologic and paleoclimatic changes. As an example of this pattern, Vasquez-Cruz & Sosa (2016) identified a main phylogeographic pattern in species of Rosaceae where the ChD was the ancestral area, followed by contraction of suitable habitat during the Last Interglacial (~120 000 – 140 000 years), followed by an expansion during the LGM (~22 000 years) to the SMO and the Central Mexican Plateau and then recolonization of the ChD during the Mid-Holocene (~6 000 years). Loera et al. (2017) argued that this pattern between SMO and ChD regions in *Ephedra compacta* was reinforced by the lack of dispersal and changes in elevation due to biogeographic barriers within this region.

Demographic history

With a patchy but extensive geographic distribution, *P. pinceana* is not considered an endangered species by the IUCN, but it is considered endangered by Mexican law based on more detailed ecological, geographic, and genetic information (SEMARNAT 059-2010).

Demographic data can help us to apply historical context to the phylogeographic question. The current genetic differentiation patterns of the ChD xeric scrubs can be explained by an initial expansion followed by processes of isolation between lineages to the present very reduced zones of secondary contact and limited migration (Scheinvar et al.,

2017; Loera et al., 2017). The results of the six models give support consistent with changes in climate and the genetic structure founded in *P. pinceana* in the ChD during the Middle Pleistocene. We suggest that this allopatric fragmentation of two lineages, together with an expansion of the xeric scrubs and pine-oak forests, after the end of the glacial inter-cycles, was crucial in shaping the present distribution of the genetic diversity in *P. pinceana*.

Quaternary dynamics in the ChD have been described with geological and biogeographic evidence. Genetic evidence reveals different levels of diversities between the northern and southern ChD (Scheinvar et al., 2020). Genetic structure and demographic history during the Pleistocene have been found for the xeric scrubs *Agave lechuguilla, A. striata,* and *A. stricta* (Scheinvar et al., 2017; Trejo et al., 2016; Martínez-Ainsworth, 2013), the rodents *Thomomys umbrinus* and *Perognathus avus*, and the turtle *Kinosternon avescens* (Mathis et al., 2014; Neiswenter & Riddle, 2010; Serb et al., 2001). The pattern of differentiation between ChD and SMO has been found in *Ephedra compacta,* where the regionalization of the biogeographic zones in SMO and ChD matches with genetic structure (Loera et al., 2017).

The estimated migration rates show an asymmetric flux between regions (Table 5). This is probably due to the current distribution, which suggests that distance works as barrier for the extreme's population. Nevertheless, we still cannot establish a cause for the percentage of admixture in individuals from San Joaquín and El Palmito despite the low connectivity with nearby populations.

The separation of the northwestern ChD in the DAPC and the $K = 3$, represents a later, more recent fragmentation from the ChD populations (Figure 3). The western populations are at the very edge of the distribution, which suggests that Nazas basin increase the isolation and restricted gene flow. This pattern of West/East divergence has been described for Chihuahuan Desert xeric scrubs like *Astrophytum* spp., *Berberis trifoliate*, *Agave victoria-reginae, Leucophyllum* spp*., Ephedra compacta* (Vázquez-Lobo et al., 2015; Angulo et al., 2017; Gándara & Sosa, 2014; Loera et al., 2017), and in the snake *Crotalus molossus* (Anderson & Greenbaum, 2012).

Genetic diversity in a temporal context

Phylogenetic divergence of *Pinus* subsection *Cembroides* occurred during the Miocene (11 Mya; Gernandt et al., 2008; Saladin et al., 2017; Jin et al., 2021). The fossil evidence for pinyon pines records large changes in the distribution range during the Pliocene along the Central Mexican Plateau (Lanner & Van Devender, 1981; Van Devender & Burgess, 1985). Conifers had a wider distribution during the Pleistocene in Mexico in the ChD, and a complex dynamic around the Holocene that diminished or extremely reduced their populations (Quiñonez-Perez et al., 2017; Jaramillo-Correa et al., 2006).

The high frequency of singletons and high diversity within populations of *P. pinceana* in this highly fragmented habitat has likely occurred recently. Their long generation time and overlap of demographic cohorts are also likely to contribute to these patterns. The prevalence of new polymorphisms in all populations is maintained by relatively large effective population sizes.

Morpho-physiological studies in *P. pinceana* detected morphological differentiation patterns, where the individuals from the ChD had thicker epicuticular waxes on the needles, faster growth, and larger primary roots, and greater volume of lateral roots in Chihuahuan Desert individuals than in individuals from the Sierra Madre Oriental, differences that suggest local morphological adaptation to aridity exposure (Martiñón-Martínez et al., 2010; Córdoba-Rodríguez et al., 2011). We could not ascribe any adaptive relevance to the differentiation of the genetic clusters, in part because of the low coverage ssequencing of the gene space regions in the *Pinus* genome, and the absence of specific candidate genes, but also because the strongly structured signatures in genetic diversity produced by changes in demographic history. An independent analysis using comparative transcriptomics for this species in both biogeographic regions detected a possible plastic response to biotic and abiotic factors present in the populations (Figueroa-Corona et al., 2021).

Historical distribution

Based on our analyses, we propose that *P. pinceana* differentiated in two lineages in the Middle Pleistocene (~627 kya; 95% CI: 584.52 - 633.37), followed by the division of both biogeographic regions (~127.7 - ~539.2 kya) during the interglacial cycles. Over time the climatic conditions in the habitat of *P. pinceana* became more common, in contrast to *Picea*

chihuahuana, which has become more restricted (Quiñonez-Perez et al., 2017; Jaramillo-Correa et al., 2006). The recent climatic changes of the ChD, in particular the increment in aridification, and the reduction of water bodies (Czaja et al., 2014) promoted the geographic expansion of the species, and the phylogeographic patterns obtained by our data, involving a gradual expansion over the ChD since the LGM. This pattern of expansion has been described in the ChD in other scrub species like *A. lechuguilla* and *E. compacta* (Scheinvar et al., 2017; Loera et al., 2017). Throughout the Mid-Holocene there was more connectivity across regions, while the Nazas Basin determined the division within the ChD, producing the separation between the northwestern and central ChD despite the lack of connectivity given the climatic conditions.

Mazapil constitutes one of the three most genetically diverse populations but also represents, according to the estimation of the habitat, the most stable region over the climate dynamics since the interglacial cycles, and thus we interpreted that this region has been a refuge for *P. pinceana* over time. This region has not been previously described as a refuge for the ChD (Vasquez-Cruz & Sosa; Loera et al., 2017; Scheinvar et al.; Gámez et al., 2017).

Conclusions

The genetic diversity in *P. pinceana* was modified by dynamics during interglacial cycles in the Pleistocene. The demographic scenarios studied resulted in a ranking of models that were useful in gauging relative support for competing hypotheses. In particular, the best model involved the divergence of two lineages in the Middle Pleistocene (~627 kya; 95% CI: 633.37 – 584.52), which later colonized two biogeographic regions, the SMO and the ChD, occurring while interglacial cycles modified the Chihuahuan Desert as the aridification increased and paleolakes reduced their area. Both phylogenomic and population genomic coalescent-based analyses indicated that *P. pinceana* diverged into two lineages. This division probably occurred during the climatic changes of the Pleistocene and was related to the glacial-interglacial cycles. Thus, the phylogeographic history of *P. pinceana* is likely explained by climate dynamics that left perceptible marks in the patterns of genetic diversity and structure observed today in the species.

Data availability statement

Raw reads are available in the SRA NCBI database (https://www.ncbi.nlm.nih.gov/sra/) deposited under the PRJNA719106 BioProject.

References

Alexander, D.H. & Lange, K. (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation, *BMC Bioinformatics, 12*, 246.

Akaike, H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control, AC- *19*, 716-723.

Anderson, C.G. & Greenbaum, E. (2012) Phylogeography of northern populations of the black-tailed rattlesnake (*Crotalus molossus* Baird and Girard, 1853), with the revalidation of *C. ornatus* Hallowell, 1854. *Herpetological Monograph, 26*, 19–57.

Angulo, D.F., Amarilla, L.D., Anton, A.M. & Sosa, V. (2017) Colonization in North American arid lands: the journey of agarito (*Berberis trifoliolata*) revealed by multilocus molecular data and packrat midden fossil remains. *PLoS One, 12*, e0128559.

Beck, R.F. & Gibbens, R.P. (1999) The Chihuahuan Desert Ecosystem, New Mexico Journal of Science, p. 45. *Proceeding of the National Academy of Sciences of the United States of America*

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 15*, 2114-2120.

Cai, M., Wen, Y., Uchiyama, K., Onuma, Y., & Tsumura, Y. (2020). Population Genetic Diversity and Structure of Ancient Tree Populations of *Cryptomeria japonica* var. *sinensis* Based on RAD-seq Data. *Forests*, *11*(11), 1192.

Castiglia, P. & Fawcett J. (2006) Large Holocene lakes and climate change in the Chihuahuan Desert. *Geology, 34*, 113–116.

Córdoba-Rodríguez, D., Vargas-Hernández, J.J., López-Upton, J. & Muñoz-Orozco, A. (2011) Crecimiento de la raíz en plantas jóvenes de *Pinus pinceana* Gordon en respuesta a la humedad del suelo. *Agrociencia, 45*, 493-506.

Cobo-Simón, I., Méndez-Cea, B., Jump, A.S., Seco, J., Gallego, F.J. & Linares, J.C. (2020) Understanding genetic diversity of relict forests. Linking long-term isolation legacies and current habitat fragmentation in *Abies pinsapo* Boiss. *Forest Ecology Managment*, *461*,117947.

Czaja, A., Palacios-Fest M.R., Estrada-Rodríguez J.L., Romero-Méndez U. & Alba- Ávila J.A. (2014) Inland dunes fauna and flora from Paleolake Irritila in the Comarca Lagunera, Coahuila, northern Mexico. *Boletin de la Sociedad Geológica Mexicana*, *66*, 541–551.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G. & Durbin, R. (2011) 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics, 27*, 2156-2158.

De la Torre, A. R., Birol, I., Bousquet, J., Ingvarsson, P. K., Jansson, S., Jones, S. J., Keeling, C. I., MacKay, J., Nilsson, O., Ritland, K., Street, N., Yanchuk, A., Zerbe, P. & Bohlmann, J. (2014). Insights into conifer giga-genomes. *Plant Physiology, 166*, 1724–1732.

De la Torre, A.R., Lin Y.C., Van de Peer Y. & Ingvarsson P.K (2015) Genome-wide analysis reveals diverged patterns of codon bias, gene expression, and rates of sequence evolution in *Picea* gene families. *Genome Biology and Evolution, 7*, 1002–1015.

Doyle, J.J. & Doyle, J.L. (1987), A rapid DNA isolation procedure from small quantities of fresh leaf tissue. *Phytochemistry Bulletin, 19*, 11-15.

EasySFS (2016) Overcast I. https://github.com/isaacovercast/easySFS

Elias, S. A. & Van Devender, T. R. (1990) Fossil insect evidence for Late Quaternary climatic change in the Big Bend region, Chihuahuan Desert, Texas, *Quaternary Research, 34*, 249–261.

Eguiluz de Antuñano, S., Aranda–García, M. & Marrett, R. (2000), Tectónica de la Sierra Madre Oriental, México: *Boletín de la Sociedad Geológica Mexicana, 53*, 1–26.

Escalante, A. E. (2001) Estructura genética de poblaciones de *Pinus pinceana* G. Gordon & Glendinning usando como marcadores moleculares microsatélites de cloroplasto (cpSSR's), dissertation Universidad Nacional Autonoma de Mexico.

Excoffier, L. & Foll, M. (2011) Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios, *Bioinformatics, 27*, 1332–1334.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C. & Foll, M. (2013) Robust demographic inference from genomic and SNP data. *PLOS Genetics, 9* e1003905.

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

Farjon, A., de la Rosa, J. A. P. & Styles, B. T. (1997) A field guide to the pines of Mexico and Central American Royal Botanic Gardens.

FASTQC tool (2010) Babraham Bioinformatics https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Fick, S.E. & Hijmans R.J. (2017) WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37 4302-4315.

Figueroa-Corona, L. (2012) Genética de la conservación en pinos piñoneros de las subsecciones *Cembroides* y *Nelsoniae* de México. Universidad Nacional Autónoma de México, Mexico.

Figueroa-Corona, L., Delgado, P., Wegrzyn J. & Piñero D. 2021. Transcriptome of weeping pinyon pine, *Pinus pinceana*, shows differences across heterogeneous habitats. *Trees, 35*, 1351–1365.

Franco, M. & Silvertown, J. (2004) A comparative demography of plants based upon elasticities of vital rates. *Ecology*, *85*, 531-538.

Gándara, E., & Sosa, V. (2014) Spatio-temporal evolution of *Leucophyllum pringlei* and allies (Scrophulariaceae): a group endemic to North American xeric regions. *Molecular Phylogenetics and Evolution, 76*, 93–101.

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

Goudet, J. (2005), Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes, 5*, 184-186.

Hamblin, M.T., Warburton, M.L. & Buckler, E.S. (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PLoS ONE 2* e1367.

Hernandez-Cerda, M., Carrasco-Anaya, G. Climatología In Luna, Isolda, Juan J. Morrone, and David Espinosa. 2004. *Biodiversidad de La Sierra Madre Oriental*. *Prensas de Ciencias*, Facultad de Ciencias, UNAM.

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

Hijmans, R. & van Etten, J. (2012). Geographic analysis and modeling with raster data. R Package version. 2. 1-25.

Jaramillo-Correa, J.P., Bagnoli, F., Grivet, D., Fady B., Aravanopoulos F.A., Vendramin G.G. & González-Martínez S.C. (2020) Evolutionary rate and genetic load in an emblematic Mediterranean tree following an ancient and prolonged population collapse. *Molecular Ecology, 29*, 4797– 4811.

Jaramillo-Correa, J.P., Beaulieu, J., Ledig, F.T. & Bousquet, J. (2006), Decoupled mitochondrial and chloroplast DNA population structure reveals Holocene collapse and population isolation in a threatened Mexican-endemic conifer. *Molecular Ecology*, *15*, 2787-2800.

Jin W.T., Gernandt, D.S., Wehenkel, C., Xia, X.M., Wei, X.X. & Wang, X.Q. (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*.

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

Lanner, R. M. & Van Devender, T. R. (1981) Late Pleistocene piñon pines in the Chihuahuan Desert, *Quaternary Research*, *15*, 278–290.

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

Ledig, F.T. (1986) Conservation strategies for forest gene resources. *Forest Ecology and Managenent, 14*, 77-90.

Leigh, J.W. & Bryant, D. (2015), popart: full-feature software for haplotype network construction. *Methods Ecology and Evolution*, *6*, 1110-1116.

Leitch, A. R. & Leitch, I. J. (2012) Ecological and genetic factors linked to contrasting genome

dynamics in seed plants. *New Phytologist*, *194*, 629–46.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009) 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*, 2078-2079.

Liu, J.J., Sniezko, R., Murray, M., Wang, N., Chen, H., Zamany, A., Sturrock, R.N., Savin, D., & Kegley A. (2016) Genetic Diversity and Population Structure of Whitebark Pine (*Pinus albicaulis* Engelm.) in Western North America. PLOS ONE 11(12): e0167986.

Loera, I., Ickert-Bond, S. M. & Sosa, V. (2017) Pleistocene refugia in the Chihuahuan Desert: the phylogeographic and demographic history of the gymnosperm *Ephedra compacta*. *Journal of Biogeography, 44*, 2706–2716.

Loera, I., Sosa, V. & Ickert-Bond S.M. (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.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M.A. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297-303.

Martiñón-Martínez, R.J., Vargas-Hernández, J., López-Upton, J., Gómez-Guerrero, A. & Vaquera-Huerta, H. (2010) Respuesta de *Pinus pinceana* Gordon a estrés por sequía y altas temperaturas. *Revista Fitotecnia Mexicana, 33*, 239-248.

Martínez-Ainsworth, N. (2013) Genética de poblaciones de *Agave stricta* Salm-Dyck, especie endémica al Valle de Tehuacán-Cuicatlán. México, Tesis de Licenciatura, Facultad de Ciencias, UNAM, México, DF, Mexico.

Mathis, V.L., Hafner, M.S. & Hafner, D.J. (2014) Evolution and phylogeography of the *Thomomys umbrinus* species complex (Rodentia: Geomyidae). *Journal of Mammalogy*, *95*, 754–771.

McKinney, G. J. Waples, R. K., Seeb, L. W. & Seeb, J. E. (2017) Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations, *Molecular Ecology Resources, 17*, 656–669.

Metcalfe, S. E. (2006) Late quaternary environments of the northern deserts and central transvolcanic belt of Mexico. *Annals of the Missouri Botanical Garden*, *93*, 258–273.

Molina-Freaner, F., Delgado, P., Piñero, D., Perez-Nasser, N. & Alvarez-Buylla, E. (2001) Do rare pines need different conservation strategies? Evidence from three Mexican species. *Canadian Journal of Botany, 79*, 131–138.

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

Morrone, J.J., Escalante, T. & Rodríguez-Tapia, G. (2017) Mexican biogeographic provinces: Map and shapefiles. *Zootaxa, 4277*, 277–279.

Neiswenter, S.A. & Riddle, B.R. (2010) Diversification of the *Perognathus avus* species group in emerging arid grasslands of western North America. *Journal Mammalogy, 91*, 348–362.

Noy-Meir, I. (1973) Desert ecosystems: Environment and producers. *Annual Review Ecology and Systematics, 4*, 25–51.

Ortega-Ramírez, J., Valiente-Banuet, A., Urrutia-Fucugauchi, J., Mortera-Gutiérrez, C.A. & Alvarado-Valdez, G. (1998) Paleoclimatic changes during the Late Pleistocene - Holocene in Laguna Babícora, near the Chihuahuan Desert, México. *Canadian Journal of Earth Sciences*, *35*, 1168-1179.

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

Otto-Bliesner, B. L., Joussaume, S., Braconnot, P., Harrison, S. P. & Abe-Ouchi, A. (2009), Modeling and Data Syntheses of Past Climates: Paleoclimate Modelling Intercomparison Project Phase II Workshop; Estes Park, Colorado, 15–19 September 2008, *Eos Transactions American Geophisical Union, 90*, 93 – 93.

Pavy, N., Namroud, M. C., Gagnon, F., Isabel, N., & Bousquet, J. (2012). The heterogeneous levels of linkage disequilibrium in white spruce genes and comparative analysis with other conifers. *Heredity*, *108* (3), 273–284.

Pavek, D.S. (1994) *Pinus cembroides*. In: Fire Effects Information System, [Online]. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory. Available: https://www.fs.fed.us/database/feis/plants/tree/pincem/all.html

Peláez, P., Ortiz-Martínez, A., Figueroa-Corona, L., Montes, J. R. & Gernandt, D. S. (2020). Population structure, diversifying selection, and local adaptation in *Pinus patula*. *American Journal of Botany*, *107*, 1555–1566.

