UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

FILOGENÓMICA DE *Mammillaria* (CACTACEAE): IMPLICACIONES EN SU DELIMITACIÓN Y CLASIFICACIÓN TAXONÓMICA

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

DELIL ANDREA CHINCOYA MARTÍNEZ

TUTORA PRINCIPAL DE TESIS:
DRA. SOFÍA SOLÓRZANO LUJANO
FACULTAD DE ESTUDIOS SUPERIORES IZTACALA, UNAM
COMITÉ TUTOR:
DR. FELIPE VACA PANIAGUA
FACULTAD DE ESTUDIOS SUPERIORES IZTACALA, UNAM
DR. FIDEL ALEJANDRO SÁNCHEZ FLORES
INSTITUTO DE BIOTECNOLOGÍA, UNAM

LOS REYES IZTACALA, TLALNEPANTLA, ESTADO DE MÉXICO, OCTUBRE, 2023

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

FILOGENÓMICA DE *Mammillaria* (CACTACEAE): IMPLICACIONES EN SU DELIMITACIÓN Y CLASIFICACIÓN TAXONÓMICA

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

DELIL ANDREA CHINCOYA MARTÍNEZ

TUTORA PRINCIPAL DE TESIS:
DRA. SOFÍA SOLÓRZANO LUJANO
FACULTAD DE ESTUDIOS SUPERIORES IZTACALA, UNAM
COMITÉ TUTOR:
DR. FELIPE VACA PANIAGUA
FACULTAD DE ESTUDIOS SUPERIORES IZTACALA, UNAM
DR. FIDEL ALEJANDRO SÁNCHEZ FLORES
INSTITUTO DE BIOTECNOLOGÍA, UNAM

LOS REYES IZTACALA, TLALNEPANTLA, ESTADO DE MÉXICO, OCTUBRE, 2023

COORDINACIÓN GENERAL DE ESTUDIOS DE POSGRADO
COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS
FACULTAD DE ESTUDIOS SUPERIORES IZTACALA
OFICIO: CGEP/CPCB/FESI/0633/2023
ASUNTO: Oficio de Jurado

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

Me permito informar a usted que en la reunión ordinaria del Subcomité de Biología Evolutiva del Posgrado en Ciencias Biológicas, celebrada el día **19 de junio de 2023** se aprobó el siguiente jurado para el examen de grado de **DOCTORA EN CIENCIAS** de la estudiante **CHINCOYA MARTÍNEZ DELIL ANDREA** con número de cuenta **310003715** con la tesis titulada **"Filogenómica de *Mammillaria* (Cactaceae): implicaciones en su delimitación y clasificación taxonómica",** realizada bajo la dirección de la **DRA. SOFÍA SOLÓRZANO LUJANO,** quedando integrado de la siguiente manera:

Presidente:	DRA. PATRICIA DOLORES DÁVILA ARANDA
Vocal:	DR. ALBERTO KEN OYAMA NAKAGAWA
Vocal:	DR. ÁNGEL SALVADOR ARIAS MONTES
Vocal:	DRA. TERESA MARGARITA TERRAZAS SALGADO
Secretario:	DR. FELIPE VACA PANIAGUA

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 17 de agosto de 2023

COORDINADOR DEL PROGRAMA

POSGRADO EN CIENCIAS BIOLOGICAS
COORDINACIÓN

DR. ADOLFO GERARDO NAVARRO SIGÜENZA

c. c. p. Expediente del alumno

AGNS/GGM/EARR/ggm

AGRADECIMIENTOS INSTITUCIONALES

Al Posgrado en Ciencias Biológicas, UNAM por haberme proporcionado las herramientas para mi desarrollo académico.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo económico que hizo posible dedicarme a tiempo completo a mis estudios de Posgrado CVU 855061.

Investigación realizada gracias al Programa UNAM-PAPIIT IN228619 e IN207523. Dra. Sofía Solórzano Lujano, responsable de estos proyectos.

A mi Tutora principal, la Dra. Sofía Solórzano Lujano y a mi Comité Tutor, conformado por el Dr. Felipe Vaca Paniagua y el Dr. Fidel Alejandro Sánchez Flores, quienes enriquecieron este proyecto de investigación y guiaron mi crecimiento académico.

AGRADECIMIENTOS A TÍTULO PERSONAL

A la Dra. Sofía Solórzano Lujano por brindarme su confianza y abrirme las puertas de su laboratorio. Doctora, le agradezco porque en incontables ocasiones su apoyo trascendió más allá del ámbito académico. Gracias por ser un ejemplo de constancia, compromiso y solidaridad.

A los miembros del Jurado, Dra. Patricia Dolores Dávila Aranda, Dr. Alberto Ken Oyama Nakagawa, Dr. Ángel Salvador Arias Montes, Dra. Teresa Margarita Terrazas Salgado y Dr. Felipe Vaca Paniagua, quienes con sus valiosos comentarios contribuyeron a mejorar la calidad de este documento.

A mi hermana, Eira, por apoyarme, escucharme y ayudarme a ver las situaciones desde una perspectiva distinta. Gracias por las bellas ilustraciones que conforman la figura 1 de esta tesis; me enorgullece y me hace feliz que tu talento y dedicación queden plasmados en este documento.

A mis padres, Angélica y Oscar, por impulsar mi crecimiento y por el empeño que pusieron en mi educación.

Al Dr. Salvador Arias Montes por proporcionarme las muestras biológicas de la mayoría de los especímenes secuenciados *de novo* en esta tesis.

A la Dra. Clara E. Díaz Velásquez y al Dr. Aldo Hugo De La Cruz Montoya por preparar y secuenciar las bibliotecas genómicas en el Laboratorio Nacional en Salud: Diagnóstico Molecular y Efecto Ambiental en Enfermedades Crónico-Degenerativas, FES Iztacala, UNAM.

Al Dr. David Prieto Torres por permitirme utilizar su equipo de cómputo para la realización de los análisis de biogeografía histórica.

A Mimi

ÍNDICE

RESUMEN

El género *Mammillaria* Haw. *sensu* Hunt registra 163 especies, por lo que es el de mayor riqueza de la familia Cactaceae. Sin embargo, los primeros estudios moleculares realizados con <10 loci de cloroplasto concluyeron que el género no es monofilético. Además, estos estudios no resolvieron las relaciones filogenéticas dentro del género, ni con otros seis géneros (*Coryphantha, Escobaria, Mammilloydia, Neolloydia, Pelecyphora* y *Ortegocactus*). Este conjunto de siete géneros constituye lo que se conoce como el clado Mammilloide. Recientemente, un estudio filogenómico analizó 93,808 pb del genoma de cloroplasto de 78 especies de cactus, 52 de ellas pertenecientes al género *Mammillaria.* Los resultados de este estudio recuperaron a los taxa de *Mammillaria* en dos clados distintos (*Mammillaria sensu stricto* y Cochemiea). El género *Cochemiea* quedaría conformado por los taxa excluidos de *Mammillaria sensu stricto* y las especies de *Neolloydia* y de *Ortegocactus*. Sin embargo, este estudio tuvo una representación taxonómica pobre de *Mammillaria*, cuyas especies estudiadas tienen un sesgo geográfico al noroeste de México. Por estos sesgos, en esta tesis se consideró que la problemática filogenética y taxonómica de *Mammillaria* aún no está resuelta y por ello el objetivo general de esta tesis fue dilucidar las relaciones filogenéticas en *Mammillaria* y respecto a otros seis géneros (*Coryphantha, Escobaria, Mammilloydia, Neolloydia, Pelecyphora* y *Ortegocactus*). Para cumplir este objetivo, se usó una perspectiva filogenómica analizando genomas de cloroplasto y de núcleo. Con el objetivo de examinar las relaciones filogenéticas, estimar los tiempos de divergencia y la distribución geográfica ancestral se analizaron 52 loci muestreados en 142 genomas de cloroplasto de 103 taxa distintos, 70 de ellos del género *Mammillaria* (capítulo 1). La filogenia obtenida permitió reconocer a un grupo monofilético dentro de *Mammillaria,* compuesto por especies de los subgéneros *Dolichothele, Krainzia, Mammillaria, Mammillopsis, Oehmea* y *Phellosperma*. La Altiplanicie Mexicana se identificó como la región en donde se originó el ancestro común para los 103 taxa estudiados hace 7 Ma y en donde ocurrieron el 53% de los eventos cladogenéticos. Esta región se postuló como el

centro de origen de *Mammillaria* y del resto del clado Mammilloide. La historia biogeográfica se identificó como un factor clave para entender los procesos de divergencia y la distribución actual de los cactus estudiados, siendo la dispersión el proceso más frecuente en los últimos 4.5 Ma. Por otra parte, los análisis con genoma nuclear iniciaron con la evaluación *in silico* de dos conjuntos de sondas universales para angiospermas, con el objetivo de examinar su utilidad en estudios filogenéticos con Cactaceae. Para ello, se procesaron los genomas completos de diez taxa (9 obtenidos del sitio del NCBI y 1 secuenciado en esta tesis (*Mammillaria huitzilopochtli*)). La evaluación se basó en los siguientes parámetros: mayor número de loci recuperados, longitud total de las secuencias, proporción de sitios parsimoniosamente informativos (PI) y la menor proporción de loci parálogos. Los resultados mostraron que el conjunto de sondas de Angiosperms353 tuvo un mejor desempeño que el conjunto Angiosperm v.1, ya que recuperó en promedio 276 loci, 123,687 pb, 4.32% de sitios PI y <3% de parálogos. Con base en estos resultados se concluyó que Angiosperms353 es una opción para realizar estudios filogenéticos en Cactaceae, por lo cual fue elegido para analizar el genoma nuclear. Los objetivos del capítulo 3, fueron analizar la discordancia filogenética entre compartimentos genómicos y entre distintos genes nucleares; y dilucidar las posibles causas de dichas discordancias. Para el cumplimiento de estos objetivos se analizaron un total de 322 loci en 47 taxa del clado Mammilloide. Los resultados permitieron identificar un alto nivel de discordancia filogenética; así como dos eventos de evolución reticulada entre las relaciones profundas de los taxa. Por lo tanto, se concluyó que, aunque los eventos de evolución reticulada son fuente de discordancia filogenética, el sorteo incompleto de linajes es el principal proceso biológico que la causa, particularmente entre especies filogenéticamente cercanas. En conclusión, los resultados de esta tesis demostraron una historia evolutiva compleja del clado Mammilloide y particularmente de *Mammillaria*. También, se obtuvo una filogenia completamente resuelta, con el mayor muestreo taxonómico de *Mammillaria*, gracias a la perspectiva filogenómica. Aún se deben profundizar los estudios filogenético-taxonómicos sumando datos morfológicos, anatómicos, biogeográficos y ecológicos.

ABSTRACT

The genus *Mammillaria* Haw. *sensu* Hunt records 163 species, thus is the genus with the highest richness of the family Cactaceae. However, the early molecular studies were carried out with <10 loci of chloroplast and concluded that this genus is not monophyletic. In addition, these studies did not resolve the phylogenetic relationships within the genus neither did with other six genera (*Coryphantha, Escobaria, Mammilloydia, Neolloydia, Pelecyphora, and Ortegocactus*). Recently, a phylogenomic study analyzed 93,808 bp of the genome of the chloroplast genome of 78 cacti species, 52 of them from *Mammillaria*. The results of this study recovered those taxa from *Mammillaria* into two distinct clades (*Mammillaria sensu stricto* and *Cochemiea*). *Cochemiea* would be composed by those excluded from *Mammillaria sensu stricto* and species from genera *Neolloydia* and *Ortegocactus.* However, this study had a poor taxonomic sampling of *Mammillaria,* whose studied species were geographically biased to northwest of Mexico. For this sampling bias, this thesis considered that phylogenetic and taxonomic issues of *Mammillaria* are still unresolved; and for this in this thesis the general objective was elucidate the phylogenetic relationships within Mammillaria and with taxa of other six genera (*Coryphantha, Escobaria, Mammilloydia, Neolloydia, Pelecyphora,* and *Ortegocactus*). To accomplish this objective, in this thesis was used a phylogenomic perspective to analyze genomes from chloroplasts and nuclei. To examine the phylogenetic relationships, to estimate the divergence times and the ancestral geographic distribution were analyzed 142 genomes of chloroplast of 103 distinct taxa and 70 of these from *Mammillaria*, each taxa sampled with 52 loci (chapter 1). In the phylogeny obtained was recognized a monophyletic group composed by taxa of *Mammillaria*, from the subgenera *Dolichothele, Krainzia, Mammillaria, Mammillopsis, Oehmea,* and *Phellosperma*. The Mexican Plateau was identified as the region where 7 mya arose the common ancestor of the 103 taxa, and where 53% of cladogenetic events occurred. In this thesis, was proposed that this region is the center of origin of the genus *Mammillaria* and the other genera of Mammilloid clade. The biogeographic history was identified as a key factor to

explain the divergence and the current geographic distribution of the cacti studied; being the dispersion the most common in the last 4.5 Ma. By other side, the analysis based on nuclear genome initiated with the test *in silico* of two sets of universal probes for angiosperms; to examine their useful in phylogenetic studies in Cactaceae. For this, the whole genomes of 10 taxa (9 downloaded from NCBI and 1 sequenced in this thesis (*Mammillaria huitzilopochtli*)) were bioinformatically processed. This test was based on: number of recovered loci, total sequence length, proportion of parsimoniously informative sites (PI), and lower proportion of paralogous loci. The results showed that Angiosperms353 was better, since it recovered in average 275.6 loci, 123,687 bp, 4.32% of PI sites, and <3% of paralogs. Based on these results it was concluded that Angiosperms353 is proper for phylogenetic studies in Cactaceae; and we choose this set to analyze the nuclear genome. The objectives of chapter 3 were to analyze phylogenetic discordances between genomic compartments, as well as between nuclear genes was evaluated, and to elucidate the causes of such discordances. To achieve these objectives a total of 47 taxa of Mammilloide clade, were sampled with 322 nuclear loci. The results identified a high level of phylogenetic discordance, as well as two events of reticulate evolution among the deep relationships of the analyzed species. It was concluded that although reticulate evolution events are a source of phylogenetic discordance, the incomplete lineage sorting is the main biological process that caused the estimated discordance, particularly between phylogenetically closest species. In conclusion, the results of this thesis showed a complex evolutionary history of Mammilloid clade, and particularly of *Mammillaria*. In this thesis was obtained a fully resolved phylogeny provided by the phylogenomic perspective and with the highest taxonomic sampling of *Mammillaria*. Still must to be deeply analyzed the phylogenetic-taxonomic studies, by integrating morphologic, anatomic, biogeographic and ecologic data.

INTRODUCCIÓN GENERAL

1. Aportaciones de la filogenética molecular en la resolución de la historia evolutiva y delimitación taxonómica en plantas

Los primeros estudios de filogenética molecular (*e. g.* Chase *et al*., 1993; Xiang *et al.*, 1998) enriquecieron el conocimiento de la columna vertebral de las relaciones evolutivas de las plantas. Durante tres décadas de desarrollo, los estudios moleculares han contribuido a desentrañar preguntas evolutivas añejas, como el "abominable misterio" que representaba para Darwin el origen y la rápida diversificación de las angiospermas durante el Cretácico (*e. g.* Ramirez-Barahona *et al.*, 2020). Una filogenia resuelta proporciona el contexto histórico para estudiar, por ejemplo, la evolución de rasgos complejos (*e. g.* Carrive *et al.*, 2020, Zhang *et al.*, 2020, Gamisch *et al.*, 2021) o los patrones espacio-temporales de la diversidad biológica actual (*e. g.* Deng *et al.*, 2018, Song *et al.* 2020). Sin embargo, algunos linajes presentan una mayor dificultad para resolver sus relaciones filogenéticas, particularmente cuando el muestreo molecular es pequeño.

La presencia de discordancia filogenética es uno de los retos para la resolución de relaciones filogenéticas; esta ocurre cuando las genealogías obtenidas del análisis de distintos loci (*i. e.* árboles de genes) no son concordantes. La discordancia puede presentarse entre distintos árboles de genes, entre los árboles de genes y el árbol de especies o entre los árboles de especies inferidos a partir de distintos compartimentos genómicos (*i. e.* cloroplasto y núcleo). Si bien, la discordancia filogenética puede deberse a errores metodológicos (Tamashiro *et al*., 2019), ésta también puede ser causada por procesos biológicos como el sorteo incompleto de linajes y los eventos de evolución reticulada. En esta tesis se examinaron estos dos procesos biológicos como fuente de discordancia en el conjunto de datos nucleares y entre los árboles de especies inferidos con cloroplasto y núcleo.

El sorteo incompleto de linajes ocurre cuando se mantienen polimorfismos ancestrales en los linajes descendientes después de los eventos de especiación. Como consecuencia,

la genealogía observada puede ser discordante con el patrón de diversificación de los taxa (Whitfield y Lockhart, 2007), es decir el gen no refleja el proceso de divergencia. Es más probable que se presente sorteo incompleto de linajes cuando los tamaños efectivos poblacionales son más grandes y cuando el tiempo, expresado en número de generaciones, entre los procesos de divergencia es menor (Pamilo y Nei, 1988). Respecto al término de eventos de evolución reticulada, éste agrupa a los procesos que causan discordancia filogenética debido a la presencia de flujo genético interespecífico (*i. e*. hibridación e introgresión). La hibridación y la introgresión son mecanismos comunes en la evolución de plantas a distintos niveles taxonómicos (*e. g.* García *et al.*, 2017, Paetzold *et a*l., 2019, Yang *et al*., 2023).

Además de la discordancia filogenética, los niveles bajos de variación molecular representan otro desafío para inferir relaciones filogenéticas. Esta situación ocurre con frecuencia en linajes de origen reciente, así como en aquellos que se originaron durante una diversificación rápida. En el caso de los linajes de origen reciente, los factores que claramente participan son el tiempo y la tasa de mutación. Estos factores pueden restringir la acumulación de suficiente variación molecular que permita disgregar a taxa cercanos (Parks *et al*., 2009), conduciendo a filogenias con politomías próximas a las terminales del árbol. En el caso de los linajes que diversificaron rápidamente, el poco tiempo que ocurre entre los eventos cladogenéticos restringe la acumulación de sustituciones (Whitfield y Lockhart, 2007). Este patrón se corresponde con filogenias con internodos demasiado cortos como para poder ser estimados con confianza.

Examinar las relaciones filogenéticas en linajes con poca variación molecular requiere aumentar el número de loci analizados. El desarrollo de la filogenómica, que consiste en el uso de datos de escala genómica para inferir relaciones evolutivas (Delsuc *et al.*, 2005, Eisen *et al*., 2003), ha sido facilitado gracias al avance y popularización de las tecnologías de secuenciación masiva. La creciente disponibilidad de datos genómicos ha mejorado notablemente el conocimiento de las relaciones filogenéticas desde niveles taxonómicos someros, como ocurre entre poblaciones de *Euphorbia balsamifera* (Villaverde *et al*, 2018) o

en especies del género *Melicope* (Rutaceae) (Paetzold *et al.*, 2019) hasta los nodos basales de angiospermas (Ruhfel *et al.*, 2014, Wickett *et al.*, 2014, Gitzendanner *et al.*, 2018). Además, el uso de este tipo de datos aplicado al análisis de la discordancia filogenética ha permitido identificar algunos de los procesos que la causan (García *et al.*, 2017, Rose *et al.*, 2021). Con ello, la discordancia deja de ser visualizada como un problema, para convertirse en una ventana al pasado que permite inferir algunos de los procesos que también forman parte de la historia evolutiva de los taxa.

La filogenómica integrada a otras disciplinas ha permitido reexaminar numerosas preguntas evolutivas en plantas. Particularmente, la integración de la filogenómica y la biogeografía histórica permite incorporar el contexto espacio-temporal al patrón de diversificación de los linajes (*e. g.* Ye *et al*., 2019, Low *et al*., 2021, Xia *et al.*, 2022). Por ejemplo, la integración de estas dos disciplinas ha permitido identificar el impacto de los eventos históricos tanto climáticos como orográficos en la configuración actual de la flora asiática. En cuanto a los eventos climáticos, algunos estudios identificaron al incremento en la intensidad de los monzones durante el Mioceno como el principal promotor de la diversificación (Low *et al*., 2021) y de la expansión de la distribución geográfica (Xia *et al.*, 2022) de la flora del este y sur de Asia. También se ha documentado la importancia de los eventos orográficos, como la elevación de la Meseta Tibetana-Qinghai (Xia *et al.*, 2022) y de las Montañas Hengduan (Ye e*t al*., 2019), en la diversificación de los linajes que conforman la flora del este asiático.

Por otra parte, establecer los límites taxonómicos ha sido un desafío frecuente en varios grupos de plantas. Por muchos años, la taxonomía se basó exclusivamente en el análisis de caracteres morfológicos, sin embargo, sin despreciar su importancia, por sí solos no han sido suficientes para resolver algunos problemas taxonómicos. En esta situación, los marcadores moleculares han contribuido a resolver problemas taxonómicos, como evaluar la monofilia (e. g. Butterworth y Wallace, 2004, Griffith y Porter, 2009, Kadereit *et al.,* 2016) y los límites entre distintos grupos taxonómicos (e.g. Chase *et al*., 2016). Por su parte, el poder de resolución de los estudios filogenómicos ha proporcionado evidencia para la

circunscripción de muchos grupos taxonómicos de plantas. Por ejemplo, a nivel de especies, el genotipado por secuenciación (GBS) permitió identificar que entre tres especies putativas del género *Lemna* (Lemnaceae), dos de ellas correspondían a una sola especie (Bog *et al.*, 2020). A nivel de género, el análisis realizado con datos provenientes de *genome skimming* de 228 especies de *Artemisia* y otros géneros cercanos*,* identificó que en el género *Artemisia* se debía incluir al género *Kaschgaria* para que se mantenga como monofilético (*e. g*. Jiao *et al*., 2023). Para el caso de la familia Cactaceae, los estudios filogenómicos tienen el potencial de refinar el tratamiento taxonómico ya realizado con caracteres morfológicos. Este refinamiento es necesario, debido a que los estudios moleculares ya han identificado la no monofilia de géneros como *Echinopsis* (Schlumpberger *et al.*, 2012), *Mammillaria* (Butterworth y Wallace, 2004), *Opuntia* (Griffith y Porter, 2009), *Pereskia* (Edwards *et al*., 2005) y *Rebutia* (Ritz *et al.*, 2007).

2. La diversificación de la familia Cactaceae y el desafío para resolver sus relaciones filogenéticas

La familia Cactaceae (Caryophyllales) representa uno de los componentes más conspicuos de las regiones áridas de América. Las 1,440 especies que se reconocen en la familia (Hunt, 2006) despliegan una gran diversidad de formas de crecimiento (Hernández-Hernández *et al*., 2011), hábitats (Mutke, 2015) y estrategias reproductivas (Mandujano *et al*., 2010). Esta gran diversidad se originó en un tiempo relativamente corto, pues Cactaceae es una de las familias más recientes de angiospermas, con una edad estimada en 28 a 35 Ma (Arakaki *et al*., 2011, Hernández-Hernández *et al*., 2014, Magallón *et al.*, 2015). Pese a los numerosos esfuerzos por resolver las relaciones filogenéticas en Cactaceae (Ritz *et al.*, 2007, Griffith y Porter, 2009, Bárcenas *et al*., 2011, Butterworth y Wallace, 2004) no ha sido posible obtener filogenias completamente resueltas para la familia. El origen reciente de la familia y la rápida diversificación de muchos de sus linajes (Arakaki *et al*., 2011, Hernández-Hernández *et al*., 2014) se postulan en esta tesis como factores que generan bajos niveles de variación molecular y que aumentan la probabilidad

de discordancia filogenética por sorteo incompleto de linajes. Sin embargo, esta discordancia también puede ser causada por eventos de hibridación interespecíficos e intergenéricos, considerados frecuentes en la familia (revisado en Machado, 2008). Por lo tanto, en esta tesis se usó una perspectiva filogenómica para elucidar las relaciones filogenéticas de Cactaceae, en particular del género *Mammillaria.*

3. El problema del género *Mammillaria*

3.1. Diversificación y patrones biogeográficos del género Mammillaria

Esta tesis se enfoca en el estudio de las especies del género *Mammillaria* (subfamilia Cactoideae) que agrupa 163 especies *sensu* Hunt (2006) o 143 de acuerdo con Korotkova *et al.* (2021). Siguiendo a Hunt (2006), *Mammillaria* es el género con la mayor riqueza, misma que se originó en los últimos 8 Ma (Arakaki *et al*., 2011, Hernández-Hernández *et al*., 2014). La diversificación de éste y de otros géneros ricos en especies como *Opuntia* (75) y *Echinocereus* (67), se ha propuesto como resultado del proceso de aridificación global que inició en el Mioceno tardío (~14 Ma) (Arakaki *et al*., 2011) y de la expansión del Desierto Chihuahuense (Hernández-Hernández *et al*., 2014).