Picard Toolkit (2018) Broad Institute http://broadinstitute.github.io/picard/

Phillips, S. J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation, *Ecography*, *31*, 161–175.

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

QGIS Development Team (2009) QGIS Geographic Information System. Open Source Geospatial Foundation. URL http://qgis.org

Quiñones-Pérez, C.Z., González-Elizondo, M. & Wehenkel, C. (2017) Ruling out genetic erosion in *Picea chihuahuana* Martínez. *New Forests, 48,* 201–215.

Rzedowski J (1978) Vegetación de México. Limusa, México, DF, Mexico

Rzedowski, J. & Calderón de Rzedowski, G. (1993). Datos sobre la dinámica de la flora fanerogámica del Valle de México, con énfasis en especies narrativas raras, en peligro de extinción y aparentemente extintas. *Acta Botanica Mexicana, 25*, 81 - 108.

Rozas, J., Ferrer-Mata, A., Sánchez-Del Barrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E. & Sánchez-Gracia, A. (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Molecular Biology and Evolution, 34*, 3299 - 3302.

Saladin, B., Leslie, A. B., Wüest, R. O., Litsios, G., Conti, E., Salamin, N., & Zimmermann, N. E. (2017) Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Ecology and Evolution, 17*, 95.

Serb, J.M., Phillips, C.A. & Iverson, J.B. (2001) Molecular phylogeny and biogeography of *Kinosternon avescens* based on complete mitochondrial control region sequences. *Molecular Phylogenetics and Evolution, 18*, 149–162.

SEMARNAT 059-2010 Modificación a la 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*.Volume Segunda sección, Secretaría de Medio Ambiente y Recursos Naturales. https://www.profepa.gob.mx/innovaportal/file/435/1/nom_059_semarnat_2010.pdf

Scheinvar, E., Gámez, N., Moreno-Letelier, A., Aguirre-Planter, E. & Eguiarte, L.E. (2020) Phylogeography of the Chihuahuan Desert: Diversification and evolution over the Pleistocene. In: Mandujano, M., Pisanty, I., Eguiarte, L. (eds) Plant Diversity and Ecology in the Chihuahuan Desert. Cuatro Ciénegas Basin: An Endangered Hyperdiverse Oasis. Springer, Cham.

Scheinvar, E., Gámez, N., Castellanos-Morales, G., Aguirre-Planter, E. & Eguiarte, L.E. (2017), Neogene and Pleistocene history of *Agave lechuguilla* in the Chihuahuan Desert. *Journal of Biogeography*, *44*, 322-334.

Sosa, V., Ruiz-Sanchez, E. & Rodriguez-Gomez, F.C. (2009), Hidden phylogeographic complexity in the Sierra Madre Oriental: the case of the Mexican tulip poppy *Hunnemannia fumariifolia* (Papaveraceae). *Journal of Biogeography, 36*, 18-27.

Stevens, K. A., Wegrzyn, J. L., Zimin, A., Puiu, D., Crepeau, M., Cardeno, C., Paul, R., Gonzalez-Ibeas, D., Koriabine, M., Holtz-Morris, A. E., Martínez-García, P. J., Sezen, U. U., Marçais, G., Jermstad, K., McGuire, P. E., Loopstra, C. A., Davis, J. M., Eckert, A., de Jong, P., … Langley, C. H. (2016). Sequence of the Sugar Pine Megagenome. *Genetics*, *204*(4), 1613–1626.

Suzan-Azpiri, H., Sanchez-Ramos, G., Martinez-Avalos, J.G., Villa-Melgarejo, S. & Franco, M. (2002) Population structure of *Pinus nelsonii* Shaw, an endemic pinyon pine in Tamaulipas, Mexico. *Forest Ecology and Management*, *165*, 193-20.

Trejo, L., Alvarado-Cárdenas, L.O., Scheinvar, E. & Eguiarte, L.E. (2016), Population genetic analysis and bioclimatic modeling in *Agave striata* in the Chihuahuan Desert indicate higher genetic variation and lower differentiation in drier and more variable environments. *American Journal of Botany*, *103*, 1020-1029.

Vásquez-Cruz, M. & Sosa, V. (2020) Assembly and origin of the flora of the Chihuahuan Desert: The case of sclerophyllous *Rosacea*e. *Journal of Biogeography, 47*, 445–459.

Van Devender, T. R. & Burgess, T. L. (1985) Late Pleistocene Woodlands in the Bolson de Mapimi: A Refugium for the Chihuahuan Desert Biota?, *Quaternary Research*, *24*, 346–353.

Vázquez-Lobo, A., Aguilar Morales, G., Arias S., Golubov, J., Hernández-Hernández T. & Mandujano M.C. (2015) Phylogeny and biogeographic history of *Astrophytum* (Cactaceae). *Systematic Botany, 40*, 1022-1030.

Villanueva-Días, J., Cerano Paredes, J., Stahle, D.W., Constante Garcíd, V., Vázquez Salem, L., Estraba Avalos, J. & Benavides-Solorio, J. (2010) Arboles longevos de México. *Revista Mexicana de Ciencias Forestales, 1*, 7-30.

Weir, B.S. & Cockerham, C.C. (1984), Estimating f-statistics for the analysis of population structure. *Evolution*, *38*, 1358-1370.

Zavala-Hurtado, J.A. & Jiménez, M. (2020) Diversity and Uniqueness at Its Best: Vegetation of the Chihuahuan Desert. In: Mandujano M., Pisanty I., Eguiarte L. (eds) Plant Diversity and Ecology in the Chihuahuan Desert. Cuatro Ciénegas Basin: An Endangered Hyperdiverse Oasis. Springer, Cham.

Biosketch

Author contributions: L.F.C., D.P. and D.S.G. formulated the project; A.M.L. analyzed the ecological niche modeling hypotheses, L.F.C. and P.P conducted the bioinformatics processing, J.W. gave guidance on bioinformatics processing; D.O.D. gave guidance on bioinformatics and demographic history processing analyses. L.E.E. gave guidance on demographic history and the biogeographic patterns in the Chihuahuan desert. L.F.C. and D.P. led the writing. All authors commented on previous versions of the manuscript read and approved the final manuscript.

Funding

Fieldwork was supported by PAPIIT-DGAPA UNAM grant IN209816 to D.S.G., DNA processing and sequencing was financed by Instituto de Ecología (D.P.). D.O.D. is supported by PAPIIT-UNAM IA200620 and a UC-MEXUS CONACYT fellowship (CN-19-29). L.F.C. was supported with CONACYT scholarship (449615).

Supplementary materials

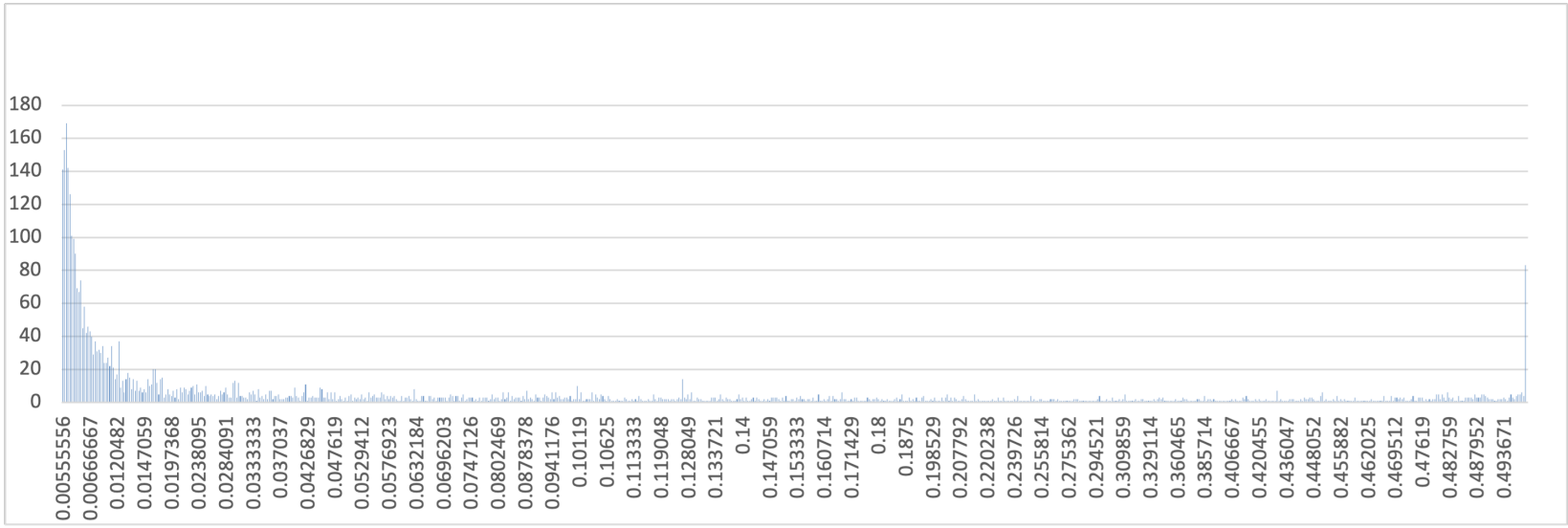

Figure S1. Site frequency spectrum for 4 141 SNPs

Table S1. F_{IS} calculation for extreme values reductions from 4 141 single nucleotide polymorphism loci

Removed	-	**≤ -0.949**	**≤ -0.899**	**≤ -0.849**	**≤ -0.799**	**≤ -0.749**	**≤ -0.699**	**≤ -0.649**	**≤ -0.59**	**≤ -0.549**	**≤ -0.499**
Average	-0.19	-0.15	-0.14	-0.13	-0.12	-0.12	-0.11	-0.1	-0.1	-0.09	-0.09
Median	-0.03	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01

To establish the impact of putative paralogs, we provide the calculation for the mean and median F_{IS} values to show the effect of reducing the extreme values

Table S2. Measures of genetic diversity for 90 individuals of *P. pinceana* from ten geographical populations calculated from 3 762 single nucleotide polymorphism loci.

	Population	H_E	H_o	F_{IS}	π	Tajima D	θ_W
SMO	Tolantongo	0.074	0.10	-0.31	0.0082	-0.42 $_{NS}$	0.10
	San Joaquín	0.079	0.10	-0.35	0.0082	-0.31 $_{NS}$	0.11
	Maguey Verde	0.099	0.14	-0.33	0.0065	-0.41 $_{NS}$	0.09
	Nuñez	0.078	0.11	-0.37	0.0088	-0.30 $_{NS}$	0.11
	La Florida	0.074	0.10	-0.35	0.0065	-0.30 $_{NS}$	0.11
ChD	El Palmito	0.079	0.10	-0.37	0.0072	-0.37 $_{NS}$	0.11
	Mazapil	0.090	0.12	-0.34	0.0072	-0.31 $_{NS}$	0.12
	La Noria	0.078	0.11	-0.35	0.0087	-0.25 $_{NS}$	0.13
	G. Cepeda	0.099	0.14	-0.36	0.0087	-0.18 $_{NS}$	0.13
	Sierra Parras	0.095	0.13	-0.38	0.0090	-0.14 $_{NS}$	0.13
Global		0.116	0.0934	-0.35	0.0066	-0.63 $_{NS}$	0.15

H_O, observed heterozygosity; H_E, expected heterozygosity within populations; F_{IS}, inbreeding coefficient; π, nucleotide diversity

Table S3. Pairwise differentiation (F_{ST}) among the ten populations sampled of *P. pinceana* with 3 762 SNPs

SMO	Tolantongo									
	San Joaquín	0.0061								
	Maguey Verde	0.0062	0.0069							
	Nuñez	0.0098	0.0173	0.0181						
	La Florida	0.0103	0.0195	0.0194	0.0104					
ChD	El Palmito	0.0451	0.0492	0.0468	0.0472	0.0487				
	Mazapil	0.0488	0.0506	0.0552	0.0545	0.0545	0.0166			
	La Noria	0.0521	0.0545	0.0586	0.0565	0.0581	0.009	0.0089		
	G. Cepeda	0.0525	0.058	0.0608	0.0603	0.0593	0.0184	0.0069	0.0043	
	Sierra Parras	0.0596	0.0588	0.0645	0.0651	0.0648	0.02	0.0053	0.0103	0.0021

Capítulo 3

The introgression truly rules a part in the diversification of large cone pinyon pines (*Pinus* subsection *Cembroides* Engelm.) from North America?

Laura Figueroa-Corona, José-Rubén Montes, Jill Wegrzyn, David S. Gernandt & Daniel Piñero

El siguiente capítulo está constituido por el artículo: The introgression truly rules a part in the diversification of large cone pinyon pines (*Pinus* subsection *Cembroides* Engelm.) from North America?, que está en proceso de elaboración y tiene el objetivo de probar la c la hipótesis planteada por Ledig et al. (2001) que sugiere que la diversidad en *P. pinceana* es resultado de la introgresión con *P. maximartinezii*. Para esto estudié la variación intraespecífica del genoma plastídico y SNPs en 858 genes nucleares de bajo número de copias, con estos dos sets de datos se reconstruyeron las relaciones filogenetica y se evaluó si existieron eventos de introgresión o migración en tres especies de pinos piñoneros de conos grandes. Los resultados obtenidos del ensamble del genoma plastídico de 18 plastomas de *Pinus pinceana, P. rzedowskii* y *P. maximartinezii* y 112,893 SNPs muestran reconstrucciones filogeneticas que coinciden con estudios evolutivos previos indicando que la relación entre los linajes de pinos piñoneros de conos grandes y reflejan la ausencia de eventos de flujo genético o reticulación entre especies. Sin embargo, en ambas reconstrucciones se obtuvo en el linaje de *P. pinceana* una divergencia que reflejan la estructura genética dentro del Desierto Chihuahuense y la Sierra Madre Oriental.

The introgression truly rules a part in the diversification of large cone pinyon pines (*Pinus* subsection *Cembroides* Engelm.) from North America?

Laura Figueroa Corona[1, 2], José-Rubén Montes[1,3], Jill Wegrzyn[4], David S. Gernandt[3] and Daniel Piñero[2]

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

[2] Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, México

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

[4] Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA

Email addresses:

LFC: laurafc25@gmail.com

JRM: ruben.montes@st.ib.unam.mx

JW: jill.wegrzyn@uconn.edu

DSG: dgernandt@ib.unam.mx

DP: pinero@unam.mx

ABSTRACT

- *Premise of the study:* We addressed whether the lineages in the three pine species experienced introgression events or migration using the intraspecific nuclear variation and plastid genome, two different set of inheritance to reconstruct the phylogenetic divergence in the large cone pinyon pines (*Pinus* subsection *Cembroides* Engelm.)
- *Methods:* We inferred the lineages relationships under coalescent-based method and Bayesian inference. We used to infer a species tree with 112,893 polymorphisms in 858 protein-coding nuclear genes and an inference with 18 assembled plastomes from

Pinus pinceana, P. rzedowskii, and *P. maximartinezii*. Finally, we performed a reticulation analysis using Dsuite applied to 112,893 nuclear biallelic SNPs across quartets.

- *Results:* Phylogenetic and coalescent results coincide with previous evolutionary studies. For the three especies of large cone pinyon pines show the same topology and the *D* test results did not detect events of gene flow or introgression within *Pinus pinceana, P. rzedowskii,* and *P. maximartinezii*. Whithin the *P. pinceana* show a lineage divided into two clades that reflects the genetic population structure within the Chihuahuan desert and the Sierra Madre Oriental regions.
- *Discussion:* Our results add evidence for divergences without introgression in the diversification inside the large cone pines. The high diversity that previously has been described as an introgession between species could be explained by the phylogeographyc pattern and isolation from two genetic groups in *P. pinceana*.

KEYWORDS: large cones pinyon pines, plastomes lineages, *Pinus pinceana*, *Pinus maximartinezii, Pinus rzedowskii.*

INTRODUCTION

Pinus subsection *Cembroides* Engelm. is a group of small to medium-sized trees and shrubs with economical and biological importance (Richardson, 2000). The North America pinyon pines have enlarged seeds with lipid-rich megagametophytes and an absent or vestigial wing. Pinyons present divergent morphology mainly in cones, seeds, and needles (Malusa, 1992). Pinyon pines have been divided into two groups, those with small cones (< 10 cm) and those with large cones (>10 cm; Malusa, 1992). The large cone group is monophyletic and comprises *Pinus maximartinezii* Rzed., *P. pinceana* Gordon, and *P. rzedowskii* Madrigal & M. Caball (Montes et al., 2019). Ortiz-Medrano et al. (2016) described a positive relationships between enlarged seeds and cones with aridity dated in the basal divergence of *P. maximartinezii* and *P. pinceana*. Phylogenetic studies suggest that wingless seeds evolved either once with a reversal to winged seeds in *P. rzedowskii*, or twice—once in the small cone clade and once in the ancestor of *P. pinceana* and *P. maximartinezii* (Gernandt et al., 2005).

A third alternative is that the MRCA of *Pinus* section *Parrya* had wingless seeds, and wings were regained in the ancestor of *Balfourianae* and in *P. rzedowskii*.

Pinus maximartinezii and *P. pinceana* are endemic to Mexico with high levels of genetic variation heterogeneously distributed among populations (Ledig et al., 2001; Molina-Freaner et al., 2001). *Pinus pinceana* is distributed in the Sierra Madre Oriental (SMO) and the Chihuahuan desert (ChD), whereas *P. maximartinezii* is confined to two Sierra Madre Occidental (SMOc) forest localities. Based on monoterpene composition (Zavarin & Snajberk, 1987), and high diversity on the cpDNA SSRs frequencies Ledig et al. (2001), proposed the hypothesis that *P. pinceana* and *P. maximartinezii* have undergone subtle events of hybridization that produce the speciation of *P. maximartinezii.*

Using chloroplast and nuclear markers (Gernandt et al., 2005; Montes et al., 2019; Willyard et al., 2007) recovered a sister group relationship between *Pinus maximartinezii* and *P. pinceana* with high statistical support. Recently, genomic approaches discovered phylogeographic patterns and detected changes in genetic diversity of *P. pinceana* during the climatic dynamic on the interglacial cycles over the Pleistocene. The best supported model involved the isolation of two lineages, one distributed in the SMO, and the other extended over the ChD (Figueroa-Corona et al., *in prep*). Here, considering that the plastome is an independent evolutionary unit, paternally inherited may carry signals of reticulation or introgression events, we hypothesize that divergence in plastid and nuclear markers will be useful to reconstruct the evolutionary changes and migration events between populations of *P. maximartinezii* and *P. pinceana* populations.

METHODS

Sampling

We sampled needles of 17 individuals from 14 locations and one haploid megagametophyte seed from the entire distribution range of *P. pinceana,* three individuals of *P. maximartinezii*, and one sample from *P. rzedowskii* as an outgroup (Table S1; Figure S1). We extracted DNA using the CTAB method (Doyle and Doyle, 1987) for diploid leaf tissue and from the Qiagen DNeasy kit protocol for the haploid tissue *P. pinceana*. DNA concentration was measured

using a Qubit fluorometer and Qubit dsDNA HS assay kit (Life Technologies, Carlsbad, California, USA) and a Nanodrop spectrophotometer (Thermo Scientific, Waltham, Massachusetts, USA).