Las regiones áridas y semiáridas de Estados Unidos de América y de México concentran el mayor número de especies de *Mammillaria* (Hernández y Gómez, 2015). En cuanto a los tipos de vegetación, este género alcanza su mayor diversidad en los matorrales xerófilos, aunque también se les encuentra en los bosques deciduos tropicales, bosques espinosos y pastizales; e incluso en los bosques de pino, bosques de pino-encino y bosques de encino (Arias *et al.*, 1997, Peters *et al*., 2008). Es posible que la distribución geográfica de la mayoría de las especies de *Mammillaria* esté determinada por atributos ecológicos, lo cual se refleja en la gran proporción de especies endémicas o con distribución altamente restringida (Hernández y Godinez, 1994).

La distribución geográfica actual del género *Mammillaria* abarca desde el sur de Estados Unidos de América, en los estados de Utah y Nevada, hasta el norte de los Andes, en los países de Colombia y Venezuela. El género también se encuentra en algunas de las

islas del Caribe (Pilbeam, 1999, Hernández y Gómez, 2015). Sin embargo, la riqueza de *Mammillaria* se concentra en el territorio mexicano, donde hasta el 88% de las especies son endémicas. Del total de especies de *Mammillaria,* solamente dos (*M. mammillaris* y *M. nivosa*) no se registran en México. En este país, la mayor riqueza se registra en la Península de Baja California, el Desierto Chihuahuense y el Valle de Tehuacán-Cuicatlán (Hernández y Gómez, 2015). Tres especies del género (*M. albilanata*, *M. columbiana y M. voburnensis*) se extienden desde México hasta América Central, éstas se distribuyen en la franja semiárida que atraviesa transversalmente a Guatemala. De estas especies, solo *M. columbiana y M. voburnensis* alcanzan el sureste de Honduras y oeste de Nicaragua. En el resto de los países de América Central no se tienen registros de *Mammillaria*. El género se registra nuevamente en Colombia y Venezuela, únicamente con dos especies: *M. colombiana* y *M. mammillaris.* Finalmente, en las islas del Caribe se documentan cuatro especies: *M. columbiana*, *M. mammillaris, M. nivosa* y *M. prolifera* (Pilbeam, 1999, Hernández y Gómez, 2015).

3.2. Delimitación y clasificación taxonómica de Mammillaria

La especie tipo de *Mammillaria* fue originalmente descrita por Linneo como *Cactus mammillaris* en 1752 y fue renombrada como *Mammillaria mammillaris* por Haworth (1812). En los 249 años que han transcurrido desde 1752 hasta el año 2001 (Anderson, 2001) las especies agrupadas en el género *Mammillaria* han tenido hasta 14 sinónimos distintos. Lo anterior refleja un problema para la delimitación taxonómica de *Mammillaria,* tanto entre grupos de especies del mismo género como con otros géneros. Particularmente, los límites taxonómicos de *Mammillaria* han sido confusos respecto a los géneros *Coryphantha, Escobaria, Mammilloydia, Neolloydia, Ortegocactu*s y *Pelecyphora*. La complejidad morfológica de *Mammillaria* y los seis géneros mencionados se ha traducido en una taxonomía poco clara, lo que se agrava por no contar con una filogenia resuelta. Por ejemplo, algunos taxa actualmente reconocidos en el género *Coryphantha,* fueron considerados inicialmente como parte de *Mammillaria* (Engelmann, 1856). Posteriormente

Coryphantha fue separado y elevado a nivel de género con base en la presencia de un surco en la superficie de los tubérculos (Lemaire, 1868); aunque algunos otros autores (*e. g.* Schumann, 1899) consideraron a *Coryphantha* como subgénero de *Mammillaria* (Anderson, 1986). Además, *Escobaria dasyacantha* y *E. tuberculosa* fueron descritas originalmente como *Mammillaria,* subgénero *Coryphantha* por Schumann; fue hasta 1923 que *Escobaria* se propuso como un género independiente a *Coryphantha* y *Mammillaria* por Britton y Rose (1923). Sin embargo, los límites entre los géneros *Escobaria* y *Coryphantha* continúan siendo controversiales hasta el día de hoy (Dicht y Lüthy, 2005, Sánchez *et al*., 2022). En cuanto al género *Neolloydia*, en el que Hunt (2006) solo reconoce dos especies, históricamente ha agrupado a cerca de 30 especies que actualmente se incluyen en los géneros *Coryphantha, Cumarina, Escobaria, Mammillaria*, *Thelocactus* y *Turbinicarpus* (Anderson, 1986). Por su parte, la única especie del género *Mammilloydia*, *M. candida,* fue descrita originalmente como *Mammillaria* y posteriormente elevada a nivel de género por Buxbaum (1951). Esta especie continuó causando controversia entre los taxónomos al considerarla como subgénero de *Mammillaria* o como género independiente (Pilbeam, 1999). Recientemente, sumando a la inestabilidad taxonómica de estas especies, se propuso incluir en el género *Cochemiea* a algunas de las especies del género *Mammillaria* y a las que se agrupan en *Neolloydia* y *Ortegocactus* (Breslin *et al.*, 2021).

Actualmente, se propone que el tipo de areola y las características superficiales de la semilla (figura 1) son diagnósticos del género *Mammillaria* (Lüthy, 1995; Hunt, 2006). En las especies de este género la areola se caracteriza por ser dimórfica; presentando dos meristemos areolares separados entre sí (figura 1a); en otros géneros ambos meristemos se encuentran juntos (figura 1b). El meristemo que se localiza en la punta del tubérculo se especializa en la producción de espinas; mientras que el meristemo localizado en la axila del tubérculo se especializa en la producción floral (Anderson, 2001). En *Mammillaria* estos dos meristemos no se encuentran conectados; en contraste, en géneros como *Coryphantha* se presenta un surco de tejido meristemático que conecta ambas porciones de la areola (figura 1c). En cuanto a las características de la semilla, en el género *Mammillaria* las

células superficiales de la testa se caracterizan por tener una depresión al centro de la pared periclinal externa (figura 1d). Este carácter distingue a las especies de *Mammillaria* del género monotípico *Mammilloydia*, el cual presenta semillas aparentemente lisas (figura 1e). En esta tesis se siguió la propuesta de Hunt (2006), quien delimitó a las especies de *Mammillaria* con base en los dos caracteres mencionados. Sin embargo, es posible que estos dos caracteres se hayan originado por convergencia morfológica en las especies de *Mammillaria*. En consecuencia, si han convergido varios linajes en lo que actualmente se agrupa en el género *Mammillaria*, implicaría la necesidad de segregar a *Mammillaria* en varios géneros independientes.

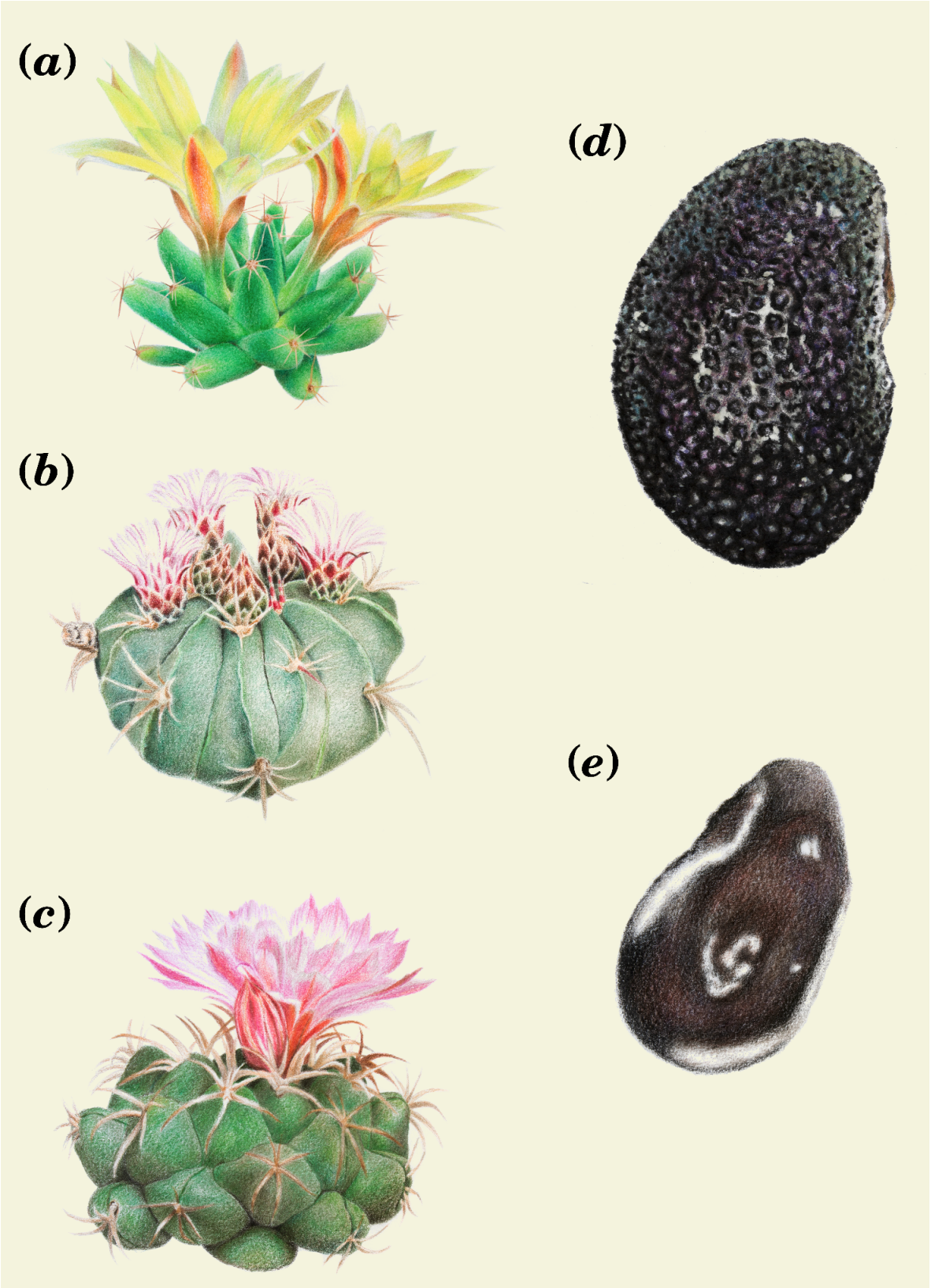

Figura 1. Caracteres diagnósticos del género *Mammillaria.* Areolas y semillas de *Mammillaria* y su comparación con las de otros géneros de cactus. *Mammillaria longimamma:* areolas dimórficas (a). *Ferocactus macrodiscus:* las flores y las espinas

surgen de la misma areola (b). *Coryphantha elephantidens:* areolas con surco (c). Semilla de *M. longimamma:* células de la testa con depresiones (d). Semilla de *Mammilloydia candida:* células de la testa sin depresiones y con apariencia lisa (e). Ilustraciones de Eira A. Chincoya. Lápiz de color.

Por otra parte, el sistema de clasificación infragenérico más utilizado para *Mammillaria* fue propuesto por Hunt (1971, 1981, 2006). Hunt (1971) estableció su primera propuesta de clasificación a partir de características como la consistencia del látex, ausencia o presencia de perispermo, morfología de la testa, color de la cubierta de semilla y distribución geográfica. Hunt (1971) reconoció dentro del género *Mammillaria* cinco subgéneros, tres secciones y 11 series a partir de la clasificación de Schumann (1899) e incorporando algunas sugerencias de Buxbaum (1951), sobre todo a nivel de serie. Posteriormente, Hunt (1981, 2006) realizó actualizaciones y cambios a su propia propuesta de clasificación infragenérica. En su última actualización (2006), Hunt agrupa 163 especies en ocho subgéneros y 15 series (Cuadro 1). En esta tesis se utilizó esta última propuesta de Hunt (2006) como referencia para las categorías taxonómicas infragenéricas.

Cuadro 1. Clasificación taxonómica infragenérica propuesta por Hunt (2006) para *Mammillaria*.