Samples with ≥ 200 ng of DNA and an A260/A280 between 2.0 and 2.2 were sent to Arbor Biosciences (Ann Arbor, MI, USA) for preparation of genomic libraries. For target enrichment (Gnirke et al., 2009), 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). A mixture or enriched and unenriched libraries were included in the sequencing runs to obtain both low-copy nuclear genes and high-copy number plastomes (Hyb-Seq; Weitemier et al., 2014). The samples were multiplexed and spread across eight Illumina paired-end sequencing runs (Table S2) that also included samples for other studies.

Marker selection and gene selection alignment

Reads were trimmed using TRIMMOMATIC ver. 0.32 (Bolger et al., 2014), retaining sequences with an 80 bp minimum length and minimum quality score of 30. The procedure was executed with FASTQC ver. 0.11.7 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and all the samples were condensed with MULTIQC ver. 1.9 (Ewels et al., 2016).

To identify SNPs, we used a total of 858 putative single-copy nuclear genes from the genome reference of *Pinus lambertiana* ver. 1.5 (available at https://treegenesdb.org/jbrowse). All high-quality reads were mapped using BWA-MEM ver. 0.7.10 (Li et al., 2009) with default parameters. The unmapped reads were discarded.

We used the Picard package (http://broadinstitute.github.io/picard/) Mark Duplicates to correct for artifacts of PCR duplication by only keeping one read or read-pair with the highest summed base quality among those of identical external coordinates and/or same insert lengths. Finally, we used Haplotypecaller and Variant Filtration tools from GATK ver. 3.7 (McKenna et al. 2010), SNP quality by depth (QD 2.0), mapping quality (MQ 40.0), Phred-scaled *P-value* using Fisher's exact test to detect strand bias (FS >60.0), Mapping

Quality Rank Sum Test (MQRankSum < -12.5), Read Pos Rank Sum Test (ReadPosRankSum <-8.0), and Strand Odds Ratio (SOR > 4.0) processing variants.

Finally, we made a filter prior to analysis using VCFTools ver. 0.1.16 (Danecek et al., 2011) to a list of biallelic SNPs and retained a total of 112 893 SNPs with a minimum mean depth > 6, Max missing by sample < 0.3, Max missing by site < 0.05, and a MAC of 1.

To qualify the structure within the samples we used Admixture ver. 1.3.0 (Alexander & Lange 2011) to estimate the population structure among individuals with admixture proportion from $K = 3$ to 17. Genotypes were converted to ordinary PLINK files (.ped) using PLINK ver. 1.9 (Purcell et al., 2007).

Plastome assembly

We used two short-read assemblers, one reference guided assembly conducted with Geneious ver. 9.2.6 (Kearse et al., 2012) and a *de novo* short read assembly conducted with SPAdes ver. 3.6.2 (Bankevich et al., 2012) to assemble plastomes. Trimmed reads were imported and processed in Geneious, were mapped to a reference plastome (*P. californiarum*; DSG1549) previously annotated from *P. monophylla* (MH612859; Gernandt et al., 2018), generating consensus sequences. Mapped reads were imported to SPAdes ver. 3.9 (Bankevich et al., 2012) in format FASTQ. The assembly in SPAdes resulted in several scaffolds per species which were imported again to Geneious. Scaffolds (<500 bp) were eliminated, and the rest were mapped to the previously generated consensus sequences. New consensus sequences were remapped to produce the final plastome sequences (see Aguirre-Dugua & Gernandt, 2017). Heterozygous sites were manually edited based on the quality of the sequences and comparison of the sites with a reference.

We eliminated plastomes with lower coverage values (<10×) prior to the phylogenetic analysis. Plastome sequences were aligned in MAFFT ver. 7.0 (Katoh et al., 2002) implemented in Geneious ver. 9.2.6. The alignment was imported to Gblocks ver. 0.91b for further editing (Castresana, 2000). To construct a haplotype network, we generated a median

joining network using PopART ver. 1.7 with the multiple alignment of plastid sequences (118.28 kb: Leigh & Bryant, 2015).

Plastome phylogenetic analyses

Based on an alignment with the 18 individuals plastomes, nucleotid substitution models using the Akaike Information Criterion (AIC) were selected in jModelTest ver. 2.1.10 to choose the best model (Darriba et al., 2012; Miller et al., 2010) to infer a phylogenetic tree under Bayesian inference with MrBayes ver. 3.2.6 (Huelsenbeck and Ronquist 2001; Miller et al., 2012) with partition blocks for coding and non-coding regions. The analysis was conducted using the General Time Reversible model (GTR), allowing 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 40,000,000 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).

Coalescent-based analyses

A species tree was inferred under a multispecies coalescent model using singular value decomposition scores for species quartets (SVDquartets) (Chifman & Kubatko, 2014) implemented in PAUP* ver. 5.0 (Swofford, 2002). SVDquartets infers phylogenetic relationships with SNPs and accommodates incomplete lineage sorting (ILS). For this method, we used a concatenated alignment with 19 samples and 112,893 SNPs from 858 low-copy nuclear genes. We removed from the analysis the sample DSG440 source of a haploid megagametophyte to only analyze biallelic SNPs. We computed the quartet-based phylogeny under the QFM algorithm (Quartet Fiduccia and Mattheyses algorithm). A maximum of 1,000,000 randomly chosen quartets were estimated and we performed 500 replicates of bootstrap.

Introgression analysis

We used *D*suite ver. 1 (Malinsky et al., 2020) to estimate *D*-statistics (also known as the ABBA-BABA statistic) to detect gene flow or introgression between lineages dividing the populations of *P. pinceana* according to biogeographic regions. The test was applied to 112 893 biallelic SNPs across quartets (populations, taxa, species) on a typical tree topology (((P1, P2), P3), O), where O is the outgroup, *P. rzedowskii,* which served to define the ancestral allele (A), while the derived alleles were denoted by (B). Site patterns were ordered as follows: BBAA (P1 and P2 sharing the derived allele), ABBA (P2 and P3 sharing the derived allele), and BABA (P1 and P3 sharing the derived alleles). Under the null model (no gene flow), ABBA and BABA patterns were expected to occur with equal frequencies, while a significant deviation from that suggested possible introgression between P3 and either P1 or P2 (Malinsky et al., 2020). The *D* value was expected to be 0 if no gene flow had occurred and significantly different from 0 if gene flow (or introgression) existed between P2 and P3.

RESULTS

Plastome assembly statistics

The mean coverage was low in all plastomes assembled (11.7 - 36.69×). The only unusually high coverage was recorded in the unique haploid sample of *P. pinceana* from Hidalgo (440; 1908.5×) in the Sierra Madre Oriental. The atypical coverage value was produced by two probes for two cpDNA genes (UMN-6920 between *cemA* and *ycf4*, and 0-8850 near *psaB*) in the first version of the probes that were removed for the later versions (Table S2). In the *P. pinceana* plastome assemblies, we found a mean coverage of 18×, and a mean length of 113,975 pair bases (bp), but substantial variation in overall plastome size among samples was detected (Figure 1). Plastomes of *P. maximartinezii* had a length 103,141 bp and a mean coverage of 14×. Finally, the sample of *P. rzedowskii* had a length of 115,157 bp and a coverage of 36.69× (Figure S2; Table 1).

The overall difference of ~1,182 bp in plastome size reflected the substantial variation in a large single-copy region (LSC) between the inverted repeat region A (IRA) and inverted repeat region B (IRB). The average GC (%) content in CDS for both species was 38.6%. Gene density (total number of genes per kb) was 0.09 for *P. maximartinezii* and *P. pinceana,* and 0.097 for *P. rzedowskii*.

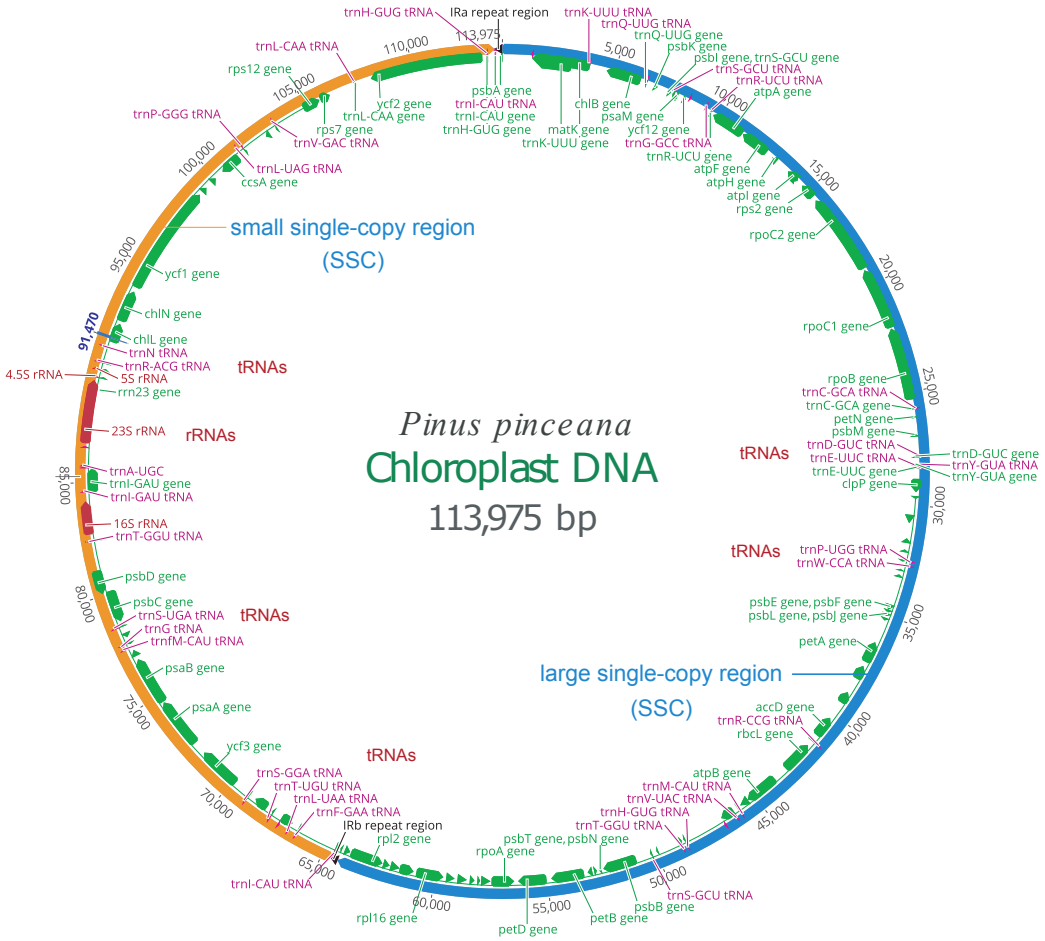

Figure 1. The plastid genome of *Pinus pinceana* (sample 440). Small and large single-copy regions (SSC, LSC) inverted repeat regions (IRA, IRB), tRNA sequences (purple), rRNA sequences (red) and gene sequences and gene clusters are shown as series in green.

Table 1. General features of the assembly of complete plastomes of large cone pinyon pines.

Species				Total length	# contigs	N50	N75	Coverage
P. maximartinezii		6499		113,141	13	13,116	5,821	11.98
		1010		114,990	7	30,308	20,725	14.73
		7799		109,503	37	4,071	2,736	7.59
P. rzedowskii		636		115,157	7	19,355	17,948	36.69
P. pinceana	**SMO**	San Joaquin 1	SaJo1	116,296	3	73,913	41,575	25.43
		San Joaquín 2	SaJo2	116,221	3	73,736	41,677	23.54
		Tolantongo	To	116,342	6	21,711	20,868	22.95
		Florida 1	Flo1	115,997	5	75,017	5,821	17.83
		Florida 2	Flo2	116,313	5	42,111	42,111	16.7
		Nuñez	Nu	116,212	7	31,814	20,424	24.38
		Maguey Verde	MV	115,470	5	41,161	41,161	35.1
		1163		112,036	21	7,550	4,696	10.8
		440		113,975	13	13,971	9,377	1,908.52
	ChD	8999		110,365	45	3,886	1,765	7.32
		Mazapil	MZ	115,878	5	41,684	41,684	22.29
		General Cepeda	GC	116,002	6	22,015	9,377	30.6
		Palmito 2	Pa2	116,313	6	59,386	20,302	19.11
		Palmito 1	Pa1	116,455	6	22,047	20,496	33.95

N50: median of lengths at 50; N75: median of lengths at 75; SMO: Sierra Madre Oriental; ChD: Chihuahuan Desert

No gene losses were detected across species. The most variable region between species was the LSC region, which had a length of 65,271 bp in *P. maximartinezii*, 65,207 bp in *P. rzedowskii* and 64,283 to 65,357 bp in *P. pinceana*. Length variation was not associated with provenance or phylogenetic relationship. Also, the *ycf1* gene presents insertions across species, ranging in length from 5,949 to 6,067 bp in *P. maximartinezii,* 5,004 to 6,217 bp in *P. pinceana* and 6,060 pb in *P. rzedowskii*. The *ycf2* region varied in length from 6,261 to 6,310 bp in *P. maximartinezii*, 4,978 to 6,294 bp in *P. pinceana* and 6,222 bp in *P. rzedowskii*. The *petB* gene length was 1,443 bp in *P. maximartinezii* and *P. rzedowskii* ranged from 1,438 to 1,443 bp in *P. pinceana* from the SMO samples and 1,443 to 1,461 bp from the ChD samples.

The haplotype network was constructed with 17 haplotypes identified from all the samples (Figure 2). Two samples from SMO locations Flo 1 and Nu share the same haplotype. Smaller distances in *P. pinceana* were observed in the SMO samples with 80 changes, similar to the distance between the closest haplotypes of *P. pinceana* and *P. maximartinezii*. A Mazapil sample (MZ) from the ChD had fewer differences with extreme south samples from the SMO than with other samples from the ChD. The greatest number of differences were found between the *P. rzedowskii* and the other haplotypes. Thus, the relationships among haplotypes correspondended to both biogeography and phylogeny.

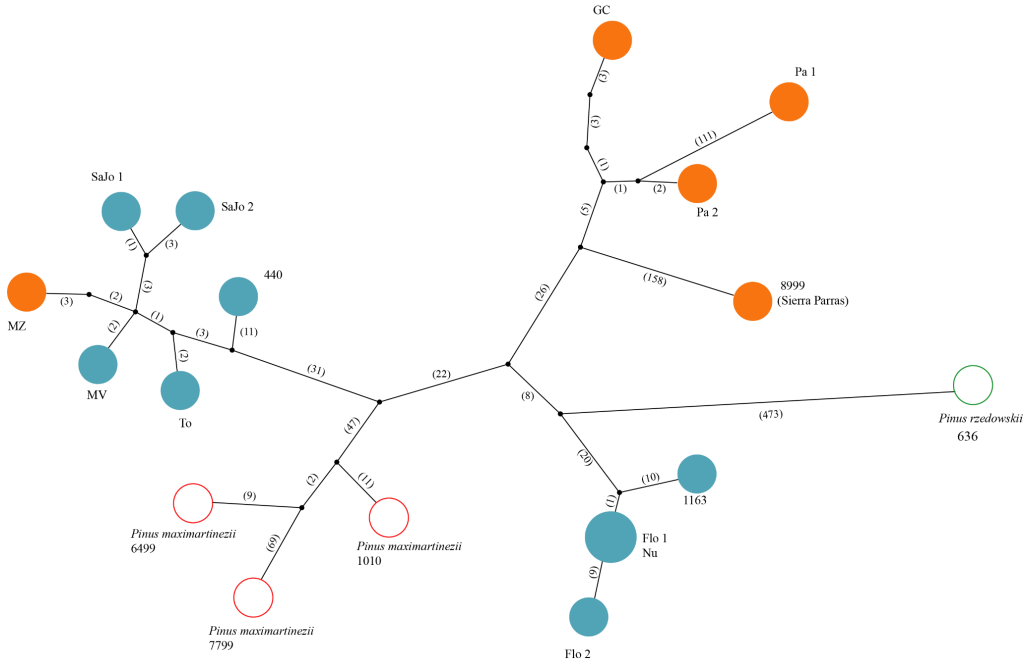

Figure 2. Haplotype network of large cone pinyon pines constructed by the complete plastome genome. The size of the dots represents the number of individuals with a given haplotype weighted by the total polymorphism shared. On blue circles shows the SMO locations on orange circles the ChD locations of *P. pinceana*, for *P. maximartinezii* the red outline circles, and the green outline circle corresponds to *P. rzedowskii*.

Phylogenetic and coalescence-based results

The plastome alignment had 18 terminals and 206 parsimony-informative characters (Figure 3). The Bayesian majority-rule tree showed that each species constitutes a well-supported (>0.9 posterior probability) monophyletic lineage. All relationships among lineages were well supported. The *P. pinceana* lineage is divided into two main clades that correspond to the biogeographic provenance of the samples, except for the Mazapil (MZ) sample. Most branches in the Bayesian-tree are short, except for *P. pinceana* PA1 and *P. maximartinezii* 7,799, that indicated few differences between taxa.

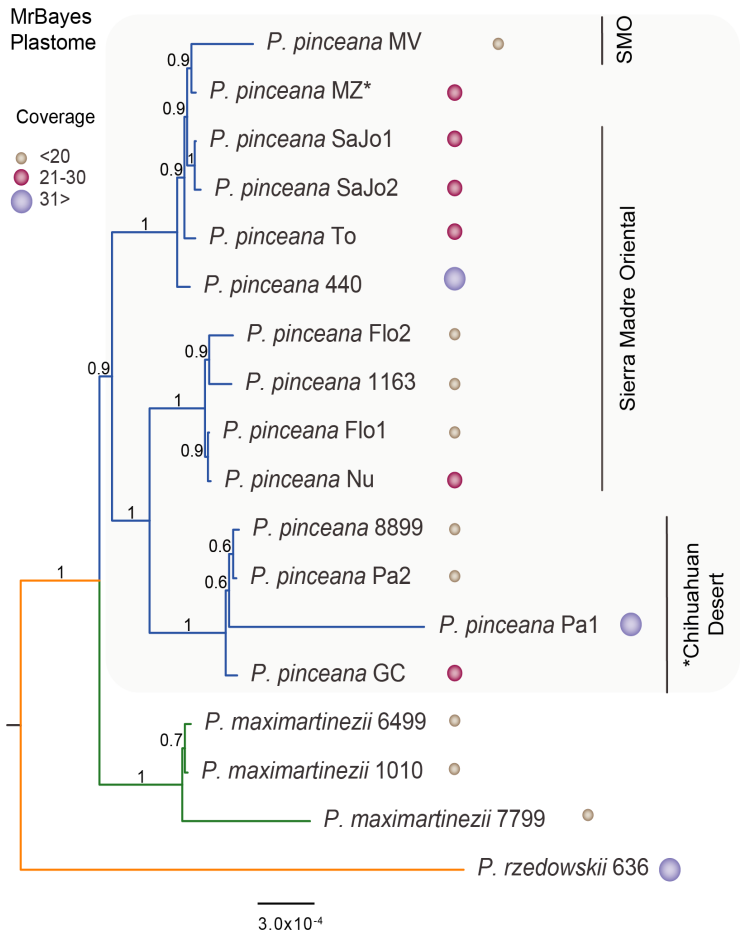

Figure 3. Consensus tree by the Bayesian inference based on 18 plastome sequences. On the right side, the sequencing coverage is shown proportional to the size of the circles. The values on branches represent the posterior probabilities.