Subgénero	**Serie**
1. Chilita	
2. Cochemiea	
3. Dolichothele	
4. Krainzia	1. Herrerae + Pectiniferae
	2. Longiflorae
5. Oehmea	
6. Mammillaria	3. Decipientes
	4. Heterochlorae
	5. Lasiacanthae

	6. Leptocladodae
	7. Leucocephalae
	8. Mammillaria
	9. Polyacanthae
	10. Polyedrae
	11. Proliferae
	12. Rhodanthae
	13. Sphacelatae
	14. Stylothelae
	15. Supertextae
7. Mamillopsis	
8. Phellosperma	

3.3. *Aproximaciones moleculares para circunscribir a Mammillaria y resolver sus relaciones filogenéticas*

La controversia respecto a la delimitación taxonómica de *Mammillaria* y géneros cercanos ha devenido en una amplia representación del género en estudios moleculares. Uno de los primeros estudios, realizado con un locus de cloroplasto (intron *rpl16*), analizó las relaciones filogenéticas de 66 especies de cactus, de las cuales 15 pertenecían al género *Mammillaria*. Este estudio acuñó el nombre de “clado Mammilloide” para referirse al clado conformado por *Mammillaria* y otros seis géneros con los que *Mammillaria* ha tenido límites taxonómicos lábiles (*Coryphantha, Escobaria, Mammilloydia, Neolloydia, Pelecyphora* y *Ortegocactus*). Si bien, este estudio obtuvo una filogenia politómica y con valores de soporte bajos, fue el primero en identificar que el género *Mammillaria* no delimita a un grupo monofilético (Butterworth *et al.,* 2002). El origen no monofilético de *Mammillaria* fue confirmado por otros estudios moleculares, como los realizados con el espaciador *psbA-trnH* y el intrón *rpl16* (Butterworth y Wallace, 2004), el intrón *matK* (Barcenas *et a*l., 2011) y 10 regiones codificantes y no codificantes del genoma de cloroplasto (Crozier,

2005). Sin embargo, el muestreo molecular de estos estudios siguió siendo insuficiente para resolver las relaciones filogenéticas de *Mammillaria* y establecer sus límites. Recientemente, el primer estudio filogenómico del género obtuvo una filogenia para 78 especies de cactus, de las cuales 52 pertenecen al género *Mammillari*a, a partir del análisis de 93,808 pb del genoma de cloroplasto (Breslin *et al.*, 2021). Con base en sus resultados los autores hicieron propuestas de reasignación taxonómica a nivel de género para el clado Mammilloide. La más drástica consiste en agrupar bajo el género *Cochemiea* a las especies agrupadas en un clado compuesto por los géneros *Neolloydia*, *Ortegocactus* y algunas especies de *Mammillaria*. Sin embargo, este estudio estuvo fuertemente sesgado a las especies de *Mammillaria* que se distribuyen en el noroeste de México y suroeste de Estados Unidos, por lo que considero que no es representativo del género.

La información presentada en las secciones anteriores demuestra la persistencia de numerosas interrogantes en el género *Mammillaria*. Los estudios realizados demuestran que para dilucidar la compleja historia evolutiva del género resulta insuficiente analizar unos pocos loci. Si bien, se ha postulado que la diversidad del género se originó en respuesta a la aridificación global (Arakaki *et al*., 2011), aún no se termina de comprender el papel de otros eventos históricos en su diversificación y en la configuración de su distribución geográfica actual. Además, la controversia respecto a su delimitación taxonómica sigue tan viva como en los más de dos siglos previos a los análisis moleculares, particularmente con la reciente propuesta de circunscripción realizada por Breslin *et al.* (2021). Desafortunadamente, la ausencia de una hipótesis sólida sobre las relaciones filogenéticas de *Mammillaria*, con un muestreo taxonómico amplio, ha planteado una barrera para abordar estas y otras problemáticas.

Por lo tanto, en esta tesis se utilizaron datos genómicos para resolver las relaciones filogenéticas de *Mammillaria* y los otros seis géneros del clado Mammilloide (*Coryphantha, Escobaria, Mammilloydia, Neolloydia, Pelecyphora* y *Ortegocactus*). Además, los resultados filogenéticos del genoma de cloroplasto se analizaron bajo el marco teórico de la biogeografía histórica para incorporar el contexto espacio-temporal de la historia evolutiva

de las especies. Los objetivos de esta tesis fueron 1) estimar las relaciones filogenéticas de las especies a partir de marcadores provenientes de distintos compartimentos genómicos (*i. e.* cloroplasto y núcleo); 2) interpretar los resultados filogenéticos en el contexto de la delimitación taxonómica a nivel de género, particularmente de *Mammillaria*; 3) identificar los posibles factores históricos que promovieron la diversificación de los cactus analizados y configuraron su distribución geográfica actual; 4) evaluar la discordancia filogenética en el clado Mammilloide e identificar los procesos biológicos que posiblemente la han causado y 5) examinar la utilidad de dos conjuntos de sondas universales para angiospermas en la realización de estudios filogenéticos en Cactaceae.

CAPÍTULO 1. Phylogenomics and Biogeography of the Mammilloid Clade Revealed an Intricate Evolutionary History Arose in the Mexican Plateau

(Artículo de requisito)

biology

MDPI

Article

Phylogenomics and Biogeography of the Mammilloid Clade Revealed an Intricate Evolutionary History Arose in the Mexican Plateau

Delil A. Chincoya [1,2,*], Salvador Arias [3], Felipe Vaca-Paniagua [4,5,6], Patricia Dávila [7] and Sofía Solórzano [1,*]

1 Laboratorio de Ecología Molecular y Evolución, UBIPRO, FES Iztacala, Universidad Nacional Autónoma de México, Avenida de los Barrios 1, Los Reyes Iztacala, Tlalnepantla de Baz 54090, Estado de México, Mexico
2 Posgrado en Ciencias Biológicas, Unidad de Posgrado, Edificio D, 1° Piso, Circuito de Posgrados, Ciudad Universitaria, Coyoacán 04510, Ciudad de México, Mexico
3 Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, Tercer Circuito Exterior, Ciudad Universitaria, Coyoacán 04510, Ciudad de México, Mexico
4 Laboratorio Nacional en Salud, Diagnóstico Molecular y Efecto Ambiental en Enfermedades Crónico-Degenerativas, FES Iztacala, Universidad Nacional Autónoma de México, Los Reyes Iztacala, Tlalnepantla de Baz 54090, Estado de México, Mexico
5 Subdirección de Investigación Básica, Instituto Nacional de Cancerología, Tlalpan 14080, Ciudad de México, Mexico
6 Unidad de Investigación en Biomedicina, FES Iztacala, Universidad Nacional Autónoma de México, Los Reyes Iztacala, Tlalnepantla de Baz 54090, Estado de México, Mexico
7 Laboratorio de Recursos Naturales, UBIPRO, FES Iztacala, Universidad Nacional Autónoma de México, Avenida de los Barrios 1, Los Reyes Iztacala, Tlalnepantla de Baz 54090, Estado de México, Mexico
* Correspondence: dela@comunidad.unam.mx (D.A.C.); solorzanols@unam.mx (S.S.)

Simple Summary: Cacti account for nearly 1440 species, most of them native to the American continent. These succulent plants are the most ubiquitous elements of the arid ecosystems. Mexico harbors the highest number of cacti species in the world (45%). Unfortunately, many of them are threatened by human activities. Although having this biodiversity relevance, presently the evolutionary processes of cacti have been poorly studied. Because the biological and conservation unit is the species, evolutionary studies provide relevant information. In this study, we analyzed how and when past events shaped the evolutionary relationships of 103 species. Our results showed that from 4.5 million years ago the arid regions of Mexico were the locations for abundant cacti speciation. From these lands, cacti have colonized most of the Mexican territories, the southern regions of the United States, as well as the Caribbean. The evolution of these plants was probably promoted by past temperatures that were comparable to the present ones. We identified different speciation and dispersal events in these fascinating plants. This study identified the Mexican Plateau as the place where the early stages of the evolutionary history of cacti occurred.

Abstract: Mexico harbors ~45% of world's cacti species richness. Their biogeography and phylogenomics were integrated to elucidate the evolutionary history of the genera *Coryphantha*, *Escobaria*, *Mammillaria*, *Mammilloydia*, *Neolloydia*, *Ortegocactus*, and *Pelecyphora* (Mammilloid Clade). We analyzed 52 orthologous loci from 142 complete genomes of chloroplast (103 taxa) to generate a cladogram and a chronogram; in the latter, the ancestral distribution was reconstructed with the Dispersal-Extinction-Cladogenesis model. The ancestor of these genera arose ~7 Mya on the Mexican Plateau, from which nine evolutionary lineages evolved. This region was the site of 52% of all the biogeographical processes. The lineages 2, 3 and 6 were responsible for the colonization of the arid southern territories. In the last 4 Mya, the Baja California Peninsula has been a region of prolific evolution, particularly for lineages 8 and 9. Dispersal was the most frequent process and vicariance had relevance in the isolation of cacti distributed in the south of Mexico. The 70 taxa sampled as *Mammillaria* were distributed in six distinct lineages; one of these presumably corresponded to this genus, which likely had its center of origin in the southern part of the Mexican Plateau. We recommend detailed studies to further determine the taxonomic circumscription of the seven genera.

Citation: Chincoya, D.A.; Arias, S.; Vaca-Paniagua, F.; Dávila, P.; Solórzano, S. Phylogenomics and Biogeography of the Mammilloid Clade Revealed an Intricate Evolutionary History Arose in the Mexican Plateau. *Biology* **2023**, *12*, 512. https://doi.org/10.3390/biology12040512

Academic Editor: Lorenzo Peruzzi

Received: 25 February 2023
Revised: 21 March 2023
Accepted: 24 March 2023
Published: 29 March 2023

Keywords: arid lands; biogeography; Cactaceae; colonization; *Mammillaria*; Mexican Plateau; Miocene; phylogenomics; Pleistocene; recent diversification

1. Introduction

The integration of the analytical frameworks of phylogenetics and biogeography allows analysis of the influence of biogeography on the evolutionary history of extant taxa, as well as to identify all those biogeographical events that promoted speciation [1]. The studies that incorporate these frameworks have inferred past biogeographical scenarios that have shaped the current geographical ranges of the species (e.g., [2,3]) and even large flora assemblies (e.g., [4–6]). Furthermore, they have enabled evaluation of the relative role of vicariance and dispersal in shaping the current geographical distribution of species [1,7,8]. Gradually, with the advances of high throughput sequencing technologies, more of these studies are using denser molecular sampling, which has made it possible to obtain confident phylogenetic trees that may serve to resolve close phylogenetic relationships [9]. Poor molecular sampling usually produces non-monophyletic trees, or discordance with phylogenies based on morphology [10]. In addition, the biogeographical data may support the establishment of taxonomic limits between species [10], and actually they have been identified as providing better auxiliary information than morphology to elucidate phylogenetic relationships [11].

In addition, it is recognized that global paleoclimatic changes have shaped the large current distribution patterns of the biota and caused extinctions at different geographical scales [12]. Furthermore, they influenced the expansion and contraction of the geographical distribution of the current extant species (e.g., [13,14]). Consequently, paleoclimate changes have been recognized as one of the most influential factors in shaping the world biodiversity patterns at large scales, but also for understanding the current local flora assemblies (e.g., [15]). On the other hand, the topography and the intricate local orography have also influenced the ecological, biogeographical, and evolutionary processes of the local biota [16]. All these events and processes that occurred in the past might have modified gene flow patterns, which gradually may cause population genetic divergence and eventually promoted speciation processes [17].

In the contemporary arid lands of the American continent, many complex assemblages of native local floras are found in which cacti taxa are the most ubiquitous elements. The nearly 1440 taxa grouped in Cactaceae [18] are recognized as a monophyletic group [19]. Today the evolutionary history of Cactaceae, particularly its origin and mode of speciation, are still considered enigmatic [20]. Due to the lack of fossil records of Cactaceae representatives, there is no direct evidence to date its origin. However, estimations based on molecular clock hypothesis have dated the origin of Cactaceae to nearly 28.8 million years ago (Mya) [21], or 32.11 Mya [22], and 35 Mya [23]. Accordingly, these estimations place the origin of Cactaceae in the Cenozoic Era, in the Paleogene period from the Late Eocene (~35 Mya) to the Middle Oligocene (~28 Mya). In addition, Arakaki et al. [23] concluded that unequal and inconstant speciation rates for 123 cacti sampled were explained by the environmental changes that occurred in the Miocene, based on the phylogenetic tree obtained with two loci, one from the nuclear genome (PHYC) and the other one from the chloroplast (trnK/matK). Accordingly, these authors suggested that there have been at least six main peaks of speciation in the evolutionary history of Cactaceae. These authors dated the earliest two speciation peaks to 25 Mya and 15 Mya, whereas the other four occurred in the last 8 million years. Furthermore, they showed that those last four peaks were contemporaneous to the decreases in atmospheric CO_2 that promoted global aridification, giving new ecological opportunities to cacti [23]. On the other hand, specialized paleoclimatic studies (e.g., [24,25]) have dated the decreases of atmospheric CO_2 from the Middle Miocene (14 Mya) to the Middle Pleistocene (0.8 Mya). These relatively low levels

of CO_2 eventually caused a cooler and drier global climate, a phenomenon recognized as an aridification process [26].

In Cactaceae, the genus *Mammillaria* Haw is notable for its diversity [27], conservation concerns [28] and unresolved phylogenetic and taxonomic issues [18]. The taxonomy of *Mammillaria* has been controversial from its original description. In 1753, Charles Linnaeus described the type specimen as *Cactus mammillaris* L. and later it was renamed as *Mammillaria* in 1812 [29]. During its history, the genus *Mammillaria* has received 14 different names [27], reflecting the difficulty in achieving a clear taxonomic circumscription based almost entirely in external morphological traits. Throughout the last two centuries, numerous attempts have been made to organize the wide infrageneric morphological variation among taxa classified as *Mammillaria* (e.g., [30–32]. The most recent infrageneric classification was proposed by Hunt [18], who recognized eight subgenera and 15 series. The subgenus *Mammillaria* contains the highest number of species (117), followed by *Chilita* Orcutt (18) and *Krainzia* Backeb. (12); whereas *Cochemiea* Brandegee, *Dolichothele* (K.Schum.) Britton & Rose, *Mammillopsis* Morren, *Oehmea* Buxb., and *Phellosperma* Britton & Rose together add 16 species. For the purposes of the present study, we follow this last infrageneric classification system; nevertheless, phylogenetic support for these infrageneric classifications has not been tested.

Today, the global geographical distribution of *Mammillaria* ranges from the southern arid lands of the United States to the north of South America. Mexico has the highest documented diversity of *Mammillaria*. Nearly 20% of the species of *Mammillaria* are distributed in the Mexican arid lands of the southern part of the Chihuahuan Desert [33]. Only two of the 163 species currently recognized [18], *M. mammillaris* (L.) H. Karst and *M. nivosa* Link ex Pfeiff. are not documented in this country [33]. The genus *Mammillaria* is a rare taxon across Central America, as only four species are recorded in Guatemala (*M. albilanata* Backeb., *M. columbiana* Salm-Dyck, *M. ericantha* Link & Otto ex Pfeiff. and *M. voburnensis* Scheer), and two of them are also distributed across Nicaragua and Honduras (*M. columbiana* and *M. voburnensis*). Two more, *M. columbiana* and *M. mammillaris*, are documented in some small localities in the north of Venezuela and Colombia. In addition, four species (*M. columbiana*, *M. mammillaris*, *M. nivosa*, and *M. prolifera* (Mill.) Haw.) are recorded in the Caribbean islands [33].

The early phylogenetic studies carried out with *Mammillaria* reignited the unsolved discussion regarding its unclear taxonomic circumscription and its limits with taxa of another six genera (*Coryphantha* (Engelm.) Lem., *Escobaria* Britton & Rose, *Mammilloydia* Buxb., *Neolloydia* Britton & Rose, *Ortegocactus* Alexander, and *Pelecyphora* Ehrenb.). These six genera and *Mammillaria* compose the Mammilloid Clade [34]. Butterworth and Wallace [35] analyzed the phylogenetic relationships of 123 species of Mammilloid Clade (113 of them grouped in *Mammillaria)* based on two plastid loci (*rpl16* intron and the intergenic spacer *psbA-trnH*). Their phylogenetic tree showed abundant polytomies and low support bootstrap values. In addition, the sampled taxa of these six genera were grouped together with those of *Mammillaria*. Hence the authors concluded that this genus has a polyphyletic origin. Later, Crozier [36] used 10 plastid loci to analyze 157 cacti taxa; only 29 of them were *Mammillaria* taxa and 10 belonged to the six genera. The results of this study did not resolve the phylogenetic relationships of the sampled taxa; it also concluded non-monophyly for *Mammillaria*. In addition, Crozier [36] concluded that the monophyly of *Mammillaria* could only be obtained if: (1) the *Mammillaria* genus was expanded to include all the species currently grouped in the six genera; or (2) the genus *Mammillaria* includes only those species of the subgenus *Mammillaria* sensu Hunt [37]. Breslin et al. [38] recently sampled 93,808 bp of the large single copy (LSC) of the chloroplast genome from 78 cacti taxa, 52 of which were *Mammillaria* and 17 from five genera (*Coryphantha*, *Escobaria*, *Neolloydia*, *Ortegocactus*, and *Pelecyphora*). These authors concluded monophyly for *Mammillaria* by excluding all those species that were grouped in a distinct clade, which was composed of taxa in *Mammillaria*, *Neolloydia*, and *Ortegocactus*. In addition, it was proposed that all the species of this clade to be placed inside the genus *Cochemiea.*

In this study, we integrated phylogenomics and historical biogeography to elucidate the controversial evolutionary history of the group of seven genera of cacti (*Coryphantha*, *Escobaria*, *Mammillaria*, *Mammilloydia*, *Neolloydia*, *Ortegocactus*, and *Pelecyphora*) sensu Hunt [18]. We hypothesized that these taxa have a monophyletic origin, whose unique ancestor arose recently and rapidly evolved in response to past decreases in global temperature. The objectives were to evaluate the phylogenetic relationships of the studied species, to estimate their divergence times, and to identify the probable ancestral geographical distribution of the taxa studied in these seven genera; to discuss the possible effects of past global temperature and orographic events in the colonization and expansion of these cacti across the arid lands of Mexico; and finally we use our results to identify the taxonomic limits of the genera studied with emphasis on the taxa sampled in the genus *Mammillaria*.

2. Materials and Methods

2.1. Taxon Sampling

A total of 142 complete chloroplast genomes (cpDNA) of 103 taxa were analyzed (Table S1), of which 141 cpDNA belong to the tribe Cacteae (Cactoideae). The non Cactoideae taxon *Blossfeldia liliputana* Werderm. was included because it was identified as the sister species for the rest of the subfamily Cactoideae [36]. We compiled these cpDNA from the following sources: seven complete cpDNA of *Mammillaria* previously published [39], as well as the raw data of 86 genomes that were downloaded from NCBI site, which were linked to BioProject PRJNA671701 [38]. In addition, the whole complete chloroplast genomes of 49 taxa were de novo sequenced in this study. The tissue samples for 47 of these taxa were provided by the collection of the Botanical Garden of the Universidad Nacional Autónoma de México, whilst the tissues of *M. napina* J.A.Purpus and *M. huitzilopochtli* D.R.Hunt were obtained from completed research projects (SS). Among these 142 genomes, 132 represented seven of the genera (i.e., Mammilloid Clade): *Coryphantha*, *Escobaria*, *Mammillaria*, *Mammilloydia*, *Neolloydia*, *Ortegocactus*, and *Pelecyphora* (Table 1).

Table 1. Taxon diversity sampled for the seven genera. Taxonomic names and the total number of recognized taxa for the levels of genus and subgenus following Hunt [18]. Total number of the taxa and number of genomes analyzed (in silico plus those de novo sequenced); and the number of genomes de novo sequenced. NA indicates that the subgenus level is not recognized.

Genus	Subgenus	Number of Recognized Taxa	Total Number of Analyzed Taxa	Number of Genomes Analyzed	Number of Genomes de novo Sequenced
1. Coryphantha		42	10	11	3
	1.1 Coryphantha	26	8	9	2
	1.2 Neocoryphantha	15	2	2	1
2. Escobaria	NA	19	8	9	2
3. Mammillaria		163	70	105	37
	3.1. Chilita *	18	16	33	3
	3.2. Cochemiea *	3	3	10	2
	3.3 Dolichothele	6	2	2	2
	3.4. Krainzia	12	5	5	2
	3.5. Mammillaria	117	37	46	26
	3.6. Mammillopsis	1	1	1	0
	3.7. Oehmea	1	1	1	1
	3.8. Phellosperma	5	5	7	1

Table 1. *Cont.*

Genus	Subgenus	Number of Recognized Taxa	Total Number of Analyzed Taxa	Number of Genomes Analyzed	Number of Genomes de novo Sequenced
4. Mammilloydia	NA	1	1	1	1
5. Neolloydia *	NA	2	2	3	2
6. Ortegocactus *	NA	1	1	2	1
7. Pelecyphora	NA	2	1	1	1

* Taxa included in *Cochemiea* according to Breslin et al. [38].

In addition, the taxonomic sampling covered the whole geographical range of five of these genera (*Coryphantha*, *Mammilloydia*, *Neolloydia*, *Ortegocactus*, and *Pelecyphora*). In contrast, the geographical range of *Mammillaria* was not sampled in South America; and for *Escobaria* was not sampled the Caribbean. Of them, 105 specimens (70 taxa) corresponded to *Mammillaria*, which are currently distributed in continental and peninsular Mexican territories, as well as the southern parts of the USA and the Caribbean. We documented the geographical distribution in the Global Biodiversity Information Facility (GBIF) ((https://www.gbif.org/ (accessed on 20 March 2022)) (Figure 1). In order to reduce record density, we used the spThin package [40] to discard all those records with <1 km in separation distance. The geographical data of those remained records were hand-curated following to Hernández and Gómez-Hinostrosa [33]. In addition, we sampled as external groups 10 specimens of eight genera: *Acharagma* (N.P. Taylor) Glass, *Ariocarpus* Scheidw., *Blossfeldia* Werderm., *Cumarinia* (Knuth) Buxb., *Lophophora* J.M. Coult., *Stenocactus* (K. Schum.) A. Berger, *Strombocactus* Britton & Rose, and *Turbinicarpus* Backeb.

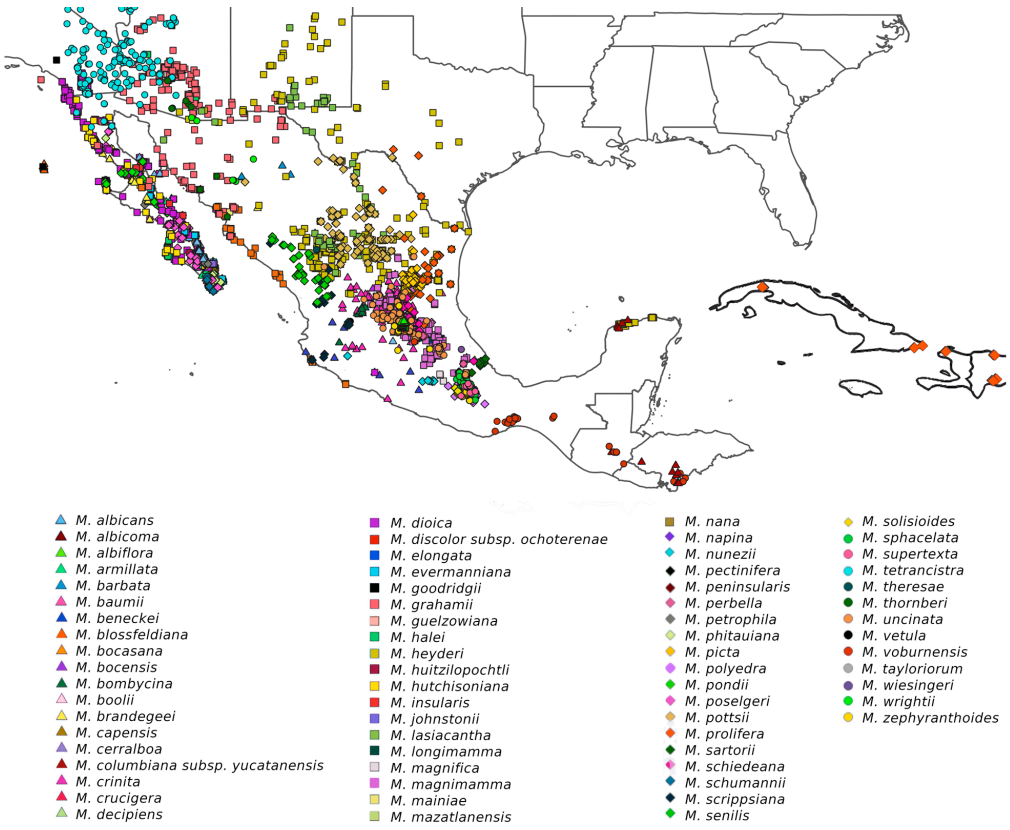

Figure 1. World geographical distribution of the 70 taxa of *Mammillaria*. Geographical distribution per taxon is showed in detail in Supplementary Materials (File S1).