For the nuclear reconstruction was used 12,521 parsimony informative SNPs from he 858 low copy nuclear genes. The bootstrap values for most branches in the phylogeny were up 60%. The *P. pinceana* individuals were divided into two sister clades corresponding to the two biogeographic regions, Sierra Madre Oriental and Chihuahuan desert (Figure 4). To compare both reconstruction we assemble a tanglegram with the consensus tree from the plastome Bayesian inference and the ultrametric coalescence reconstruction of SVDquartets using the default parametters in Dendroscope (Huson & Scornavacca, 2012; Figure S3).

The admixture analysis shows results for $K = 3$, the best supported number of clusters. Were only SaJo2 shows admixture with the 3 groups. The *P. maximartinezii* samples share complete similarity with *P. pinceana* GC sample and partially with SaJo samples. The remaning samples of *P. pinceana* samples have 100% similarity with two groups, without a geographical sense except for the 8899 sample, that share admixture between these two groups (Figure 4B).

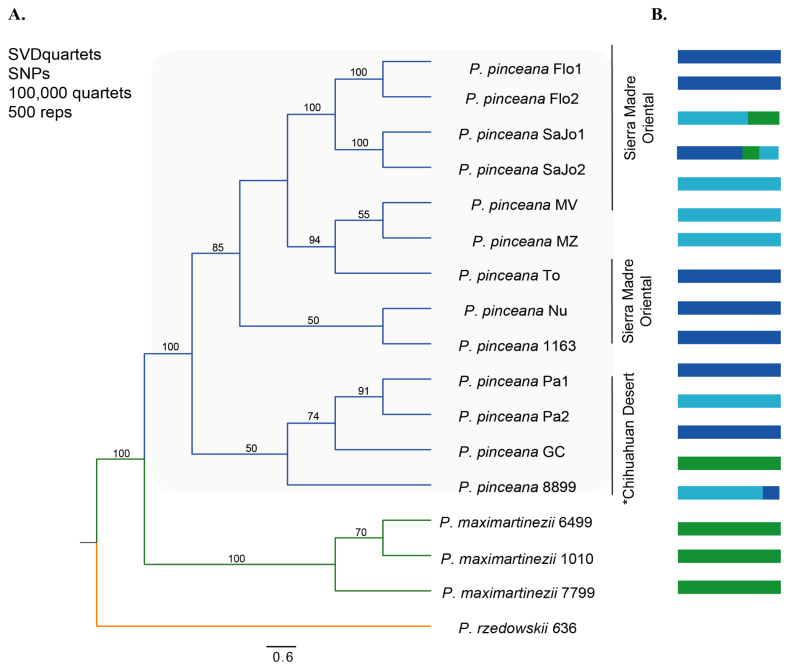

Figure 4. Ultrametric tree contructed by SVDquartets using 12,521 SNPs for nuclear genes concatenated from 858 lowcopy genes for 17 samples of large cone pinyon pines **A.** Maximum likelihood tree inferred. **B.** shows the admixture composition for the best supported clustering ($K = 3$).

Reticulation test

D-tests between lineages or populations based on 12,521 SNPs produced extremely low *D*-values (*D* = 0.00142 on average) and a non-significant *p*-value between lineages (Table 2), rejecting the hypothesis of introgression between *P. pinceana* and *P. maximartinezii*. Identical values were obtained among *P. pinceana* and *P. maximartinezii*, which can be attributed to the equal number of ancestral polymorphisms shared between the two lineages of *P. pinceana* and *P. maximartinezii*.

Table 2. *D*-statistics for large cone pinyon pine clade lineages and genetic cluster using 12 521 polymorphisms from 858 low copy nuclear genes. Using a topology (((P1, P2), P3), O), where O is the outgroup, *P. rzedowskii,* which served to define the ancestral allele (A), while the derived alleles were denoted by (B).

P1	P2	P3	*D* statistic	*p*-value
P. pinceana SMO	*P. pinceana* ChD	*P. maximartinezii*	0.00127	0.240213
P .pinceana ChD	*P. pinceana* SMO	*P. maximartinezii*	0.00127	0.240213
P. maximartinezii	*P. pinceana* SMO	*P. pinceana* ChD	0.00173	0.117542

DISCUSSION

Plastome assembly statistics

Evolutionary studies using whole plastids in pines has increased in the last decade (Parks et al., 2012; Gernandt et al., 2018; Montes et al., *in prep*). The quality of sequences, total reads, and percent of coverage in pines depends on the library preparation and the sequencing platform (Cronn et al., 2008; Parks et al., 2012; Gernandt et al., 2018) and the assembly quality depends on data and assembler methods (Abbas et al., 2014). Using the Hyb-Seq method, we recovered 18 plastomes of 3 species with a mean length of 114,664 ± 2,145.5 pb.

In contrast, the mean length for the recovered by Parks et al. (2012) using the Illumina platform was 117,157 ± 3,634 bp in *P. thunbergii*, 2,493 bp more than reportered here. Similarly, the total length of the plastome alignment after filtering was slightly shorter than previous studies (see Gernandt et al. 2018; Montes et al. *in prep*) but substantially shorter than the genus-wide alignment of Parks et al. (2012), which had a total length of 141,265 bp. Since the species studied here are closely related, the informative sites were less than reported in others previous studies of pines (15,151 infs: Parks et al., 2012; 1,833 infs: Gernandt et al., 2018; 11,281 infs: Montes et al., *in prep*). However, our results add more information at population level.

Intraspecific vs interspecific differences

The sequence of plastidic genome within the large cone pinyon pines shows low divergence; this outcome has been reported in other whole chloroplast assemblies (Whitall et al., 2010) as a result of the reported low substitution rate in *Pinus* (7.28×10^{-10} substitution per site per year; De la Torre et al., 2017), combined with rapid diversification events during the Last Miocene / Pleistocene (Montes et al., *in prep*). Nevertheless, the topology found in the nuclear and plastid is congruent between species phylogenies. The absence of a signal of hybridization in this clade of pines indicates that reticulate evolution is not a main cause in the large cone pinyon pines.

One of the results to emerge from our analysis of nearly plastidic genome is low within these pine species. The most variable region was the large single-copy region, which play a vital role in regulation of expression, and the intron or gene losses that have been found in other chloroplast genomes (Xu et al., 2020; Liu et al., 2018; Shrestha et al., 2019).

The separation of *P. pinceana* into two lineages with both nuclear and plastid DNA reflects the genetic structuring within the biogeographic regions. This result was reported in an earlier study using cpDNA SSR loci (Escalante, 2001). More recently it has been associated with demographic events during the Quaternary, when the interglacial dynamics increased aridity producing changes over the distribution of *P. pinceana* distribution (Figueroa-Corona et al., *in prep*). The Chihuahuan Desert and Sierra Madre Oriental

populations are distinct and fully differentiated in both plastomes and nuclear genomes, with topologies that are fully congruent with other phylogenetic studies of subsection *Cembroides* (Gernandt et al., 2005; Whitall et al., 2010; Montes et al., 2019). This is shown graphically as a tanglegram constructed with default parametts in Dendroscope (Huson & Scornavacca, 2012; Figure S3).

Our results cannot distinguish introgression events within the large cone pinyon pines. The admixture analysis indicates a degree of recombination within and between species, However, the support in the phylogenetic reconstruction shows a robust and clear divergence. Within the *P. pinceana* samples, the divergence between two different clades shows divergence in accordance with the biogeographical regions, with rare migration events between the two groups. The Mazapil sample (MZ) shows reticulation between the phylogenetic reconstruction between plastomes and nuclear markers, thus could by explain by two hypothesis, one, because it is one of the most genetically diverse populations, and projections along the Pleistocene indicated that this is the most stable region for *P. pinceana* constituting a Plestocene refugia along the interglacial cycles working as reservoir the most ancient genetical diversity (Figueroa-Corona et al., *in prep*), and second, could be evidence of plastome capture produced by the transit of pollen across population from the different biogeographical regions.

Given the results obtained and the high agreement with other phylogenetic reconstructions obtained with different markers, we find no evidence to support the introgression hypothesis proposed by Ledig et al., (2001), that the diversity inside *P. pinceana* was a reflection of past hybridization events with the sister species *P. maximartinezii.*

ACKNOWLEDGMENTS

We want to thank Carmen Bautista, Patricia Delgado, Celestino Flores, Socorro González, Azalea Guerra, Ulises Manzanilla, Rodolfo Salas-Lizana, Alejandra Vázquez-Lobo and Alfredo Villarruel for fieldwork assistance. To the Institute for Systems Genomics Computational Biology Core, UCONN for access to software and hardware support.

Fieldwork funding was provided by PAPIIT-DGAPA, UNAM Grant IN209816. This work constitutes a partial fulfillment of the requeriments of the Posgrado de Ciencias Biológicas at the Universidad Nacional Autónoma de México (UNAM) for JRM and LFC.

LITERATURE CITED

Abbas, M. M., Q. M. Malluhi, and P. Balakrishnan. 2014. "Assessment of de novo assemblers for draft genomes: a case study with fungal genomes". BMC genomics 15: S10.

Aguirre-Dugua, Xitlali, and David S. Gernandt. 2017. "Complete Plastomes of Three Endemic Mexican Pine Species (*Pinus* Subsection *Australes*)." Mitochondrial DNA. Part B, Resources 2 (2): 562–65.

Alexander, David H., and Kenneth Lange. 2011. "Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation." BMC Bioinformatics 12 (June): 246.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012a. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." Journal of Computational Biology: A Journal of Computational Molecular Cell Biology 19 (5): 455–77.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." Bioinformatics 30 (15): 2114–20.

Castresana, J. 2000. "Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis." Molecular Biology and Evolution 17 (4): 540–52.

Chifman, Julia, and Laura Kubatko. 2014. "Quartet Inference from SNP Data under the Coalescent Model." Bioinformatics 30 (23): 3317–24.

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.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. "The Variant Call Format and VCFtools." Bioinformatics 27 (15): 2156–58.

Darriba, Diego, Guillermo L. Taboada, Ramón Doallo, and David Posada. 2012. "jModelTest 2: More Models, New Heuristics and Parallel Computing." Nature Methods 9 (8): 772–772.

De la Torre, Amanda R., Yao-Cheng Lin, Yves Van de Peer, and Pär K. Ingvarsson. 2015. "Genome-Wide Analysis Reveals Diverged Patterns of Codon Bias, Gene Expression, and Rates of Sequence Evolution in Picea Gene Families." Genome Biology and Evolution 7 (4): 1002–15.

Doyle, J. J., and J. L. Doyle. 1987. "A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue." RESEARCH. https://worldveg.tind.io/record/33886/.

Ewels, Philip, Måns Magnusson, Sverker Lundin, and Max Käller. 2016. "MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report." Bioinformatics 32 (19): 3047–48.

Farjon, Aljos. 1996. "Biodiversity of Pinus (Pinaceae) in Mexico: Speciation and Palaeo-Endemism." Botanical Journal of the Linnean Society. Linnean Society of London 121 (4): 365–84.

———. 2010. A Handbook of the World's Conifers (2 Vols.). BRILL.

Farjon, Aljos, Brian T. Styles, and Others. 1997. "Flora Neotropica. *Pinus* (Pinaceae).". https://www.cabdirect.org/cabdirect/abstract/19980604262.

Gernandt, David S., Xitlali Aguirre Dugua, Alejandra Vázquez-Lobo, Ann Willyard, Alejandra Moreno Letelier, Jorge A. Pérez de la Rosa, Daniel Piñero, and Aaron Liston. 2018. "Multi-Locus Phylogenetics, Lineage Sorting, and Reticulation in *Pinus* Subsection *Australes*." American Journal of Botany 105 (4): 711–25.

Gernandt, David S., Gretel Geada López, Sol Ortiz García, and Aaron Liston. 2005. "Phylogeny and Classification of *Pinus*." Taxon 54 (1): 29–42.

Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, et al. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature Biotechnology 27: 182–9.

Huelsenbeck, J. P., and F. Ronquist. 2001. "MRBAYES: Bayesian Inference of Phylogenetic Trees." Bioinformatics 17 (8): 754–55.

Huson, Daniel H., and Celine Scornavacca. 2012. "Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks." Systematic Biology 61 (6): 1061–67.

Katoh, Kazutaka, Kazuharu Misawa, Kei-Ichi Kuma, and Takashi Miyata. 2002. "MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." Nucleic Acids Research 30 (14): 3059–66.

Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, et al. 2012. "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." Bioinformatics 28 (12): 1647–49.

Ledig, F. T., M. A. Capó-Arteaga, P. D. Hodgskiss, H. Sbay, C. Flores-López, M. Thompson Conkle, and B. Bermejo-Velázquez. 2001. "Genetic Diversity and the Mating System of a Rare Mexican piñon, *Pinus pinceana*, and a Comparison with *Pinus maximartinezii* (Pinaceae)." American Journal of Botany 88 (11): 1977–87.

Leigh, Jessica W., and David Bryant. 2015. "Popart: Full-feature Software for Haplotype Network Construction." Methods in Ecology and Evolution / British Ecological Society 6 (9): 1110–16.

Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform." Bioinformatics 25 (14): 1754–60.

Liu, L-X, P. Li, H-W Zhang, and J. R. P. Worth. 2018. "Whole Chloroplast Genome Sequences of the Japanese Hemlocks, *Tsuga diversifolia* and *T. sieboldii*, and Development of Chloroplast Microsatellite Markers Applicable to East Asian *Tsuga*." Journal of Forest Research 23 (5): 318–23.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297-303.

Malinsky, Milan, Michael Matschiner, and Hannes Svardal. 2021. "Dsuite - Fast D-Statistics and Related Admixture Evidence from VCF Files." Molecular Ecology Resources 21 (2): 584–95.

Malusa, James. 1992. "Phylogeny and Biogeography of the Pinyon Pines (*Pinus* Subsect. *Cembroides*)." Systematic Botany 17 (1): 42–66.

Martinon-Martinez, Rodrigo J., J. J. Vargas-Hernández, Javier López-Upton, Armando Gomez-Guerrero, Humberto Vaquera-Huerta, and Others. 2010. "Response of *Pinus pinceana* Gordon to Drought and High Temperature Stress." Revista Fitotecnia Mexicana 33 (3): 239–48.

Miller, Mark A., Wayne Pfeiffer, and Terri Schwartz. 2012. "The CIPRES Science Gateway: Enabling High-Impact Science for Phylogenetics Researchers with Limited Resources." In Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the Campus and beyond, 1–8. XSEDE '12 39. New York, NY, USA: Association for Computing Machinery.

Molina-Freaner, Francisco, Patricia Delgado, Daniel Piñero, Nidia Perez-Nasser, and Elena Alvarez-Buylla. 2001. "Do Rare Pines Need Different Conservation Strategies? Evidence from Three Mexican Species." Canadian Journal of Botany. Journal Canadien de Botanique 79 (2): 131–38.

Montes, José-Rubén, Pablo Peláez, Ann Willyard, Alejandra Moreno-Letelier, Daniel Piñero, and David S. Gernandt. 2019. "Phylogenetics of *Pinus* Subsection *Cembroides* Engelm. (Pinaceae) Inferred from Low-Copy Nuclear Gene Sequences." Systematic Botany 44 (3): 501–18.

Montes, José-Rubén, P. Peláez, A. Moreno-Letelier, and D. S. Gernandt. *In prep*. "Coalescent-based species delimitation in North American pinyon pines using low-copy nuclear genes and plastomes".

Ortiz-Medrano, Alejandra, Daniel Patrick Scantlebury, Alejandra Vázquez-Lobo, Alicia Mastretta-Yanes, and Daniel Piñero. 2016. "Morphological and Niche Divergence of Pinyon Pines." Ecology and Evolution 6 (9): 2886–96.

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.

Perry, Jesse P., Jr, and Others. 1991. The Pines of Mexico and Central America. Timber Press, Inc.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." American Journal of Human Genetics 81 (3): 559–75.

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

Rambaut, Andrew, Alexei J. Drummond, Dong Xie, Guy Baele, and Marc A. Suchard. 2018. "Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7." Systematic Biology 67 (5): 901–4.

Ramirez-Herrera, C., T. Beardmore, and J. Loo. 2008. "Overcoming Dormancy of *Pinus pinceana* Seeds." Seed Science and Technology 36 (1): 1–20.

Richardson, D. M. 2000. "Ecology and Biogeography of *Pinus*." https://books.google.ca/books?hl=en&lr=&id=YawYOzQmcHEC&oi=fnd&pg=PR13&ots=SJh9tm0zPx&sig=h6VEWtqlLC7R0xsR3cp4lPxsZeA.

Saladin, Bianca, Andrew B. Leslie, Rafael O. Wüest, Glenn Litsios, Elena Conti, Nicolas Salamin, and Niklaus E. Zimmermann. 2017. "Fossils Matter: Improved Estimates of Divergence Times in *Pinus* Reveal Older Diversification." BMC Evolutionary Biology 17 (1): 95.

Shrestha, Bikash, Mao-Lun Weng, Edward C. Theriot, Lawrence E. Gilbert, Tracey A. Ruhlman, Shawn E. Krosnick, and Robert K. Jansen. 2019. "Highly Accelerated Rates of Genomic Rearrangements and Nucleotide Substitutions in Plastid Genomes of *Passiflora* Subgenus *Decaloba*." Molecular Phylogenetics and Evolution 138 (September): 53–64.

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

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.

Wei-Tao, J., Gernandt, D. S., Wehenkel, C., Xia, X.-M., Wei, X.-X. & Wang, X.-Q. (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*(20). https://doi.org/10.1073/pnas.2022302118

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 Suppl 1 (March): 100–114.

Willyard, Ann, John Syring, David S. Gernandt, Aaron Liston, and Richard Cronn. 2007. “Fossil Calibration of Molecular Divergence Infers a Moderate Mutation Rate and Recent Radiations for *Pinus*.” Molecular Biology and Evolution 24 (1): 90–101.

Xu, Jingya, Xiaofeng Shen, Baosheng Liao, Jiang Xu, and Dianyun Hou. 2020. “Comparing and Phylogenetic Analysis Chloroplast Genome of Three *Achyranthes* Species.” Scientific Reports 10 (1): 1–13.

Zavarin, Eugene, and Karel Snajberk. 1987. “Monoterpene Differentiation in Relation to the Morphology of *Pinus culminicola, Pinus nelsonii, Pinus pinceana and Pinus maximartinezii*.” Biochemical Systematics and Ecology 15 (3): 307–12.

Supplementary material

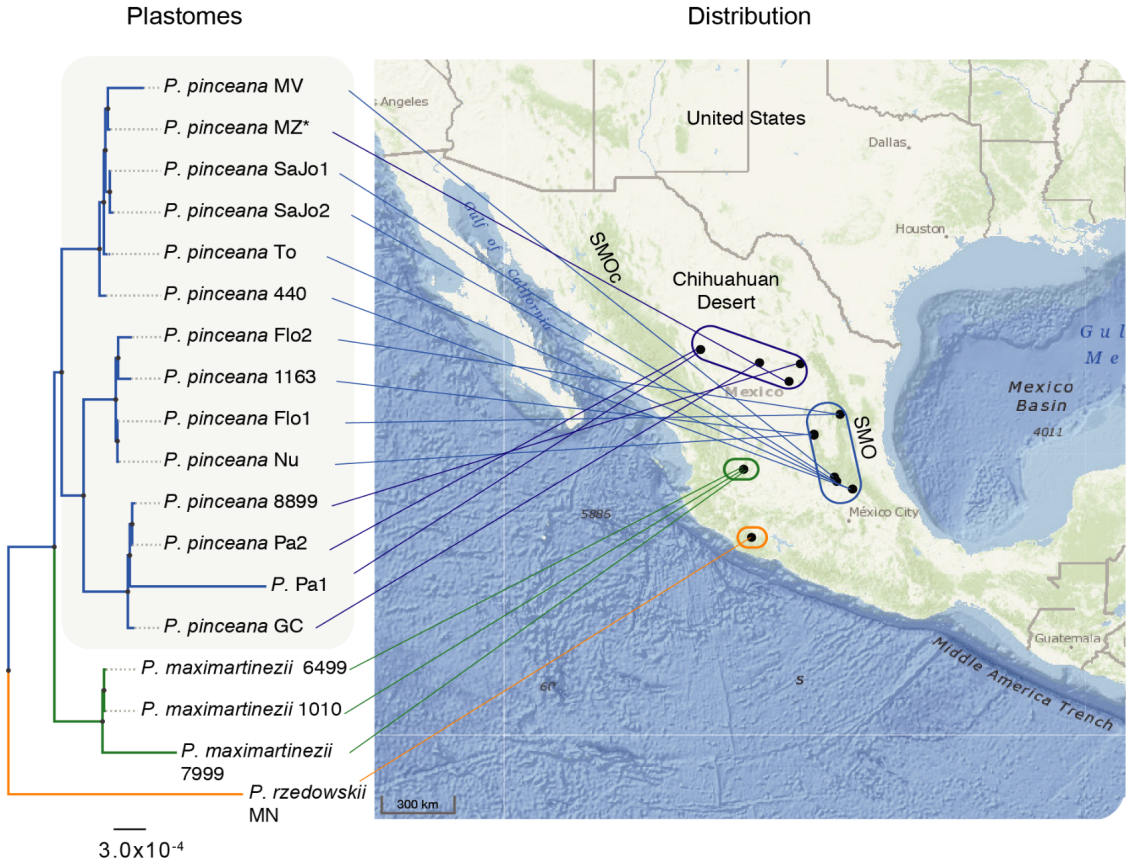

Figure S1. Geographic distribution of samples tracking the phylogenetic reconstruction with the chloroplast genome.

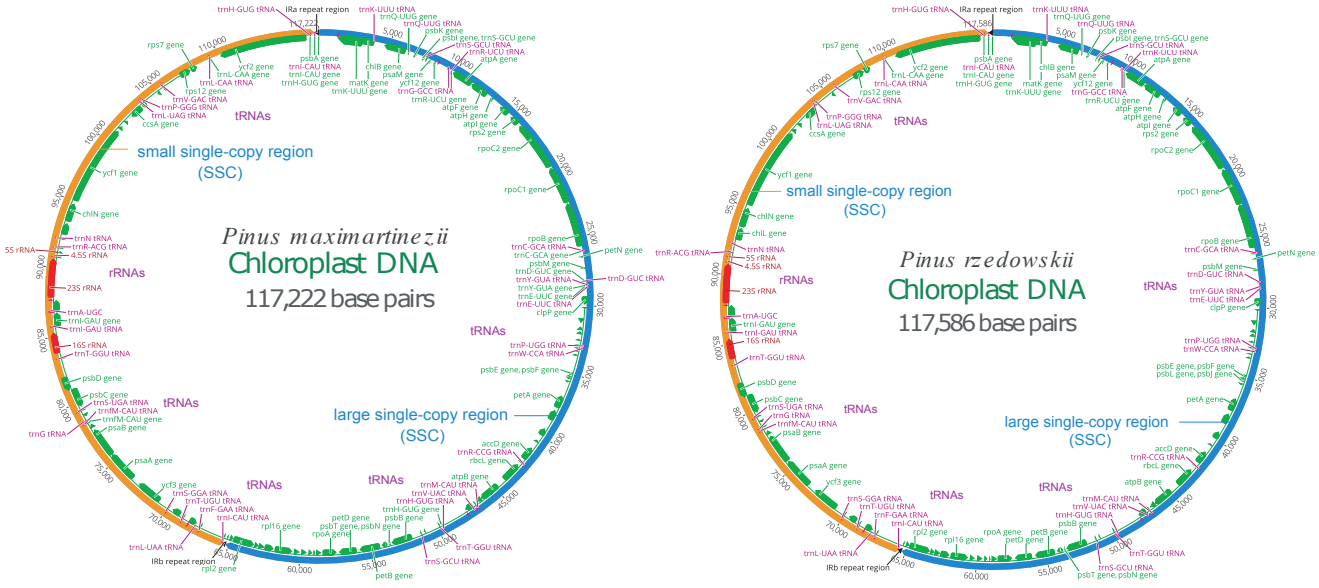

Figure S2. The complete chloroplast genome of *Pinus maximartinezii and P. rzedowskii*

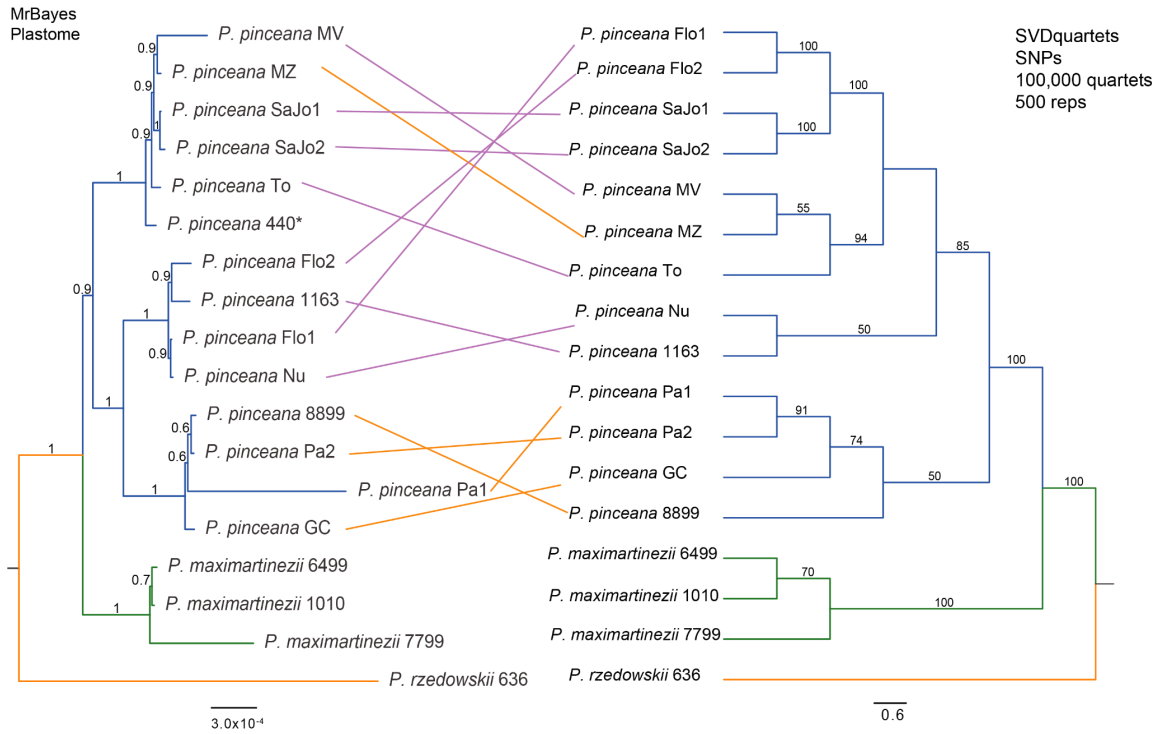

Figure S3. Tanglegram of Bayesian majority-rule consensus tree and SVDquartets lineages tree. A. Maximum likelihood tree based on concatenated alignment with 858 nuclear genes. B. Coalescent-based tree based on 1 000 000 quartets.

Table 1S. Sampled sites

Location		Condition	Latitude (N)	Longitude (W)
Tolantongo	To	SMO	20.63	98.99
Maguey Verde	MV	SMO	21.09	99.695
Nuñez	Nu	SMO	22.671	100.483
San Joaquín	SaJo	SMO	20.91	99.633
DSG 440		SMO	20.635	98.99
DSG 1163		SMO	22.618	100.488
Florida	Flo	SMO	23.4	99.503
Mazapil	MZ	ChD	24.59	101.472
DSG 8999		ChD	25.255	101.02
General Cepeda	GC	ChD	25.301	102.607
Palmito	Pal	ChD	25.741	104.881
P. maximartinezii				
DSG6499			19.493	98.848
DSG7799			21.35	103.225
DSG1010			19.317	99.195
P. rzedowskii				
DSG636			18.821	102.926

Table S2. Conditions and concentrations of the parallel sequencing per sample. Ratio of unenriched (unenriched:enriched).

Lane	Probe version	ID collector	Sample	DNA	Samples/lane	sequencing depth	Illumina Sequence Lengths
GP02	V1	DSG440	*Pinus pinceana*	Genomic, 1n from megagametophyte	48	20:80	100
GP03	V2	DSG1163	*Pinus pinceana*	Genomic, 2n from needle	48	30:70	100
GP03	V2	DSG8999	*Pinus pinceana*	Genomic, 2n from needle	48	30:70	100
GP03	V2	DSG07799	*Pinus maximartinezii*	Genomic, 2n from needle	48	30:70	100
GP03	V2	DSG1010	*Pinus maximartinezii*	Genomic, 2n from needle	48	30:70	100
GP03	V2	DSG6499	*Pinus maximartinezii*	Genomic, 2n from needle	48	30:70	100
GP03	V2	DSG636	*Pinus rzedowskii*	Genomic, 2n from needle	48	30:70	100
GP06	V2	Sierra de Parras	*Pinus pinceana*	Genomic, 2n from needle	48	40:60	125
GP06	V2	Mazapil	*Pinus pinceana*	Genomic, 2n from needle	48	40:60	125
GP06	V2	Nuñez	*Pinus pinceana*	Genomic, 2n from needle	48	40:60	125
GP06	V2	Palmito1	*Pinus pinceana*	Genomic, 2n from needle	48	40:60	125
GP06	V2	Palmito2	*Pinus pinceana*	Genomic, 2n from needle	48	40:60	125
GP06	V2	Maguey Verde	*Pinus pinceana*	Genomic, 2n from needle	48	40:60	125
GP06	V2	General Cepeda	*Pinus pinceana*	Genomic, 2n from needle	48	40:60	125
GP06	V2	Tolantongo	*Pinus pinceana*	Genomic, 2n from needle	48	40:60	125
GP07	V3	Florida 1	*Pinus pinceana*	Genomic, 2n from needle	96	40:60	150
GP07	V3	Florida 2	*Pinus pinceana*	Genomic, 2n from needle	96	40:60	150
GP07	V3	San Joaquín 1	*Pinus pinceana*	Genomic, 2n from needle	96	40:60	150
GP07	V3	San Joaquín 2	*Pinus pinceana*	Genomic, 2n from needle	96	40:60	150

Discusión general

La arquitectura genética, es decir, las contribuciones genéticas a un fenotipo son variables en tiempo y espacio, influenciadas por la mutación, la migración, la recombinación, la deriva genética y la selección (Timpson et al., 2018). La variación genética se fundamenta como el sujeto de la evolución y conforma a largo plazo un efecto en el fenotipo. Esta variación genética comprende el número de polimorfismos genéticos que influyen en el rasgo, las frecuencias de los polimorfismos en la población, la magnitud de sus efectos, además de sus interacciones entre genes y con el ambiente (Savolainen et al., 2007; Timpson et al., 2018).

Este proyecto fue planteado con tres aproximaciones genómicas en una de las coníferas más ampliamente distribuidas en condiciones de extrema aridez, *Pinus pinceana*, con el objetivo de combinar estrategias de secuenciación de alto rendimiento para estudiar tanto en regiones neutrales como en genes altamente conservados por el efecto intenso de la selección purificadora y la actividad transcripcional. Con tres diferentes aproximaciones genéticas, analicé 1) la expresión transcripcional en *P. pinceana*, y cómo responde fisiológicamente al medio ambiente, dada sus interacciones a factores bióticos y abióticos; 2) el efecto de su historia demográfica y los patrones filogeográficos en su diversidad genética; y 3) los patrones de divergencia dentro de los linajes de *P. pinceana* y sus especies hermanas.

1. Transcriptómica en *Pinus pinceana*

La aproximación funcional en este proyecto, planteanda a partir de la identificación la respuesta a la aridez por medio de los cambios en la expresión genética, la describo en el Primer Capítulo, donde planteé entender la relación fenotipo-ambiente utilizando siete individuos de localidades geográficas con distintos patrones de precipitación y temperatura en el año. El ensamble *de novo* del transcriptoma de *P. pinceana* se caracterizó entre 27,217 y 42,590 transcritos (CDS) en los individuos analizados. La varianza transcripcional entre localidades responde esencialmente a dos factores: 1) la naturaleza de los estudios de RNA-seq, como una muestra puntual, local y específica a las condiciones del organismo en el

momento de la colecta; 2) la variación entre las localidades y el diseño experimental usado no estuvieron sujetas a condiciones controladas u homogéneas, ya que se priorizó colectar individuos en toda el área de distribución sin utilizar réplicas locales que permitieran una mejor representatividad local. Consecuentemente, esta varianza transcripcional repercute en las rutas metabólicas caracterizadas, influye en los niveles de expresión cuantificados y evidencía la diversidad de relaciones ecológicas locales con organismos parásitos y endófitos.

A pesar de la alta heterogeneidad, el número de transcritos caracterizados está de acuerdo con lo reportado en estudios previos de transcriptomas de acículas de coníferas y del número estimado de genes en los genomas de coníferas. Por ejemplo, el transcriptoma de *Sequoiadendron giganteum* con 35,183 CDS; *P. cembroides* con 20,656 CDS; o en *P. sylvestris* con 24,780 CDS (Scott et al., 2020; Webster et al., *in prep.*; Ojeda et al., 2019).

La variabilidad detectada en los perfiles de expresión genética podría reflejar la respuesta a factores bióticos y abióticos, a pesar de que en el estudio no pudimos atribuir ninguna relevancia adaptativa, en parte debido a la cobertura relativamente baja en la comparación entre muestras, y a la ausencia de genes candidatos específicos. Las rutas metabólicas caracterizadas en el transcriptoma de *P. pinceana* ayudaron a identificar la respuesta transcripcional en las localidades naturales regulados por la expresión génica de rutas metabólicas de respuesta al estrés biótico o abiótico, que involucran la tolerancia al estrés e incluyen mecanismos que limitan y reparan el daño a los componentes celulares.

Las respuestas fisiológicas de respuesta al efecto ambiental y las adaptaciones a la aridez están influenciadas por mecanismos internos y factores ambientales externos y la interacción entre ambos. Las respuestas fisiológicas son variables en duración e intensidad, afecta la síntesis, concentración, transporte y almacenamiento de metabolitos primarios y secundarios (Brodribb et al., 2005). Los ajustes ocurren en múltiples vías metabólicas para restaurar los desequilibrios químicos y energéticos.

A partir de los resultados obtenidos en el Primer Capítulo, caractericé los transcritos presentes en 7 árboles de su distribución natural, y evalué el perfil de expresión génica, comparándolos entre regiones ambientales y con dos árboles fuera de rango altitudinal y de

precipitación en el área natural de distribución. Con ambas comparaciones, describí la participación de genes que actúan en protección estructural relacionados con las condiciones de aridez y estrés, como osmoprotectores, acuaporinas, chaperonas, proteínas de embriogénesis tardía y proteínas de choque térmico; factores que promueven la eliminación de especies reactivas de oxígeno (ROS); y reguladores del metabolismo de la pared celular y de la protección a las estructuras celulares. A nivel hormonal, encontré reguladores en la síntesis del ácido salicílico (SA), ácido jasmónico (JA), y genes *pr-1* ortólogos identificados previamente en brasinoesteroides, implicados en las respuestas al estrés por aridez, congelación, salinidad y osmótico (Akbudak et al., 2020; Jirage et al., 1999). Corcuera et al. (2012a, b) correlacionaron en *P. pinaster* los niveles de JA con la humedad relativa del aire y con la presencia de patógenos, puesto que también son inductores de la producción de JA en lesiones locales (Rojo et al., 2003; Koo & Howe, 2009).

La presencia de intermediarios en la regulación por ABA resulta igualmente interrelacionada con la respuesta al estrés abiótico. Para las coníferas el transporte a distancia de reguladores de ABA puede no resultar eficiente a fin de provocar el cierre de los estomas a tiempo y evitar déficits de agua en los tejidos (Brodribb et al., 2014). Perks et al. (2002) encontraron que en *P. sylvestris* que el cierre de los estomas precedió al aumento de la concentración de ABA en respuesta a un aumento en el déficit de agua del suelo. Es importante resaltar que los factores reguladores de SA encontrados ha sido descritos como inductores de la síntesis de osmoprotectores, como la prolina bajo estímulos de estrés hídrico (Umebese et al., 2009; Yang & Miao, 2010; Xiao et al., 2008, 2009). Li et al. (2014) describieron en *Torreya grandis* que la acumulación de SA y prolina inducen actividad de la cadena respiratoria, la síntesis de clorofila y la reducción de ROS, además de prevenir daño en membrana durante el estrés por salinidad.