2.2. DNA Extraction, cpDNA Enrichment, and High-Throughput Sequencing

For each of the 49 species de novo sequenced, 30–100 mg of frozen tissue was obtained to isolate 1 ug of gDNA with A260/280 ratio $\geq$ 1.7. The tissue samples were individually processed with the DNeasy plant mini kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. To obtain an enriched proportion of chloroplast genome, these gDNAs were processed with the NEBNext Microbiome DNA Enrichment Kit (New England BioLabs, Ipswich, MA, USA) according to the kit's instructions. These enriched DNAs were used to prepare pair-end (PE) genomic libraries with the Nextera XT kit, with mean insert size of 400 bp, and were sequenced in MiSeq 2 $\times$ 300 cycles.

2.3. De Novo Assembly of Chloroplast Genomes

We assembled de novo the raw data of the 86 genomes attached to Breslin et al. [38] (Table S1); as well as the raw data of the 49 taxa de novo sequenced. These 135 genomes were filtered, trimmed and adapters were removed with TrimGalore version 0.4.3 [41]. The recovered reads with PHRED quality score $\geq$ 15 and length $\geq$ 80 bp were assembled with Get Organelle version v1.7.1 [42], using as a seed the cpDNA of *M. supertexta* (GenBank accession: MN508963.1) previously published [39].

2.4. Phylogenetic Relationships and Divergence Times

The 142 genomes (Table 1) were analyzed with BLAST version 2.5.0 [43] to identify common loci based on sequence similarity. All these common loci were aligned with MAFFT version v7.310 [44]. Because the genomes analyzed showed different structural arrangements, some sequences were not recovered, giving alignments with a high proportion of missing data; and other alignments showed low molecular variation. Thus, these two types of alignments were discarded before further phylogenetic analysis. Accordingly, only those alignments with sequences present for $\geq$70% of the studied taxa, $\geq$15% of proportion of informative sites (PIS) and a length of $\geq$200 bp were obtained. Accordingly, a total of 52 orthologous loci (Table S2) were identified and concatenated in a matrix of 48,869 bp used for the phylogenomic analysis and estimation of times of divergence. The matrix partitions and substitution models were estimated with ModelFinder [45], implemented in IQ-TREE2 version 2.1.4-beta [46]. The phylogenetic tree was generated with IQ-TREE2 using *B. liliputana* as the outgroup and running 10,000 ultra-fast bootstrap (UFBoot) replicates. Then, we estimated the evolutionary times of divergences using two secondary calibrations from previous estimations for the Cactaceae family [22]. In our first calibration, we used the crown age of 12.67–24.46 Mya estimated for the whole Cactoideae subfamily, and for the second calibration, we used the crown age of 4.86–10.63 Mya for the clade composed of the seven focus genera (*Coryphantha*, *Escobaria*, *Neolloydia*, *Mammillaria*, *Mammilloydia*, *Ortegocactus*, and *Pelecyphora*). We estimated the divergence times with BEAST version v2.6.6 [47], whose specific input file was constructed with BEAUti. In this input file was specified the GTR + I + Γ substitution model, which was estimated with Modeltest [48] according to AICc, a lognormal relaxed molecular clock, calibration points as uniform distributions, a Yule process tree prior, and 200,000,000 generations with a sampling frequency of each 2000 generations. In addition, convergence of parameter estimation was corroborated with Tracer version v1.7.2 [49], and the trees were summarized in a maximum clade credibility tree with TreeAnnotator version v2.6.3 [50]; 10% of the trees were discarded based on this final analysis.

2.5. Biogeographical Analysis

For the biogeographical analysis we documented the current geographical distribution for each of the 141 specimens (102 taxa) native to the arid lands of North America. As *B. liliputana* is endemic to South America it was discarded from the biogeographical analysis. The geographical data were compiled from GBIF ((https://www.gbif.org/ (accessed on 20 March 2022)). These data were verified by checking the geographical distribution of taxa reported from different sources [33,51,52]. The geographical distribution range of the

141 specimens was classified into the respective Mexican Floristic Provinces proposed by Rzedowski [53]. We estimated the ancestral geographical ranges based on the dated tree using the R package BioGeoBEARS version 1.1.1 [54] implemented in RASP4 v4.0 [55]. We evaluated four distinct models of the geographical range evolution for the 141 specimens: both the model of Dispersal-Extinction-Cladogenesis (DEC); and the likelihood version of Dispersal-Vicariance Analysis (DIVALIKE) were tested under two conditions, with and without the assumption of Founder-Event Speciation (+J) parameter. Finally, we plotted the changes estimations in the global surface air temperature (ΔT) in relation to the current values, previously published in the supplementary material (S4) of Herbert et al. [26].

3. Results

3.1. Evolutionary History of Cacti: Recent Divergence and Intricate Biogeography

The topologies of the phylogenetic tree (ML tree) (Figure 2) and the chronogram (BI tree) (Figure 3) were highly concordant; the only difference was found in the relationships of the small clade composed of *Ariocarpus*, *Strombocactus*, and *Turbinicarpus*. In the ML tree, the clade *Turbinicarpus-Strombocactus* was sister to *Ariocarpus*, whereas in the BI tree, the *Turbinicarpus-Ariocarpus* clade was sister to *Strombocactus*. In the biogeographical analysis, the DEC model (without +J parameter) was selected according to the value of AICcWt (File S2), however, this value was slightly higher than that obtained for the DEC+J model. In addition, these two models provided very similar estimations of the ancestral geographical distribution (Figures S1 and S2). The biogeographical analysis estimated a total of 135 dispersal events and 13 vicariant events (Figure 4).

The phylogenetic results (Figure 2) clearly identified for the 102 taxa (141 specimens) of the Cacteae tribe a common ancestor, which arose in the Mexican Plateau in the Late Miocene ~12.08 Mya (95% HPD: 7.73–16.82) (Figure 3). According to the temperatures taken from the bibliography [26], between 15 and 9 Mya there was a drastic decrease in global temperature of ΔT~8 °C. In this period, our results showed two key phylogenetic splits in Cacteae: the first one that separated *Stenocactus* from the remaining 101 taxa; followed by the second one that separated *Ariocarpus*, *Strombocactus* and *Turbinicarpus* from the remaining 96 taxa (Figure 4). Later, during a short period of nearly 2.7 million years (from 9 to 6.3 Mya), the temperature stayed stable (ΔT~0.1), and during this period two splits occurred. The first one is represented by the separation of *Acharagma* and *Lophophora*; the second one consists of the separation of the ancestor of *Cumarinia*. In this period (9 to 6.3 Mya), we identified the beginning of the complex evolutionary history of the 93 taxa belonging to the Mammilloid Clade (Figure 4). Moreover, these taxa continued their diversification processes during the next period of two million years (between 6.3 and 4.3 Mya), when the temperature again declined (ΔT~4.4 °C). These diversification processes continued and intensified during the last 4.3 Mya. In this last 4.3 million years, the temperature has not been steady; from 4.3 to 1 Mya (Late Pliocene to Pleistocene) a slight increase in temperature (ΔT~1 °C) was documented. However, in the last 1 million years, the temperature has decreased (ΔT~0.5 °C) (Figure 4).

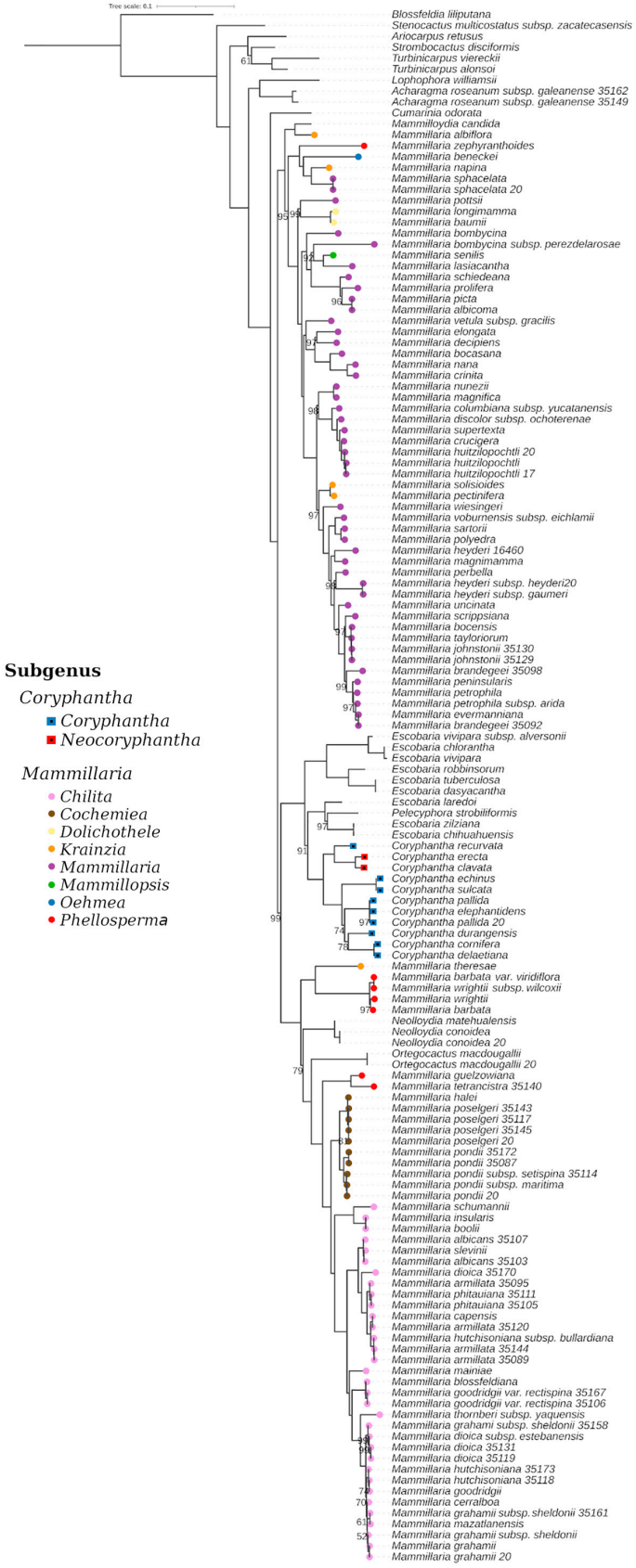

Figure 2. Phylogenetic tree estimated with IQ-TREE2 for the 103 taxa (142 specimens) using *B. liliputana* as the outgroup. Numbers below the nodes indicate UFBoot values < 100. Colored circles and squares indicate subgenera for *Mammillaria* and *Coryphantha*, respectively.

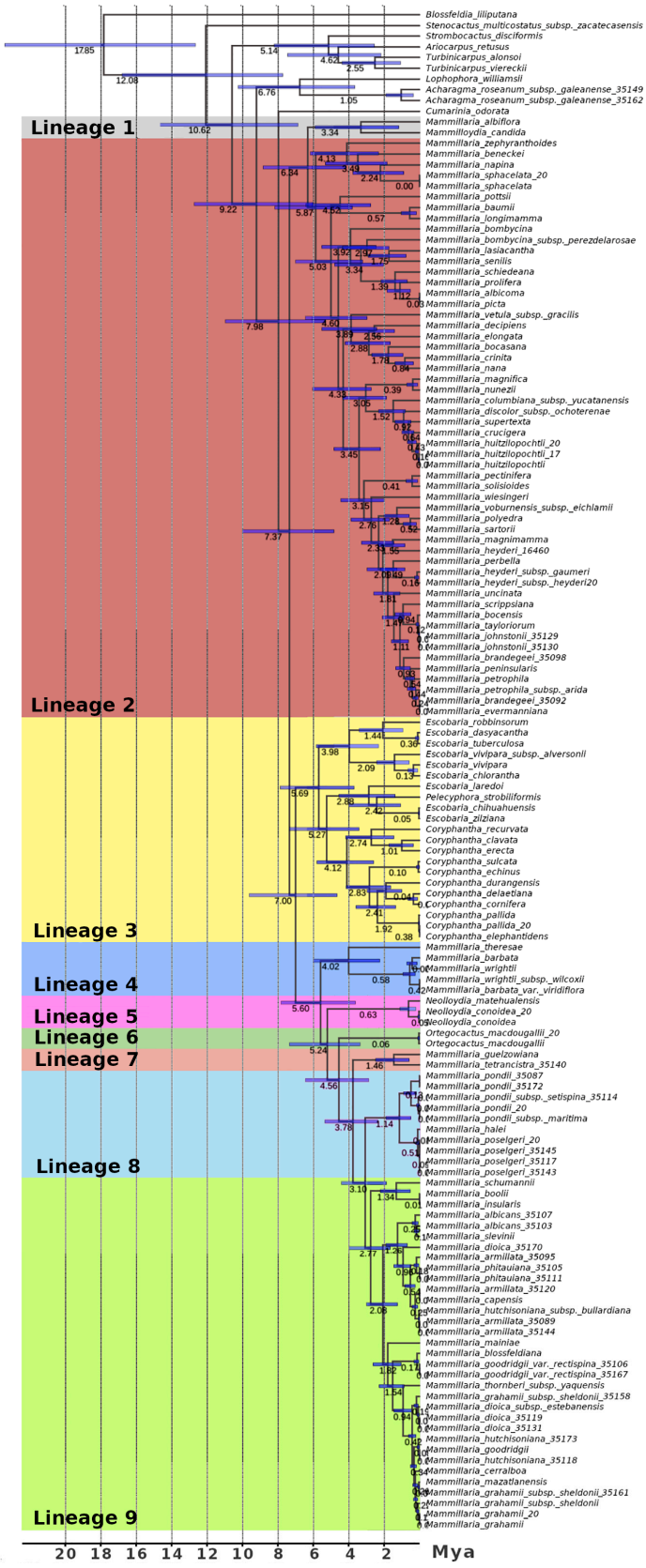

Figure 3. Chronogram estimated for the 103 taxa (142 specimens). The maximum clade credibility tree shows the divergence times estimated in BEAST. Blue bars represent 95% HPD intervals for the node ages. Shadow colors show the nine evolutionary lineages identified.

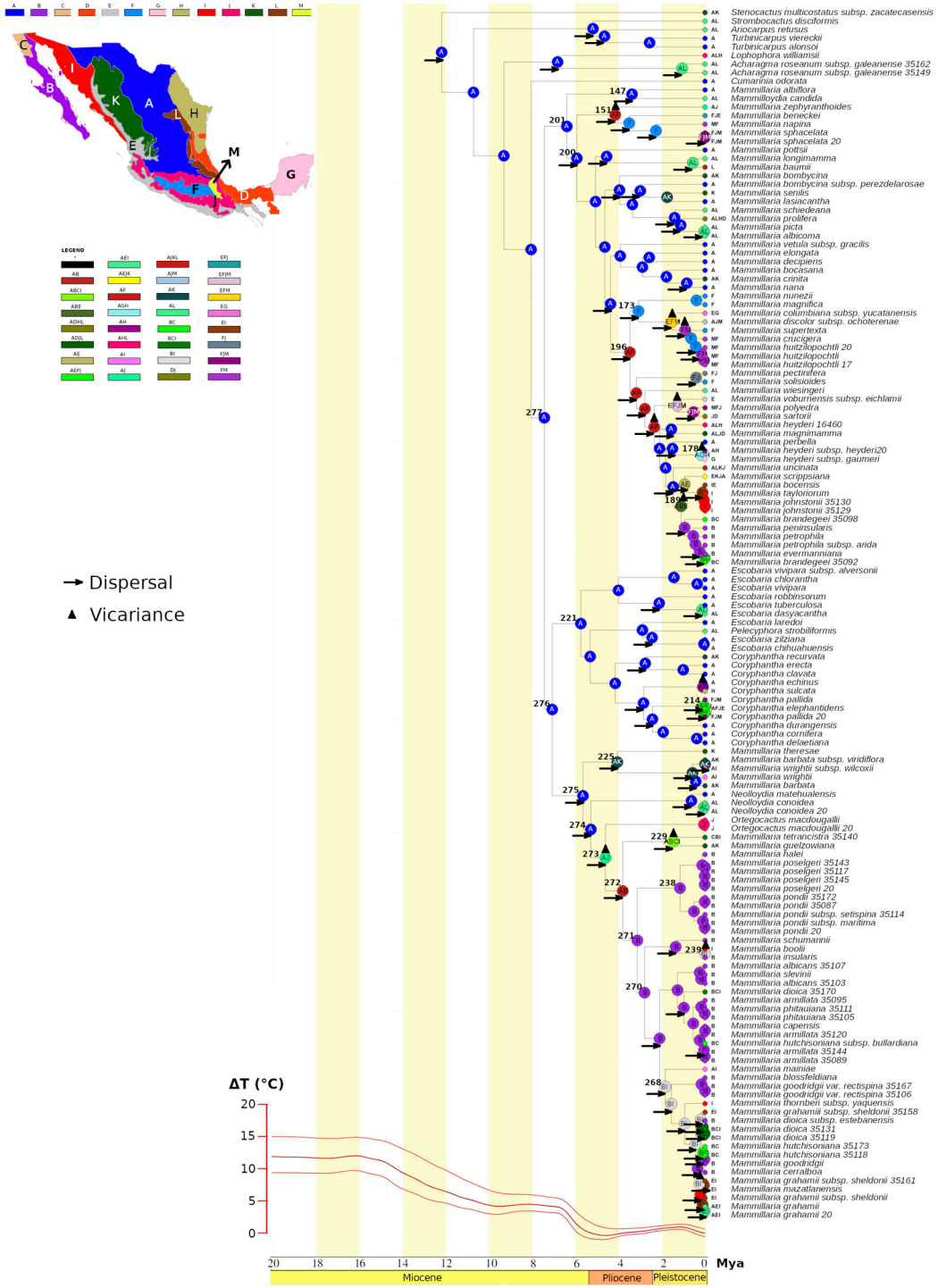

Figure 4. Ancestral geographical distribution of the North American taxa estimated on the chronogram. The estimation of the most likely ancestral distribution is represented in a colored circle for each node of the tree. The letters in the map and in the tree corresponded to Baja California (B), Balsas Basin (F), California (C), Gulf of Mexico Coast (D), Mexican Plateau (A), Meridional Serranias (J), Northeast Coastal Plain (H), Northwest Coastal Plain (I), Pacific Coast (E), Sierra Madre Occidental (K), Sierra Madre Oriental (L), Tehuacan Valley (M), and Yucatan Peninsula (G). The nodes of the main evolutionary events were numbered (see the text). The estimated events of dispersal and vicariance are indicated by arrows and triangles, respectively. The letters besides the tips indicate the current geographical distribution of the taxa. At the bottom of the figure were drawn the changes estimations in the global surface air temperature (ΔT).

The chronogram showed that the ancestor of the Mammilloid Clade originated nearly 7.37 Mya (95% HPD: 4.86–10.02 Mya). This early ancestor (node 277, Figure 4) arose at the end of the Miocene when the value ΔT of the temperature was low (ΔT~0.1 °C). This ancestor had as its probable ancestral geographical area the Mexican Plateau (Figure 4). Eventually, from this ancestor nine independent evolutionary lineages were derived (Figure 3); which profusely diversified in the last 4.3 Mya, when little increase-decrease of temperature occurred (Figure 4). This early common ancestor diverged into two new ancestors, one of which (node 276, Figure 4) was dated nearly 7 Mya (95% HPD: 4.67–9.64 Mya) (Figure 3). This ancestor had as its ancestral biogeographical scenario the Mexican Plateau (Figure 4), and from it evolved those taxa currently grouped in the six genera (*Coryphantha*, *Escobaria*, *Mammillaria*, *Neolloydia*, *Ortegocactus*, and *Pelecyphora*). The other ancestor (node 201, Figure 4) arose 6.34 Mya (95% HPD: 4.17–8.86 Mya) (Figure 3), probably also in the Mexican Plateau. From this lat ancestor evolved those taxa that were grouped in two genera: *Mammillaria* and *Mammilloydia*.

A conspicuous result obtained was that those 70 taxa (105 specimens) sampled as *Mammillaria* were distributed in two main independent clades (nodes 201 and 275, Figure 4). The immediate ancestors of these two clades originated in the past Mexican Plateau. However, those taxa derived from them clearly differ in their evolutionary history (Figure 4). Accordingly, the taxa sampled as genus Mammillaria, in fact, were distributed in six different and independent evolutionary lineages, each one with its own evolutionary history (Figures 3 and 4).

3.2. Evolutionary History of the Nine Lineages

3.2.1. Evolutionary Lineage 1

The short evolutionary lineage 1 was composed only of *Mammilloydia candida* (Scheidw.) Buxb. and *Mammillaria albiflora* Backeb, whose immediate ancestor was dated to nearly 3.34 Mya (95% HPD: 1.19–5.92 Mya). This ancestor (node 147, Figure 4) probably arose on the Mexican Plateau, during the Middle Pliocene, when the temperature underwent a slight increase (ΔT~1 °C) (Figure 4). At the present time, *M. candida* and *M. albiflora* are distributed in a small region on the southern region of the Mexican Plateau, and *M. candida* extends its geographical range to the northwest of this biogeographical area (Figure 4). Additionally, lineage 1 was identified as the phylogenetic sister to lineage 2 (Figure 2).

3.2.2. Evolutionary Lineage 2

The most probable ancestral geographical area for the immediate ancestor of lineage 2 was the Mexican Plateau (node 200, Figure 4). Lineage 2 arose nearly 5.87 Mya (95% HPD: 3.82–8.2), at the end of the Late Miocene, when the temperature decreased (ΔT~4.4 °C). However, most of the divergent processes in this lineage occurred in the last 4.3 million years, when a slight increase in the temperature (ΔT~1 °C) was followed by a slight decrease (ΔT~0.5 °C). This lineage grouped 45 of the sampled taxa, of which 37 taxa (82%) correspond to the subgenus *Mammillaria*, whereas the other eight taxa belong to five different subgenera (Figure 2). Three of these taxa (*M. napina*, *M. pectinifera* F.A.C. Weber, and *M. solisioides* Backeb.) corresponded to the subgenus *Krainzia*; two (*M. baumii* Boed and *M. longimamma* DC.) to the subgenus *Dolichothele*; one (*M. senilis* Lodd. ex Salm-Dyck) to *Mammilliopsis*; one (*M. beneckei* Ehrenb.) to the subgenus *Oehmea*; and finally, one species (*M. zephyranthoides* Scheidw.) to *Phellosperma*. Currently, 21 of these 45 taxa are endemic only to one of the 13 biogeographical areas (right-side letters beside the taxa in Figure 4); with the Mexican Plateau the area that has the highest number of endemics (eight taxa). In addition, most of the divergent events occurred in three biogeographical areas, which was the unique ancestral area or in conjunction with other ones: the Mexican Plateau (A) involved 66 % of the divergence events, the Balsas Basin (F) 35%, and the Tehuacan Valley (M) 13%.

Furthermore, biogeographical results suggested that such divergence processes were closely associated to the taxa dispersal towards new areas inside and outside of the Mexican

Plateau (Figure 4). Accordingly, in the lineage 2 the long-distance dispersal has been a common phenomenon during the last ~4 million years (Figure 4). During these long-distance dispersal events, it seems that the ancestors moved out of the Mexican Plateau and eventually displaced along different routes, either via continental arid lands or crossing the sea (Gulf of Mexico and Gulf of California). During the Pliocene, we identified two independent events of colonization (nodes 151 and 196; Figure 4) to the arid southern Mexican territories (F, J and M; Figure 4), where the colonizers eventually speciated in situ (nodes 151 and 173; Figure 4). The first long-dispersal event occurred 4.13 Mya, and the second one was dated 3.45 Mya. These two events occurred during a slight increase of temperature (ΔT~1 °C) (Figure 4). These two colonization events took place from the Mexican Plateau (A) to the Balsas Basin (F), and from there to the adjacent areas of Tehuacan Valley (M) and Meridional Serranias (J) (Figure 4). In addition, we identified that during the Pleistocene, another two independent colonization events occurred towards the northern Mexican territories. The results indicate (Figure 4) that from the Mexican Plateau there was another dispersal route that took place along the foothills of Sierra Madre Occidental (A, K), crossing it, and reaching the Pacific slope of this Sierra. In addition, we identified a recent dispersal event dated nearly 1.11 Mya (node 189, Figure 4), in which an ancestor undertook a vicariant event, which separated two lineages, one of which diversified in the Baja California Peninsula (B) and the other in the northwest of continental Mexico (A, E) (Figure 4). On the other hand, on the eastern side of Mexico, another independent long-distance dispersal event was identified (node 178, Figure 4), crossing the Gulf of Mexico and reaching the Yucatán Peninsula nearly 0.16 Mya.