La regulación hacia arriba de transcritos reguladores de la respuesta al estrés oxidativo que se han estudiado a partir del análisis de fenotipos tolerantes. Sánchez-Gómez et al. (2017) han descrito que la respuesta al déficit hídrico provoca una reducción de la tasa fotosintética, en la conductancia estomática y en la eficiencia de las cadenas respiratorias. De María et al. (2020) describieron en *P. pinaster* respuestas diferenciales para la tolerancia al

estrés hídrico y la regulación del estrés oxidativo con proteínas (RING-H2), que de acuerdo con los resultados del Primer Capítulo se encuentran reguladas positivamente en individuos de la región más árida del área de distribución de la especie.

La regulación hacia arriba de reguladores de la respuesta al estrés abiótico no fue exclusiva de las regiones de mayor aridez. También encontramos una mayor acumulación de transcritos que codifican para MYB4 y varias proteínas de choque térmico en SMO, factores de transcripción reportados en la tolerancia al estrés por calor, el equilibrio hídrico de las hojas y la respuesta a patógenos virales o fúngicos (Wang et al., 2019). MYB4 es un factor de transcripción que participa en la regulación de la síntesis de skimato y flavonoides, osmoprotectores, ABA, jasmonatos y brassinosteroides (Wang et al., 2020).

De manera conspicua, también encontramos transcritos relacionados directamente con la respuesta al estrés biótico. Entre las familias ortólogas génicas mejor representadas están las proteínas NBS-LRR, relacionadas con la inmunidad innata a través del reconocimiento de patógenos, que incluyen bacterias, virus, hongos, nemátodos, insectos y oomicetes (Meyers et al., 2002; Eitias et al., 2008; McHale et al., 2006). La presencia de estas proteínas concuerda con la incidencia de secuencias encontradas de insectos patógenos, de levaduras asociadas con ellos y de hongos patógenos y endófitos.

2. La diversidad genética e historia demográfica de *Pinus pinceana*

Más allá de la plasticidad y de la posible respuesta adaptativa en la regionalización ambiental actual de *P. pinceana,* exploré la relación genotipo-ambiente en el Segundo Capítulo de esta tesis, en el que se caractericé su diversidad genética en un contexto histórico.

Trabajos previos habían estudiado a *P. pinceana* usando aloenzimas y cpDNA SSRs, reportaron una alta variación genética al compararla con otras coníferas de amplia distribución. Estos estudios detectaron un patrón filogeográfico que divide a las poblaciones en dos linajes sin evidencia de migración, aparentemente aislados por la distancia entre las poblaciones (Ledig et al., 2001; Molina-Freaner et al., 2001; Escalante, 2001; Figueroa-Corona, 2012). A diferencia de estos trabajos, en esta tesis utilicé polimorfismos en el

genoma completo para la caracterización de la diversidad genética, con el objetivo de muestrear polimorfismos en regiones codificantes, reguladoras y neutrales.

Encontré una alta representación de alelos en baja frecuencia, y una alta diversidad dentro de las poblaciones en su hábitat fragmentado. La prevalencia de nuevos polimorfismos mantenidos por tamaños poblacionales efectivos relativamente grandes podría explicarse por dos escenarios diferentes: 1) la retención de nuevas variantes genéticas y, 2) la retención de parálogos y pseudogenes.

La diferenciación biogeográfica entre ChD y SMO ha sido reportado en varias especies de la comunidad biótica del Desierto Chihuahuense y, de manera similar, ha sido correlacionada con los cambios geológicos y paleoclimáticos. Otros organismos estudiados incluyen plantas como *Agave lechuguilla, A. striata,* y *A. stricta,* roedores como *Thomomys umbrinus, Perognathus avus,* y la tortuga *Kinosternon avescens* (Mathis et al., 2014; Neiswenter & Riddle, 2010; Serb et al., 2001; Scheinvar et al., 2017; Trejo et al., 2016; Martínez-Ainsworth, 2013; Scheinvar et al., 2020).

Para *Pinus pinceana,* los procesos de aislamiento y la baja abundancia de las condiciones climáticas favorables han mantenido la distribución a una zona muy reducida, por lo que queda limitado el contacto secundario y la migración. La separación de El Palmito, con una ubicación noroeste, aparentemente representa una fragmentación más reciente de las poblaciones del Desierto Chihuahuense. Este patrón de divergencia oeste/este en el Desierto Chihuahuense que también se ha descrito en organismos de matorrales xéricos como *Astrophytum* spp., *Berberis trifoliate*, *Agave victoria-reginae*, *Leucophyllum* spp. y *Ephedra compacta* (Vázquez-Lobo et al., 2015; Angulo et al., 2017; Gándara & Sosa, 2014; Loera et al., 2017), y en la serpiente *Crotalus molossus* (Anderson & Greenbaum, 2012). Esta regionalización de la diversidad genética en *P. pinceana* concuerda con sus patrones históricos en su demografía y con la dinámica climática modelada durante el Cuaternario. La separación entre sus dos linajes ocurrió durante el Pleistoceno medio durante los ciclos interglaciares (~627 kya; 95% CI: 584.52 - 633.37).

Los procesos de aridificación durante el Pleistoceno redujeron los cuerpos de agua en la región, y es posible que la cuenca de Nazas determinara la división dentro del ChD

produciendo la separación entre el noroeste y el centro. Con el incremento de la ardificación, las condiciones climáticas favorables para *P. pinceana* se volvieron más comunes en contraste con lo que se han planteado en la región para *Picea chihuahuana*, que actualmente las condiciones climáticas favorables para esta especie son escasas (Quiñonez-Pérez et al., 2017; Jaramillo-Correa et al., 2006).

De acuerdo con las proyecciones climáticas a lo largo del tiempo, la región con mayor estabilidad es Mazapil, población que en concordancia con los resultados del segundo capítulo es una de las tres poblaciones más diversas genéticamente. Así, con estas dos evidencias interpretamos que esta región ha funcionado como un refugio para *P. pinceana*. Sin embargo, esta región no ha sido descrita previamente como refugio plestocenico en otros estudios filogeográficos en el Desierto Chihuahuense (*e.g.*, Vasquez-Cruz & Sosa et al., 2016; Loera et al., 2017; Scheinvar et al., 2017; Gámez et al., 2017).

Los cambios climáticos que han afectado la distribución de *P. pinceana* han ocurrido desde los ciclos interglaciares, y han generado una expansión gradual sobre el Desierto Chihuahuense. Actualmente, las diferencias del clima entre ChD y SMO se dan por cambios en la precipitación y la temperatura. Siendo las poblaciones en ChD suelen ser mucho más secas, con un menor índice de contenido de agua, por lo que muestran una menor acumulación de agua en el suelo durante el año, en comparación con las poblaciones en SMO.

Con los resultados hasta ahora mencionados, la heterogeneidad y regionalización en *P. pinceana* se puede apreciar tanto a nivel de expresión genética, y como a nivel de polimorfismos genéticos en el genoma completo. La divergencia entre los dos linajes se debe a su historia demográfica que ha modelado el espectro de frecuencias, que evidencian esta diferenciación entre las poblaciones. Abonando a esta divergencia, estudios morfo-fisiológicos han detectado que los individuos del Desierto Chihuahuense tienen una cubierta cerosa más gruesa en las acículas, y un crecimiento más rápido en la raíz primaria y mayor volumen en raíces laterales que individuos de la Sierra Madre Oriental (Martiñón-Martínez et al., 2010; Córdoba-Rodríguez et al., 2011).

La respuesta a la aridez constituye un rasgo poligénico y combina estrategias ecofisiológicas plásticas y adaptativas (Bewley, 1979). El estudio de la adaptación y la

plasticidad conforman uno de los desafíos actuales más interesantes de la biología evolutiva (Seehausen et al., 2014). Por ello, elaborar planteamientos genómicos es importante para entender la organización y efecto en regiones funcionales, codificantes o neutrales. Los estudios de los efectos de la selección pocas veces se han detectado por la fijación de alelos únicos que contribuyan con un efecto mayor, y más frecuentemente se han descrito a través de cambios en las frecuencias de variantes relacionadas funcionalmente (Barrett & Schluter, 2008; Rockman, 2012; Berg & Coop, 2014; Wellenreuther & Hansson, 2016).

La respuesta transcripcional, conjunta y ortóloga, hacia factores bióticos y abióticos en el ambiente natural representa un reto para reconocer *per se* la tolerancia a la aridez. Sin embargo, reconocer patrones de respuesta en árboles adultos, con una longevidad por encima de los 80 años, ya establecidos y bajo las relaciones ecológicas locales naturales, me permitió establecer una base para entender la diversidad y heterogeneidad que concurre en la distribución natural de *P. pinceana*.

3. Variación intraespecífica de los pinos piñoneros de conos grandes

Las diferencias genéticas y morfológicas descritas en *P. pinceana* fueron el fundamento de una hipótesis de divergencia que señala que existe un patrón de hibridación y divergencia entre *P. pinceana* y *P. maximartinezii* (Ledig et al., 2001). Con una perspectiva filogenética, en el Tercer Capítulo analicé la divergencia en el clado ancestral de los pinos piñoneros, utilizando comparaciones entre genes altamente conservados y la reconstrucción del genoma del cloroplasto completo. De acuerdo con las reconstrucciones filogenéticas, la divergencia de *Pinus* subsección *Cembroides* ocurrió durante el Mioceno (hace 11 Ma; Gernandt et al., 2008; Saladin et al., 2017; Wei-Tao et al., 2021). La evidencia fósil de pinos piñoneros registra grandes cambios en el rango de distribución durante el Plioceno en el Altiplano Mexicano (Lanner & Van Devender, 1981; Van Devender & Burgess, 1985). Las coníferas tuvieron una distribución más amplia durante el Pleistoceno en México en el Desierto Chihuahuense (Quiñonez-Pérez et al., 2017; Jaramillo-Correa et al., 2006).

Los genomas plastídicos de las especies reportados en el Tercer Capítulo también serán útiles en futuros estudios filogenéticos, y para poder analizar finamente la

diversificación de los pinos piñoneros de la subsección *Cembroides*. Las secciones del genoma plastídico en la región IRB tendrán que compararse con las regiones correspondientes en los genomas plastídicos de otras especies dentro de la subsección.

Para *P. pinceana*, la divergencia entre las poblaciones del Desierto Chihuahuense y la Sierra Madre Oriental es clara, tanto los polimorfismos en el DNA plastídico como los genes núcleares altamente conservados que muestran una divergencia entre los individuos de estas regiones biogeográficas. A nivel de especie, las reconstrucciones filogenéticas por SNPs nucleares concatenados y el genoma plastídico mostraron concordancias con las topologías construidas con otros estudios de la subsección *Cembroides* (Gernandt et al., 2005; Whitall et al., 2010; Montes et al., 2019). Nuestros resultados no pueden distinguir entre el flujo de genes, eventos de captura de cloroplasto o los eventos de introgresión dentro de las muestras de *P. pinceana* y *P. maximartinezii.*

La divergencia entre los dos clados obtenidos tuvo un buen soporte y solo se observó para la muestra de Mazapil un evento de reticulación, que puede ser explicado a partir de las evidencias en las reconstrucciones climáticas durante el Cuaternario, que describen la zona como estable en la distribución de la especie en este intervalo de tiempo, además de ser una de las poblaciones con mayor diversidad genética.

De esta manera, la hipótesis de la introgresión propuesta por Ledig et al., (2001), de acuerdo con nuestros diferentes análisis filogenéticos ha sido descartada. Sin embargo, esta tesis demuestra que existe una variación intraespecífica en *P. pinceana,* a lo largo de su distribución natural. Estos dos grupos, como planteé en el Segundo Capítulo, se han mantenido aislados, que se mantiene por una baja migración entre ellos, y no es un reflejo de eventos de hibridación pasados con su especie hermana *P. maximartinezii*. Hasta el momento no existe aislamiento reproductivo entre los dos linajes de *P. pinceana*, y aunque el flujo genético parece ser limitado, no existe evidencia de la acción de algún mecanismo de especiación, o de un posible efecto de un proceso que potencialmente produzca un cambio adaptativo entre los dos linajes ligado a las diferentes condiciones ambientales presentes en la distribución.

Conclusiones y perspectivas generales

De acuerdo con la evidencia reportada en los cambios de nivel de expresión génica, es posible concluir que la expresión diferencial a los factores abióticos y bióticos en *P. pinceana* es muy heterogénea en su ambiente natural. Particularmente, la vía de regulación por fitohormonas ABA y JA, y la regulación de ROS fueron las vías de respuesta mejor caracterizadas para el control y regulación del estrés. Sin embargo, las interrelaciones de factores estresantes no hacen posible distinguir propiamente la respuesta exclusivamente en los cambios en la respuesta a la aridez.

Si bien hay una respuesta diferencial a los cambios en el ambiente, a nivel transcripcional parece ser un mecanismo plástico que cubre un espectro ante numerosos factores de estrés. Determinar si esta respuesta es adaptativa requerirá que realizar el análisis bajo condiciones homogéneas, junto con una comparación cruzada de condiciones, además de evaluar la heredabilidad de los fenotipos variables.

A partir de la diversidad genética, identifique dos grupos genéticos considerando regiones aleatorias del genoma, genes altamente conservados y el plastoma completo. La divergencia puede explicarse por el efecto de procesos filogeográficos históricos que han dejado huella en el patrón de diversidad, y es posible que esta diferenciación no incluya procesos adaptativos.

La diversidad genética en *P. pinceana* fue afectada por la dinámica interglaciar, que dividió a los clústers genéticos durante los cambios climáticos del Pleistoceno medio relacionados con los ciclos glaciar-interglaciar (~627 kya), seguido de la subfragmentación del Desierto Chihuahuense y de la Sierra Madre Oriental (~127.7 - ~539.2 kya), mientras que la aridificación aumentaba y los paleolagos reducían su área.

Aparentemente, la población de Mazapil tuvo una alta estabilidad de hábitat durante el Pleistoceno medio y de acuerdo con la diversidad genética identificada, interpreto que esta región representó un refugio durante los ciclos interglaciares. Esta hipótesis puede también explicar la reticulación entre las construcciones filogenéticas de los dos clústers genéticos entre las regiones biogeográficas.

Las relaciones filogenéticas a partir del plastoma y la concatenación de SNPs en genes altamente conservados recapturan reconstrucciones previas de la divergencia de los piñoneros de conos grandes. Sin embargo, esta tesis logró incorporar la divergencia de los grupos genéticos, descartando en cualquiera de ellos eventos de introgresión con *P. maximartinezii* como previamente se había sugerido.

Entre los dos clústers genéticos no existen evidencias de aislamiento reproductivo, ya fuera como una respuesta transcripcional o por diferencias genéticas debidas a la selección natural o a algún mecanismo adaptativo local, o bien, un proceso de especiación. Los grupos genéticos parecen ser resultado de la historia demográfica y de los cambios climáticos durante el Pleistoceno medio, lo que ha generado la diferenciación de las poblaciones.

Referencias

Aitken, S. N. & Hannerz, M. (2001). Genecology and Gene Resource Management Strategies for Conifer Cold Hardiness. In F. J. Bigras & S. J. Colombo (Eds.), *Conifer Cold Hardiness* (pp. 23–53). Springer Netherlands.

Akbudak, M. A., Yildiz, S. & Filiz, E. (2020). Pathogenesis related protein-1 (PR-1) genes in tomato (*Solanum lycopersicum* L.): Bioinformatics analyses and expression profiles in response to drought stress. *Genomics*, *112*(6), 4089–4099.

Anderson, C. G. & Greenbaum, E. (2012). Phylogeography of Northern Populations of the Black-Tailed Rattlesnake (*Crotalus molossus* Baird & Girard, 1853), With the Revalidation of *C. ornatus* (Hallowell, 1854). *Herpetological Monograph*, *26*(1), 19–57.

Angulo, D. F., Amarilla, L. D., Anton, A. M. & Sosa, V. (2017). Colonization in North American Arid Lands: The Journey of Agarito (*Berberis trifoliolata*) Revealed by Multilocus Molecular Data and Packrat Midden Fossil Remains. *PloS One*, *12*(2), e0168933.

Bacelar, E. L. V. A., Moutinho-Pereira, J. M., Gonçalves, B. M. C., Brito, C. V. Q., Gomes-Laranjo, J., Ferreira, H. M. F. & Correia, C. M. (2012). Water Use Strategies of Plants Under Drought Conditions. In R. Aroca (Ed.), *Plant Responses to Drought Stress: From Morphological to Molecular Features* (pp. 145–170). Springer Berlin Heidelberg.

Balderas-Hernández, V. E., Alvarado-Rodríguez, M. & Fraire-Velázquez, S. (2013). Conserved versatile master regulators in signalling pathways in response to stress in plants. *AoB Plants*, *5*, lt033.

Barrett, R. D. H. & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*(1), 38–44.

Berg, J. J. & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS Genetics*, *10*(8), e1004412.

Bewley, J. D. (1979). Physiological aspects of desiccation tolerance. *Annual Review of Plant Physiology*, *30*(1), 195–238.

Brodribb, T. J., Holbrook, N. M., Zwieniecki, M. A. & Palma, B. (2005). Leaf hydraulic capacity in ferns, conifers and angiosperms impacts on photosynthetic maxima. *The New Phytologist*, *165*(3), 839–846.

Brown, T. A. (2002). *Transcriptomes and Proteomes*. Wiley-Liss.

Cañas, R. A., Canales, J., Muñoz-Hernández, C., Granados, J. M., Ávila, C., García-Martín, M. L. & Cánovas, F. M. (2015). Understanding developmental and adaptive cues in pine through metabolite profiling and co-expression network analysis. *Journal of Experimental Botany*, *66*(11), 3113–3127.

Chen, C., Mitchell, S. E., Elshire, R. J., Buckler, E. S. & El-Kassaby, Y. A. (2013). Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. *Tree Genetics & Genomes*, *9*(6), 1537–1544.

Corcuera, L., Gil-Pelegrin, E. & Notivol, E. (2012a). Aridity promotes differences in proline and phytohormone levels in *Pinus pinaster* populations from contrasting environments. *Trees*, *26*(3), 799–808.

Corcuera, L., Gil-Pelegrín, E. & Notivol, E. (2012b). Differences in hydraulic architecture between mesic and xeric *Pinus pinaster* populations at the seedling stage. *Tree Physiology*, *32*(12), 1442–1457.

Córdoba-Rodríguez, D., Vargas-Hernández, J. J., López-Upton, J. & Muñoz-Orozco, A. (2011). Crecimiento de la raíz en plantas jóvenes de *Pinus pinceana* Gordon en respuesta a la humedad del suelo. *Agrociencia*, *45*(4), 493–506.

Creelman, R. A. & Mullet, J. E. (1995). Jasmonic acid distribution and action in plants: regulation during development and response to biotic and abiotic stress. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(10), 4114–4119.

Cuenca, A., Escalante, A. E. & Piñero, D. (2003). Long-distance colonization, isolation by distance, and historical demography in a relictual Mexican pinyon pine (*Pinus nelsonii* Shaw) as revealed by paternally inherited genetic markers (cpSSRs). *Molecular Ecology*, *12*(8), 2087–2097.

Davies, P. J. (2010). The Plant Hormones: Their Nature, Occurrence, and Functions. In P. J. Davies (Ed.), *Plant Hormones: Biosynthesis, Signal Transduction, Action!* (pp. 1–15). Springer Netherlands.

De la Torre, A. R., Li, Z., Van de Peer, Y. & Ingvarsson, P. K. (2017). Contrasting Rates of Molecular Evolution and Patterns of Selection among Gymnosperms and Flowering Plants. *Molecular Biology and Evolution*, *34*(6), 1363–1377.

De la Torre, A. R., Lin, Y.-C., Van de Peer, Y. & Ingvarsson, P. K. (2015). Genome-wide analysis reveals diverged patterns of codon bias, gene expression, and rates of sequence evolution in picea gene families. *Genome Biology and Evolution*, *7*(4), 1002–1015.

De la Torre, A. R., Roberts, D. R. & Aitken, S. N. (2014). Genome-wide admixture and ecological niche modelling reveal the maintenance of species boundaries despite long history of interspecific gene flow. *Molecular Ecology*, *23*(8), 2046–2059.

de María, N., Guevara, M. Á., Perdiguero, P., Vélez, M. D., Cabezas, J. A., López-Hinojosa, M., Li, Z., Díaz, L. M., Pizarro, A., Mancha, J. A., Sterck, L., Sánchez-Gómez, D., Miguel, C., Collada, C., Díaz-Sala, M. C. & Cervera, M. T. (2020). Molecular study of drought response in the Mediterranean conifer *Pinus pinaster* Ait.: Differential transcriptomic profiling reveals constitutive water deficit-independent drought tolerance mechanisms. *Ecology and Evolution*, *10*(18), 9788–9807.

De Micco, V. & Aronne, G. (2012). Morpho-Anatomical Traits for Plant Adaptation to Drought. In R. Aroca (Ed.), *Plant Responses to Drought Stress: From Morphological to Molecular Features* (pp. 37–61). Springer Berlin Heidelberg.

Delgado, P., Piñero, D., Chaos, A., Pérez-Nasser, N. & Alvarez-Buylla, E. R. (1999). High population differentiation and genetic variation in the endangered Mexican pine *Pinus rzedowskii* (Pinaceae). *American Journal of Botany*, *86*(5), 669–676.

Depaulis, F., Mousset, S. & Veuille, M. (2003). Power of neutrality tests to detect bottlenecks and hitchhiking. *Journal of Molecular Evolution*, *57 Suppl 1*, S190–S200.

Depaulis, F., Orlando, L. & Hänni, C. (2009). Using classical population genetics tools with heterochroneous data: time matters! *PloS One*, *4*(5), e5541.

Díaz-Sala, C. *et al.* The uniqueness of conifers. in *From Plant Genomics to Plant Biotechnology* (2013).

Eckert, A. J., Bower, A. D., González-Martínez, S. C., Wegrzyn, J. L., Coop, G. & Neale, D. B. (2010). Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology*, *19*(17), 3789–3805.

Eckert, A. J., van Heerwaarden, J., Wegrzyn, J. L., Nelson, C. D., Ross-Ibarra, J., González-Martínez, S. C. & Neale, D. B. (2010). Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics*, *185*(3), 969–982.

Eckert, R., Voigt, C. & Kien, T. T. (2009). Developing guidelines for energy and climate efficient urban structures a new planning instrument for adapting Ho Chi Minh City to the impacts of climate change. *Fifth Urban Research Symposium*, 28–30.

Eitas, T. K., Nimchuk, Z. L. & Dangl, J. L. (2008). Arabidopsis TAO1 is a TIR-NB-LRR protein that contributes to disease resistance induced by the *Pseudomonas syringae* effector AvrB. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(17), 6475–6480.

Escalante Hernández, Ana Elena. (2001). Estructura genética de poblaciones de *Pinus pinceana* usando como marcadores moleculares microsatélites de cloroplasto (cpSSR's). (Tesis de Licenciatura). Universidad Nacional Autónoma de México, México.

Farjon, A. (1996). Biodiversity of Pinus (Pinaceae) in Mexico: speciation and palaeo-endemism. *Botanical Journal of the Linnean Society. Linnean Society of London*, *121*(4), 365–384.

Farjon, A. (2010). *A Handbook of the World's Conifers* (2 Vols.). BRILL.

Farjon, A. & Styles, B. T. (1997). *Pinus (Pinaceae)*. Organization for Flora Neotropica.

Farjon, A., Styles, B. T. & Others. (1997). Flora Neotropica. Pinus (Pinaceae). *Flora Neotropica. Pinus (Pinaceae)*. https://www.cabdirect.org/cabdirect/abstract/19980604262

Farooq, M., Hussain, M., Wahid, A. & Siddique, K. H. M. (2012). Drought Stress in Plants: An Overview. In R. Aroca (Ed.), *Plant Responses to Drought Stress: From Morphological to Molecular Features* (pp. 1–33). Springer Berlin Heidelberg.

Farooq, M., Wahid, A., Kobayashi, N., Fujita, D. & Basra, S. M. A. (2009). Plant Drought Stress: Effects, Mechanisms and Management. In E. Lichtfouse, M. Navarrete, P. Debaeke, S. Véronique & C. Alberola (Eds.), *Sustainable Agriculture* (pp. 153–188). Springer Netherlands.

Fay, J. C. & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, *155*(3), 1405–1413.

Figueroa-Corona, L., Delgado Valerio, P., Wegrzyn, J. & Piñero, D. (2021). Transcriptome of weeping pinyon pine, *Pinus pinceana*, shows differences across heterogeneous habitats. *Trees* . https://doi.org/10.1007/s00468-021-02125-8

Figueroa Corona, Laura Alicia (2012). Genética de la conservación en pinos piñoneros de las subsecciones Cembroides y Nelsoniae de México. (Tesis de Licenciatura). Universidad Nacional Autónoma de México, México.

Figueroa Corona, Laura Alicia (2016). Identificación de Adaptación local en *Pinus pinceana*. (Tesis de Maestría). Universidad Michoacana de San Nicolas de Hidalgo, México.

Flexas, J. & Medrano, H. (2002). Drought-inhibition of Photosynthesis in C3 Plants: Stomatal and Non-stomatal Limitations Revisited. *Annals of Botany*, *89*(2), 183–189.

Gámez, N., Nihei, S. S., Scheinvar, E. & Morrone, J. J. (2017). A temporally dynamic approach for cladistic biogeography and the processes underlying the biogeographic patterns of North American deserts. *Journal of Zoological Systematics and Evolutionary Research = Zeitschrift Fur Zoologische Systematik Und Evolutionsforschung*, *55*(1), 11–18.

Gándara, E. & Sosa, V. (2014). Spatio-temporal evolution of *Leucophyllum pringlei* and allies (Scrophulariaceae): a group endemic to North American xeric regions. *Molecular Phylogenetics and Evolution*, *76*, 93–101.

Garnier-Géré, P. H. & Ades, P. K. (2001). Environmental surrogates for predicting and conserving adaptive genetic variability in tree species. *Conservation Biology: The Journal of the Society for Conservation Biology*, *15*(6), 1632–1644.

Gernandt, D. S., Aguirre Dugua, X., Vázquez-Lobo, A., Willyard, A., Moreno Letelier, A., Pérez de la Rosa, J. A., Piñero, D. & Liston, A. (2018). Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection Australes. *American Journal of Botany*, *105*(4), 711–725.

Gernandt, D. S., López, G. G., García, S. O. & Liston, A. (2005). Phylogeny and classification of *Pinus*. *Taxon*, *54*(1), 29–42.

Giles Pérez, Gustavo Ibrahim. (2017). Estructura genética y divergencia de dos especies de oyamel (Abies, Pinaceae) en la faja volcánica transmexicana. (Tesis de Maestría). Universidad Nacional Autónoma de México, México.

Hand, B. K., Lowe, W. H., Kovach, R. P., Muhlfeld, C. C. & Luikart, G. (2015). Landscape community genomics: understanding eco-evolutionary processes in complex environments. *Trends in Ecology & Evolution*, *30*(3), 161–168.

Hernández Leal, Mariana Susana. (2009). Variación genética de *Taxodium mucronatum* en las cuencas de México. (Tesis de Licenciatura). Universidad Nacional Autónoma de México, México.

Hernandez, I., Cela, J., Alegre, L.& Munn-Bosch, S. (2012) Antioxidant defenses against drought stress. Plant Responses to Drought Stress: From Morphological to Molecular Features.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A. & Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, *188*(4), 379–397.

Hochberg, U., Rockwell, F. E., Holbrook, N. M. & Cochard, H. (2018). Iso/Anisohydry: A Plant–Environment Interaction Rather Than a Simple Hydraulic Trait. *Trends in Plant Science*, *23*(2), 112–120.

Holliday, J. A., Ritland, K. & Aitken, S. N. (2010). Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *The New Phytologist*, *188*(2), 501–514.

Huang, G.-T., Ma, S.-L., Bai, L.-P., Zhang, L., Ma, H., Jia, P., Liu, J., Zhong, M. & Guo, Z.-F. (2012). Signal transduction during cold, salt, and drought stresses in plants. *Molecular Biology Reports*, *39*(2), 969–987.

Hsu, C.-Y., Wu, C.-S. & Chaw, S.-M. (2016). Birth of Four Chimeric Plastid Gene Clusters in Japanese Umbrella Pine. *Genome Biology and Evolution*, *8*(6), 1776–1784.

Jaramillo-Correa, J. P., Beaulieu, J. & Bousquet, J. (2004). Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Molecular Ecology*, *13*(9), 2735–2747.

Jardón-Barbolla, L., Delgado-Valerio, P., Geada-López, G., Vázquez-Lobo, A. & Piñero, D. (2011). Phylogeography of *Pinus* subsection *Australes* in the Caribbean Basin. *Annals of Botany*, *107*(2), 229–241.

Jirage, D., Tootle, T. L., Reuber, T. L., Frost, L. N., Feys, B. J., Parker, J. E., Ausubel, F. M. & Glazebrook, J. (1999). Arabidopsis thaliana PAD4 encodes a lipase-like gene that is important for salicylic acid signaling. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(23), 13583–13588.

Johri, P., Charlesworth, B., Howell, E. K., Lynch, M. & Jensen, J. D. (2021). Revisiting the notion of deleterious sweeps. *Genetics*. https://doi.org/10.1093/genetics/iyab094

Jones, M. R. & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, *25*(1), 185–202.

Karhu, A., Vogl, C., Moran, G. F., Bell, J. C. & Savolainen, O. (2006). Analysis of microsatellite variation in *Pinus radiata* reveals effects of genetic drift but no recent bottlenecks. *Journal of Evolutionary Biology*, *19*(1), 167–175.

Kelley, J. L., Madeoy, J., Calhoun, J. C., Swanson, W. & Akey, J. M. (2006). Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Research*, *16*(8), 980–989.

Kiani, S. P., Talia, P., Maury, P., Grieu, P., Heinz, R., Perrault, A., Nishinakamasu, V., Hopp, E., Gentzbittel, L., Paniego, N. & Sarrafi, A. (2007). Genetic analysis of plant water status and osmotic adjustment in recombinant inbred lines of sunflower under two water treatments. *Plant Science: An International Journal of Experimental Plant Biology*, *172*(4), 773–787.

Koo, A. J. K. & Howe, G. A. (2009). The wound hormone jasmonate. *Phytochemistry*, *70*(13-14), 1571-1580.

Kramer, P. J. & Boyer, J. S. (1995) Water Relations of Plants and Soils. *Water Relations of Plants and Soil*.

Kuparinen, A., Savolainen, O. & Schurr, F. M. (2010). Increased mortality can promote evolutionary adaptation of forest trees to climate change. *Forest Ecology and Management*, *259*(5), 1003–1008.

Lanner, R. M. & Van Devender, T. R. (1981). Late Pleistocene piñon pines in the Chihuahuan Desert. *Quaternary Research*, *15*(3), 278–290.

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

Lawlor, D. W. & Cornic, G. (2002). Photosynthetic carbon assimilation and associated metabolism in relation to water deficits in higher plants. *Plant, Cell & Environment*, *25*(2), 275–294.

Le, T.-N. & McQueen-Mason, S. J. (2006). Desiccation-tolerant plants in dry environments. *Re/views in Environmental Science and Biotechnology 5*(2-3), 269–279.

Ledig, F. T., Capó-Arteaga, M. A., Hodgskiss, P. D., Sbay, H., Flores-López, C., Thompson Conkle, M. & Bermejo-Velázquez, B. (2001). Genetic diversity and the mating system of a rare Mexican piñon, *Pinus pinceana*, and a comparison with *Pinus maximartinezii* (Pinaceae). *American Journal of Botany*, *88*(11), 1977–1987.

Leigh, E. G., Jr. (2007). Neutral theory: a historical perspective. *Journal of Evolutionary Biology*, *20*(6), 2075–2091.

Leitch, A. R. & Leitch, I. J. (2012). Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *The New Phytologist*, *194*(3), 629–646.

Li, Tingting, Yuanyuan Hu, Xuhua Du, Hui Tang, Chaohua Shen, & Jiasheng Wu. (2014) Salicylic Acid Alleviates the Adverse Effects of Salt Stress in *Torreya grandis* Cv. Merrillii Seedlings by Activating Photosynthesis and Enhancing Antioxidant Systems." PloS One 9 (10): e109492.

Li, Yu-Lan, Jun Li, Nai-Li Wang, & Xin-Sheng Yao. (2008) Flavonoids and a New Polyacetylene from Bidens Parviflora Willd." Molecules 13 (8): 1931–41.

Linnen, C. R. & Hoekstra, H. E. (2009). Measuring natural selection on genotypes and phenotypes in the wild. *Cold Spring Harbor Symposia on Quantitative Biology*, *74*, 155–168.

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

Loera, I., Ickert-Bond, S. M. & Sosa, V. (2017). Pleistocene refugia in the Chihuahuan Desert: the phylogeographic and demographic history of the gymnosperm *Ephedra compacta*. *Journal of Biogeography*, *44*(12), 2706–2716.

Lorenz, W. W., Sun, F., Liang, C., Kolychev, D., Wang, H., Zhao, X., Cordonnier-Pratt, M.-M., Pratt, L. H. & Dean, J. F. D. (2006). Water stress-responsive genes in loblolly pine (*Pinus taeda*) roots identified by analyses of expressed sequence tag libraries. *Tree Physiology*, *26*(1), 1–16.

Lorenz, W. W., Alba, R., Yu, Y.-S., Bordeaux, J. M., Simões, M. & Dean, J. F. D. (2011). Microarray analysis and scale-free gene networks identify candidate regulators in drought-stressed roots of loblolly pine (*P. taeda* L.). *BMC Genomics*, *12*, 264.

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

Luna-Cavazos, M., Romero-Manzanares, A. & García-Moya, E. (2010). Factors affecting the floristic relationships among Mexican pinyon woodlands. *Phytocoenologia*, *40*(4), 255–266.

MacKay, J. & Dean, J. (2011). Transcriptomics. 10.1201/b11075-9.

Mackay, J., Dean, J. F. D., Plomion, C., Peterson, D. G., Cánovas, F. M., Pavy, N., Ingvarsson, P. K., Savolainen, O., Guevara, M. Á., Fluch, S., Vinceti, B., Abarca, D., Díaz-Sala, C. & Cervera, M.-T. (2012). Towards decoding the conifer giga-genome. *Plant Molecular Biology*, *80*(6), 555–569.

Malusa, J. (1992). Phylogeny and Biogeography of the Pinyon Pines (*Pinus* Subsect. *Cembroides*). *Systematic Botany*, *17*(1), 42–66.

Martínez Ainsworth, Natalia Elena. (2013). Genética de poblaciones de *Agave stricta* salm-dyck, especie endémica al Valle de Tehuacán-Cuicatlán, México. (Tesis de Licenciatura). Universidad Nacional Autónoma de México, México.

Martiñón-Martínez, R. J., Vargas-Hernández, J., López-Upton, J., Gómez-Guerrero, A. & Vaquera-Huerta, H. (2010). Respuesta de *Pinus pinceana* Gordon a estrés por sequía y altas temperaturas. *Revista Fitotecnia Mexicana*, *33*(3), 239–248.

Mathis, V. L., Hafner, M. S. & Hafner, D. J. (2014). Evolution and phylogeography of the *Thomomys umbrinus* species complex (Rodentia: Geomyidae). *Journal of Mammalogy*, *95*(4), 754–771.

McDowell, N., Pockman, W. T., Allen, C. D., Breshears, D. D., Cobb, N., Kolb, T., Plaut, J., Sperry, J., West, A., Williams, D. G. & Yepez, E. A. (2008). Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? *The New Phytologist*, *178*(4), 719–739.

McHale, L., Tan, X., Koehl, P. & Michelmore, R. W. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biology*, *7*(4), 212.

Méndez-González, I. D., Jardón-Barbolla, L. & Jaramillo-Correa, J. P. (2017). Differential landscape effects on the fine-scale genetic structure of populations of a montane conifer from central Mexico. *Tree Genetics & Genomes*, *13*(1), 30.

Meyers, B. C., Morgante, M. & Michelmore, R. W. (2002). TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. *The Plant Journal: For Cell and Molecular Biology*, *32*(1), 77–92.

Mitton, J. B. (1989). Physiological and demographic variation associated with allozyme variation. In D. E. Soltis, P. S. Soltis & T. R. Dudley (Eds.), *Isozymes in Plant Biology* (pp. 127–145). Springer Netherlands.

Molina-Freaner, F., Delgado, P., Piñero, D., Perez-Nasser, N. & Alvarez-Buylla, E. (2001). Do rare pines need different conservation strategies? Evidence from three Mexican species. *Canadian Journal of Botany. Journal Canadien de Botanique*, *79*(2), 131–138.

Montes, J. R., Peláez, P., Willyard, A., Moreno-Letelier, A., Piñero, D. & Gernandt, D. S. (2019). Phylogenetics of *Pinus* Subsection *Cembroides* Engelm. (Pinaceae) Inferred from Low-Copy Nuclear Gene Sequences. *Systematic Botany*, *44*(3), 501–518.

Moran, E., Lauder, J., Musser, C., Stathos, A. & Shu, M. (2017). The genetics of drought tolerance in conifers. *The New Phytologist*, *216*(4), 1034–1048.

Moran Lauter, A. N., Muszynski, M. G., Huffman, R. D. & Scott, M. P. (2017). A Pectin Methylesterase ZmPme3 Is Expressed in Gametophyte factor1-s (Ga1-s) Silks and Maps to that Locus in Maize (*Zea mays* L.). *Frontiers in Plant Science*, *8*, 1926.

Moreno-Letelier, A., Mastretta-Yanes, A. & Barraclough, T. G. (2014). Late Miocene lineage divergence and ecological differentiation of rare endemic *Juniperus blancoi*: clues for the diversification of North American conifers. *The New Phytologist*, *203*(1), 335–347.

Moreno-Letelier, A. & Piñero, D. (2009). Phylogeographic structure ofPinus strobiformisEngelm. across the Chihuahuan Desert filter-barrier. *Journal of Biogeography*, *36*(1), 121–131.

Morgenstern, E.K. (1996). Geographic variation in Forest Trees. Univ. British Columbia Press. Vancouver, B.C. Canada.

Mosca, E., Gugerli, F., Eckert, A. J. & Neale, D. B. (2016). Signatures of natural selection on *Pinus cembra* and *P. mugo* along elevational gradients in the Alps. *Tree Genetics & Genomes*, *12*(1). https://doi.org/10.1007/s11295-015-0964-9

Neale D.B. & Sederoff R.R. (1989) Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theoretical and Applied Genetics*, 77, 212– 216.

Neale, D. B., McGuire, P. E., Wheeler, N. C., Stevens, K. A., Crepeau, M. W., Cardeno, C., Zimin, A. V., Puiu, D., Pertea, G. M., Sezen, U. U., Casola, C., Koralewski, T. E., Paul, R., Gonzalez-Ibeas, D., Zaman, S., Cronn, R., Yandell, M., Holt, C., Langley, C. H., … Wegrzyn, J. L. (2017). The Douglas-Fir Genome Sequence Reveals Specialization of the Photosynthetic Apparatus in Pinaceae. *G3 Genes, Genomes, Genetics*, *7*(9), 3157–3167.

Neale, D. B. & Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends in Plant Science*, *9*(7), 325–330.

Neale, D. B., Wegrzyn, J. L., Stevens, K. A., Zimin, A. V., Puiu, D., Crepeau, M. W., Cardeno, C., Koriabine, M., Holtz-Morris, A. E., Liechty, J. D., Martínez-García, P. J., Vasquez-Gross, H. A., Lin, B. Y., Zieve, J. J., Dougherty, W. M., Fuentes-Soriano, S., Wu, L.-S., Gilbert, D., Marçais, G., … Langley, C. H. (2014). Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology*, *15*(3), R59.

Nei, M., Maruyama, T. & Chakraborty, R. (1975). The Bottleneck Effect And Genetic Variability In Populations. *Evolution; International Journal of Organic Evolution*, *29*(1), 1–10.

Neiswenter, S. A. & Riddle, B. R. (2010). Diversification of the *Perognathus flavus* species group in emerging arid grasslands of western North America. *Journal of Mammalogy*, *91*(2), 348–362.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, *39*, 197–218.

Nielsen, R. & Slatkin, M. (2013). An introduction to population genetics. *Sunderland, MA: Sinauer Associates*. https://he.palgrave.com/resources/Product-Page-Downloads/N/Nielsen-Population-Genetics/NielsenSlatkin-TOC.pdf

Nosil, P., Funk, D. J. & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*(3), 375–402.

Ojeda, D. I., Mattila, T. M., Ruttink, T., Kujala, S. T., Kärkkäinen, K., Verta, J.-P. & Pyhäjärvi, T. (2019). Utilization of Tissue Ploidy Level Variation in de Novo Transcriptome Assembly of *Pinus sylvestris*. *G3: Genes, Genomes, Genetics*, *9*(10), 3409–3421.

Oney-Birol, S., Fitz-Gibbon, S., Chen, J.-M., Gugger, P. F. & Sork, V. L. (2018). Assessment of shared alleles in drought-associated candidate genes among southern California white oak species (*Quercus* sect. *Quercus*). *BMC Genetics*, *19*(1), 88.

Ortiz-Medrano, A., Scantlebury, D. P., Vázquez-Lobo, A., Mastretta-Yanes, A. & Piñero, D. (2016). Morphological and niche divergence of pinyon pines. *Ecology and Evolution*, *6*(9), 2886–2896.

Ozkur, O., Ozdemir, F., Bor, M. & Turkan, I. (2009). Physiochemical and antioxidant responses of the perennial xerophyte *Capparis ovata* Desf. to drought. *Environmental and Experimental Botany*, *66*(3), 487–492.

Paiva, J. A. P., Garnier-Géré, P. H., Rodrigues, J. C., Alves, A., Santos, S., Graça, J., Le Provost, G., Chaumeil, P., Da Silva-Perez, D., Bosc, A., Fevereiro, P. & Plomion, C. (2008). Plasticity of maritime pine (*Pinus pinaster*) wood-forming tissues during a growing season. *The New Phytologist*, *179*(4), 1180–1194.

Pavy, N., Pelgas, B., Laroche, J., Rigault, P., Isabel, N. & Bousquet, J. (2012). A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. *BMC Biology*, *10*, 84.

Perks, M. P., Irvine, J. & Grace, J. (2002). Canopy stomatal conductance and xylem sap abscisic acid (ABA) in mature Scots pine during a gradually imposed drought. *Tree Physiology*, *22*(12), 877–883.

Perry, J. P., Jr & Others. (1991). *The pines of Mexico and central America*. Timber Press, Inc.

Perry Jr, J.P., Graham, A. Richardson, D. M.(2000)The history of pines in Mexico and Central America. Ecology and biogeography of *Pinus*:137.

Pittermann, J., Stuart, S. A., Dawson, T. E. & Moreau, A. (2012). Cenozoic climate change shaped the evolutionary ecophysiology of the Cupressaceae conifers. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(24), 9647–9652.

Plomion, C., Bousquet, J. & Kole, C. (2011). *Genetics, Genomics and Breeding of Conifers*. CRC Press.

Pompa-García, M., González-Cásares, M., Acosta-Hernández, A. C., Camarero, J. J. & Rodríguez-Catón, M. (2017). Drought Influence over Radial Growth of Mexican Conifers Inhabiting Mesic and Xeric Sites. *Forests, Trees and Livelihoods*, *8*(5), 175.

Porth, I. & El-Kassaby, Y. A. (2014). Assessment of the Genetic Diversity in Forest Tree Populations Using Molecular Markers. *Diversity*, *6*(2), 283–295.

Porth, I., Hamberger, B., White, R. & Ritland, K. (2011). Defense mechanisms against herbivory in Picea: sequence evolution and expression regulation of gene family members in the phenylpropanoid pathway. *BMC Genomics*, *12*, 608.

Prunier, J., Laroche, J., Beaulieu, J. & Bousquet, J. (2011). Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology*, *20*(8), 1702–1716.

Przeworski, M. (2002). The signature of positive selection at randomly chosen loci. *Genetics*, *160*(3), 1179–1189.

Quiñones-Pérez, C. Z., Sáenz-Romero, C. & Wehenkel, C. (2014). Genetic diversity and conservation of Picea chihuahuana Martínez: A review. *African Journal of Biotechnology*, *13*(28). https://www.ajol.info/index.php/ajb/article/view/121094

Ramirez-Herrera, C., Beardmore, T. & Loo, J. (2008). Overcoming dormancy of *Pinus pinceana* seeds. *Seed Science and Technology*, *36*(1), 1–20.

Rebolledo Camacho, V., Jardón Barbolla, L., Ramírez Morillo, I., Vázquez-Lobo, A., Piñero, D. & Delgado, P. (2018). Genetic variation and dispersal patterns in three varieties of *Pinus caribaea* (Pinaceae) in the Caribbean Basin. *Plant Ecology and Evolution*, *151*(1), 61–76.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M. & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370.

Reyes Galindo V. Genómica de poblaciones de *Juniperus blancoi* y *J. monticola* : diversidad genética y variación altitudinal (Tesis de Licenciatura) Universidad Nacional Autónoma de México.

Richardson, D. M. (2000). *Ecology and biogeography of Pinus*. https://books.google.ca/books?hl=en&lr=&id=YawYOzQmcHEC&oi=fnd&pg=PR13&ots=SJh9tm0zPx&sig=h6VEWtqlLC7R0xsR3cp4lPxsZeA

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution; International Journal of Organic Evolution*, *66*(1), 1–17.

Rojo, E., Solano, R. & Sánchez-Serrano, J. J. (2003). Interactions between signaling compounds involved in plant defense. *Journal of Plant Growth Regulation*, *22*(1), 82–98.

Romero, I. G., Ruvinsky, I. & Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews. Genetics*, *13*(7), 505–516.

Rout, N. P. & Shaw, B. P. (2001). Salt tolerance in aquatic macrophytes: possible involvement of the antioxidative enzymes. *Plant Science: An International Journal of Experimental Plant Biology*, *160*(3), 415–423.

Saladin, B., Leslie, A. B., Wüest, R. O., Litsios, G., Conti, E., Salamin, N. & Zimmermann, N. E. (2017). Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evolutionary Biology*, *17*(1), 95.

Sánchez-Gómez, D., Mancha, J. A. & Cervera, M. T. (2017). Inter-genotypic differences in drought tolerance of maritime pine are modified by elevated [CO2]. *Annals of*. https://academic.oup.com/aob/article-abstract/120/4/591/3916731

Savolainen, O., Pyhäjärvi, T. & Knürr, T. (2007). Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics*. https://doi.org/10.1146/annurev.ecolsys.38.091206.095646

Scandalios, J. G. (2005). Oxidative stress: molecular perception and transduction of signals triggering antioxidant gene defenses. *Brazilian Journal of Medical and Biological Research = Revista Brasileira de Pesquisas Medicas E Biologicas / Sociedade Brasileira de Biofisica... [et Al.]*, *38*(7), 995–1014.

Scheinvar, E., Gámez, N., Castellanos-Morales, G., Aguirre-Planter, E. & Eguiarte, L. E. (2017). Neogene and Pleistocene history of Agave lechuguilla in the Chihuahuan Desert. *Journal of Biogeography*, *44*(2), 322–334.

Scheinvar, E., Gámez, N., Moreno-Letelier, A., Aguirre-Planter, E. & Eguiarte, L. E. (2020). Phylogeography of the Chihuahuan Desert: Diversification and Evolution Over the Pleistocene. In M. C. Mandujano, I. Pisanty & L. E. Eguiarte (Eds.), *Plant Diversity and Ecology in the Chihuahuan Desert: Emphasis on the Cuatro Ciénegas Basin* (pp. 19–44). Springer International Publishing.

Schlichting, C. D. & Levin, D. A. (1986). Phenotypic plasticity: an evolving plant character. *Biological Journal of the Linnean Society. Linnean Society of London*, *29*(1), 37–47.

Schlötterer, C., Tobler, R., Kofler, R. & Nolte, V. (2014). Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nature Reviews. Genetics*, *15*(11), 749–763.

Scott, A. D., Zimin, A. V., Puiu, D., Workman, R., Britton, M., Zaman, S., Caballero, M., Read, A. C., Bogdanove, A. J., Burns, E., Wegrzyn, J., Timp, W., Salzberg, S. L. & Neale, D. B. (2020a). The giant sequoia genome and proliferation of disease resistance genes.

Scott, A. D., Zimin, A. V., Puiu, D., Workman, R., Britton, M., Zaman, S., Caballero, M., Read, A. C., Bogdanove, A. J., Burns, E., Wegrzyn, J., Timp, W., Salzberg, S. L. & Neale, D. B. (2020b). A Reference Genome Sequence for Giant Sequoia. *G3*, *10*(11), 3907–3919.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, A., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., & Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews. Genetics*, *15*(3), 176–192.

Serb, J. M., Phillips, C. A. & Iverson, J. B. (2001). Molecular phylogeny and biogeography of Kinosternon flavescens based on complete mitochondrial control region sequences. *Molecular Phylogenetics and Evolution*, *18*(1), 149–162.

Sun Y, Pri-Tal O, Michaeli D & Mosquna A (2020) Evolution of Abscisic Acid Signaling Module and Its Perception. Front. Plant Sci. 11:934.

Taiz, L., Zeiger, E., Møller, I. M., Murphy, A. & Others. (2015). *Plant physiology and development*. Sinauer Associates Incorporated.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595.

Tardieu, F. & Simonneau, T. (1998). Variability among species of stomatal control under fluctuating soil water status and evaporative demand: modelling isohydric and anisohydric behaviours. *Journal of Experimental Botany*, *49*, 419–432.

Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J. & Richards, J. B. (2018). Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nature Reviews. Genetics*, *19*(2), 110–124.

Trejo, L., Alvarado-Cárdenas, L. O., Scheinvar, E. & Eguiarte, L. E. (2016). Population genetic analysis and bioclimatic modeling in *Agave striata* in the Chihuahuan Desert indicate higher genetic variation and lower differentiation in drier and more variable environments. *American Journal of Botany*, *103*(6), 1020–1029.

Umebese, C. E., Olatimilehin, T. O. & Ogunsusi, T. A. (2009). *Salicylic acid protects nitrate reductase activity, growth and proline in amaranth and tomato plants during water deficit*. https://ir.unilag.edu.ng/handle/123456789/5815

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S. M., Schlichting, C. D. & Van Tienderen, P. H. (1995). Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution*, *10*(5), 212–217.

Vázquez-Lobo, A., Morales, G. A., Arias, S., Golubov, J., Hernández-Hernández, T. & Mandujano, M. C. (2016). Phylogeny and Biogeographic History of Astrophytum (Cactaceae). *Systematic Botany*, *40*(4), 1022–1030.

Van Devender, T. R. & Burgess, T. L. (1985). Late Pleistocene woodlands in the Bolson de Mapimi: A refugium for the Chihuahuan Desert Biota? *Quaternary Research*, *24*(3), 346–353.

Vásquez-Cruz, M. & Sosa, V. (2020). Assembly and origin of the flora of the Chihuahuan Desert: The case of sclerophyllous Rosaceae. *Journal of Biogeography*, *47*(2), 445–459.

Vilagrosa, A., Chirino, E., Peguero-Pina, J. J., Barigah, T. S., Cochard, H. & Gil-Pelegrín, E. (2012). Xylem Cavitation and Embolism in Plants Living in Water-Limited Ecosystems. In R. Aroca (Ed.), *Plant Responses to Drought Stress: From Morphological to Molecular Features* (pp. 63–109). Springer Berlin Heidelberg.

Wan, T., Liu, Z.-M., Li, L.-F., Leitch, A. R., Leitch, I. J., Lohaus, R., Liu, Z.-J., Xin, H.-P., Gong, Y.-B., Liu, Y., Wang, W.-C., Chen, L.-Y., Yang, Y., Kelly, L. J., Yang, J., Huang, J.-L., Li, Z., Liu, P., Zhang, L.,

& Wang, X.-M. (2018). A genome for gnetophytes and early evolution of seed plants. *Nature Plants*, *4*(2), 82–89.

Wang, X.-C., Wu, J., Guan, M.-L., Zhao, C.-H., Geng, P. & Zhao, Q. (2020). Arabidopsis MYB4 plays dual roles in flavonoid biosynthesis. *The Plant Journal: For Cell and Molecular Biology*, *101*(3), 637–652.

Warren, R. L., Keeling, C. I., Yuen, M. M. S., Raymond, A., Taylor, G. A., Vandervalk, B. P., Mohamadi, H., Paulino, D., Chiu, R., Jackman, S. D., Robertson, G., Yang, C., Boyle, B., Hoffmann, M., Weigel, D., Nelson, D. R., Ritland, C., Isabel, N., Jaquish, B., … Bohlmann, J. (2015). Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *The Plant Journal: For Cell and Molecular Biology*, *83*(2), 189–212.

Watkinson, J. I., Sioson, A. A., Vasquez-Robinet, C., Shukla, M., Kumar, D., Ellis, M., Heath, L. S., Ramakrishnan, N., Chevone, B., Watson, L. T., van Zyl, L., Egertsdotter, U., Sederoff, R. R. & Grene, R. (2003). Photosynthetic acclimation is reflected in specific patterns of gene expression in drought-stressed loblolly pine. *Plant Physiology*, *133*(4), 1702–1716.

Wellenreuther, M. & Hansson, B. (2016). Detecting Polygenic Evolution: Problems, Pitfalls, and Promises. *Trends in Genetics: TIG*, *32*(3), 155–164.

Wegrzyn, J. L., Liechty, J. D., Stevens, K. A., Wu, L.-S., Loopstra, C. A., Vasquez-Gross, H. A., Dougherty, W. M., Lin, B. Y., Zieve, J. J., Martínez-García, P. J., Holt, C., Yandell, M., Zimin, A. V., Yorke, J. A., Crepeau, M. W., Puiu, D., Salzberg, S. L., Dejong, P. J., Mockaitis, K., … Neale, D. B. (2014). Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics*, *196*(3), 891–909.

Whittall, J. B., Syring, J., Parks, M., Buenrostro, J., Dick, C., Liston, A. & Cronn, R. (2010). Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology*, *19 Suppl 1*, 100–114.

Willyard, A., Syring, J., Gernandt, D. S., Liston, A. & Cronn, R. (2007). Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for Pinus. *Molecular Biology and Evolution*, *24*(1), 90–101.

Wu, C.S., Lin C.P., Hsu C.Y., Wang R.J. & Chaw S.M. (2011). Comparative chloroplast genomes of Pinaceae: insights into the mechanism of diversified genomic organizations. *Genome Biol Evol*. 3:309–319.

Wu, C.S., Wang, Y.N., Hsu C.Y., Lin C.P. & Chaw S.M. (2011). Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Genome Biol Evol*. 3:1284–1295.

Xiao, X., Xu, X. & Yang, F. (2008). Adaptive responses to progressive drought stress in two *Populus cathayana* populations. *Silva Fennica*, *42*(5). https://doi.org/10.14214/sf.224

Xiao, X., Yang, F., Zhang, S., Korpelainen, H. & Li, C. (2009). Physiological and proteomic responses of two contrasting *Populus cathayana* populations to drought stress. *Physiologia Plantarum*, *136*(2), 150–168.

Xing, W. & Rajashekar, C. B. (1999). Alleviation of water stress in beans by exogenous glycine betaine. *Plant Science: An International Journal of Experimental Plant Biology*, *148*(2), 185–192.

Yang, F. & Miao, L.-F. (2010). Adaptive responses to progressive drought stress in two poplar species originating from different altitudes. *Silva Fennica*, *44*(1). https://doi.org/10.14214/sf.160

Zhang, L., Liu, L., Zhao, H., Jiang, Z. & Cai, J. (2020). Differences in Near Isohydric and Anisohydric Behavior of Contrasting Poplar Hybrids (I-101 (*Populus alba* L.) × 84K (*Populus alba* L. × *Populus glandulosa* Uyeki)) under Drought-Rehydration Treatments. *Forests, Trees and Livelihoods*, *11*(4), 402.

Zhu, J.-K. (2002). Salt and drought stress signal transduction in plants. *Annual Review of Plant Biology*, *53*, 247–273.

Zimin, A., Stevens, K. A., Crepeau, M. W., Holtz-Morris, A., Koriabine, M., Marçais, G., Puiu, D., Roberts, M., Wegrzyn, J. L., de Jong, P. J., Neale, D. B., Salzberg, S. L., Yorke, J. A. & Langley, C. H. (2014). Sequencing and assembly of the 22-gb loblolly pine genome. *Genetics*, *196*(3), 875–890.