3.2.3. Evolutionary Lineage 3

We estimated the origin of the ancestor of this lineage (node 221, Figure 4) was in the Mexican Plateau (A) at 5.69 Mya (95% HPD: 3.72–7.89 Mya) (Figure 3). This lineage grouped 19 of the sampled taxa belonging to the genera *Coryphantha* (10), *Escobaria* (8), and *Pelecyphora* (1). Currently, 17 of these 19 taxa are distributed in the northern region of the Mexican Plateau (Figure 4), and only two taxa of *Coryphantha* are distributed in the southern arid lands (F, J and M, Figure 4), suggesting that a long-distance dispersal event allowed *Coryphantha* to reach the southern arid lands of Mexico (Figure 4). Therefore, in this lineage most of the past divergent processes were identified as on the Mexican Plateau, and eventually moving to northern and southern Mexico (Figure 4). The majority of these processes were dated to the Late Pliocene, when there was a slight temperature increase (ΔT~1 °C) (Figure 4). Clearly, the phylogenetic relationships of this clade were fully resolved. These results recovered *Coryphantha* as monophyletic, whereas *Escobaria* is paraphyletic with respect to *Coryphantha* and *P. strobiliformis* (Figure 2).

These findings showed that the lineage 3 was the phylogenetic sister of a clade comprising six lineages (lineages 4–9) (Figure 3) that evolved from a common ancestor (node 275, Figure 4). This ancestor was dated to nearly 5.60 Mya (95% HPD: 3.62–7.84; Figure 3) and arose in the ancestral arid lands of the Mexican Plateau (Figure 4).

3.2.4. Evolutionary Lineage 4

The lineage 4 was derived from an ancestor (node 225, Figure 4) dated nearly 4.02 Mya (95% HPD: 2.27–6.01 Mya), which had its ancestral geographical area on the Mexican Plateau (A) and the foothills of the Sierra Madre Occidental (K). We identified an early split that separated subgenus *Krainzia* (*M. theresae* Cutak, Figure 2) from *Phellosperma* (*M. barbata* Engelm. and *M. wrightii* Engelm.). In this last subgenus, a recent divergence (0.58 Mya, Figure 3) was identified; presently the taxa of this lineage are distributed in the northwestern territories of the Mexican Plateau, the Sierra Madre Occidental and the Northwest Coastal Plain (A, K and I, Figure 4).

3.2.5. Evolutionary Lineage 5

Our results revealed a common ancestor (node 274, Figure 4) that originated lineage 5 and the other four independent lineages identified as lineages 6–9 (Figure 3). This ancestor was dated to 5.24 Mya (95% HPD: 3.38–7.38 Mya) and probably arose in the ancestral lands of the Mexican Plateau (Figure 4). Although the phylogenetic split of the ancestor (node 274, Figure 4) was dated to nearly 5 Mya, the origin of lineage 5 is very recent, as it was dated to 0.63 Mya (95% HPD: 0.23–1.13 Mya) (Figure 3), when a slight temperature decrease (ΔT~0.5 °C) occurred (Figure 4). The ancestral geographical area of this lineage was also the Mexican Plateau (A, Figure 4). In addition, this lineage currently groups the two taxa recognized in the genus *Neolloydia*, which are distributed in the Mexican Plateau (Figure 4). However, *N. matehualensis* Backeb. is endemic to the center of the Mexican Plateau, and *N. conoidea* (DC) Britton & Rose ranges from the southern to the northern range of the Mexican Plateau and reaches the southern arid lands of the USA.

3.2.6. Evolutionary Lineage 6

Lineage 6 was composed of the unique species recognized in the genus *Ortegocactus*. This lineage was derived from an old ancestor (node 273, Figure 4) dated nearly 4.56 Mya (95% HPD: 2.9–6.45 Mya) in the Early Pliocene, when the cooling period ended (Figure 4). This ancestor had as its probable ancestral geographical areas the Mexican Plateau and Meridional Serranias (A and J, Figure 4), and presently this lineage is endemic to the Meridional Serranias (J, Figure 4). Lastly, the results showed that the two sampled specimens of *O. macdougallii* Alexander recently diverged about 51,000 years ago (Figure 3).

3.2.7. Evolutionary Lineage 7

The ancestor of lineage 7 (node 229, Figure 4) was dated to 1.46 Mya (95% HPD: 0.6–2.48 Mya), during a time when the temperature increased slightly (ΔT~0.5 °C), and for this were estimated four probable ancestral areas (A, B, C and I, Figure 4). This lineage consists of only two northern native taxa; *M. guelzowiana* Werderm., which is endemic to the northwestern part of the continental Mexican territories; and *M. tetrancistra* Engelm. that is distributed in Baja California (B) and California (C), northwestern continental Mexican territories (I), and reaches the southern USA.

3.2.8. Evolutionary Lineage 8

Lineages 8 and 9 had a common ancestor (node 271, Figure 4) that arose in Baja California 3.1 Mya (95% HPD: 1.9–4.43 Mya) (B, Figure 4). In particular, the immediate ancestor of lineage 8 (node 238, Figure 4) was dated to 1.14 Mya (0.52–1.91 Mya) during the Pleistocene, concurrently with a small increase of temperature (ΔT~1 °C) (Figure 4). Lineage 8 grouped three taxa (*M. halei* Brandegee, *M. pondii* Greene, and *M. poselgeri* Hildm.) belonging to the *Cochemiea* subgenus.

3.2.9. Evolutionary Lineage 9

Lineage 9 grouped 16 taxa, all pertaining to the subgenus *Chilita* (Figure 2). Its immediate ancestor was dated to 2.77 Mya (95% HPD: 1.67–3.97 Mya) and its most probable ancestral area was Baja California (B, Figure 4). This lineage developed in the Late Pliocene when there was a slight increase in temperature (ΔT~1 °C). It diversified abundantly in Baja California. In this lineage, we identified two independent dispersal events from peninsular territories to the continental Northwest Coastal Plain (I, Figure 4). One of them occurred 1.82 Mya (node 268, Figure 4) and the other was dated to 0.01 Mya (node 239, Figure 4).

4. Discussion

4.1. Origin and Diversification of the Mammilloid Clade

The findings of this study revealed that the evolutionary history of the Mammilloid Clade (*Coryphantha*, *Escobaria*, *Mammillaria*, *Mammilloydia*, *Neolloydia*, *Ortegocactus*, and *Pelecyphora*) started ~7.5 Mya in the Miocene. During this epoch, there was a cooling

trend, although the global temperature was still approximately 4–15 °C warmer than it is today [26]. The early and scarce divergence events that occurred in the Miocene were geographically restricted to the Mexican Plateau. However, during the last 4.5 million years, the cacti profusely diversified and expanded their distribution range to new areas when the global temperature was more similar to the present. Particularly, the main past colonization to new geographical areas (e.g., California, Northwest coastal plain, Pacific coast, Tehuacán Valley, and Yucatan peninsula) were dated to the last ~2.5 million years, in the Pleistocene. During this epoch various oscillations in temperature occurred [56] and have been associated with an aridity increase (e.g., [57,58]). Consequently, these cacti are modern taxa, with most of their evolutionary history occurring during the Plio-Pleistocene. In fact, the climatic oscillations in the Pleistocene were recognized as diversification driving forces for other land plants (e.g., [59–62]). Particularly for cacti, glacial [63,64] and interglacial [65,66] periods have been proposed as drivers of population processes, causing geographic contraction, isolation, and population divergence. Therefore, probably these climatic oscillations also promoted the diversification of the cacti studied here.

The Mexican Plateau has been considered geologically and climatically stable since ~15 Mya (Middle Miocene) [67]. Hence, we consider that such stability promoted the prolific speciation and colonization of cacti. However, there is a gradient of aridity along the Mexican Plateau, with the northern portion being drier than the central–southern one [68]. As cacti do not prosper in hyper-arid conditions [69], the relative "higher-humidity" at the southern end of the Mexican Plateau likely foster the ecological conditions for their abundance and speciation, which eventually led to geographic expansion. Accordingly, we postulated that the center of origin for the lineages 1 and 2 was the southern region of the Mexican Plateau, which previously was named as the Queretano-Hidalguense arid zone [70]. We based this hypothesis on the early phylogenetic split identified in these two lineages, and on their current geographical distribution. Consistent with this assumption, nearly 20% of the richness of the genus *Mammillaria* (sensu Hunt [18]) inhabit the arid lands of the Queretano-Hidalguense arid zone (Hidalgo, Guanajuato, and Querétaro) [33].

Based on similar reasoning, we inferred that the possible center of origin of lineage 3 might be the north of the Mexican Plateau. However, we recognized that more extensive taxonomic sampling is necessary to elucidate this issue. On the other hand, our results revealed that the taxa grouped in lineages 4, 7, 8 and 9 had an ecological and biogeographical affinity to northwestern Mexico. Considering that the Baja California area was the probable ancestral geographical area of lineages 8 and 9, these results suggest that this area was probably the center of origin and diversification for these lineages. Lastly, we do not discard the possibility that the small and enigmatic lineages 5 and 6 represent relicts of some phylogenetic lines that are mostly extinct. Population approaches may serve to elucidate their closest phylogenetic frontiers and recent hybridization (e.g., [17,71]). Therefore, we recommend application of this perspective to lineages 5 and 6, as well as for the sister lineages 1 and 2.

Our results also revealed that dispersal, not vicariance, was the most important past biogeographical process in these cacti. The abundant dispersal events may be related to the capacity of cacti to colonize and tolerate hostile environments (e.g., [72]), or successful seed dispersal strategies [73]. However, the data indicate that also vicariance had a relevant role in the taxa that currently are distributed in southern Mexico. Because the central portion of the Trans-Mexican Volcanic Belt (TMVB) was formed during the last three million years (plate 1 [74]), we address the TMVB as a biogeographical barrier for cacti. In fact, most of the events of colonization to southern Mexican territories were identified prior to 3 million years ago, thus the TMVB interrupted the connectivity between the arid lands of the Mexican Plateau and those of the Balsas Basin, Tehuacán Valley, and southern Meridional Serranias. In addition, the floristic affinities between the arid lands of the north and south of Mexico have been documented [53], suggesting that prior to the TMVB, the Mexican arid lands were connected from north to the south. Finally, our results showed that the arrival of cacti to the Baja California peninsula was due to dispersal and not by vicariance, as

the colonization occurred later than the opening of the Gulf of California, which occurred nearly 12–6 Mya [75].

4.2. Taxonomic Contributions of the Phylogenetic Results

The findings of this study have explained the phylogenetic relationships of the 103 taxa, particularly the 70 taxa sampled from the genus *Mammillaria* sensu Hunt [18] were polyphyletic, as was identified previously (e.g., [35,36]). However, based on our results, the monophyly of this genus can be identified within a subset of the 70 taxa sampled as *Mammillaria*. We consider that the putative genus *Mammillaria* is represented by lineage 2, in which 85% of the taxa were from subgenus *Mammillaria*. Accordingly, monophyletic *Mammillaria* is not restricted exclusively to the *Mammillaria* subgenus, as Crozier [36] proposed, but also includes taxa of another five subgenera: *Dolichothele*, *Krainzia*, *Mammillopsis*, *Oehmea*, and *Phellosperma*. Recently, Breslin et al. [38] proposed a monophyletic circumscription of the genus *Mammillaria*, based on massive sequencing of the chloroplast genome and 52 taxa assumed to be from members of the genus. Because these 52 taxa exhibited polyphyletic relationships, these authors decided to exclude a substantial number of them in order to reach a monophyletic group.

In addition, Breslin et al. [38] proposed that the 36 taxa of the genus *Mammillaria* that were placed out of the monophyletic group, should be placed in the genus *Cochemiea* together with *N. conoidea* and *O. macdougallii*, although the species of these genera exhibit strong morphological variation (Table S1) [21]. Our results showed that lineages 4 to 9 were grouped in a distinct clade, independent of the clade that grouped lineages 1 and 2. These six lineages composed a monophyletic group (lineages 4–9). Although, these six lineages were grouped similarly to the clade named as *Cochemiea* by Breslin et al. [38] we do not agree to put together the taxa of these six lineages as our results showed strong disparities in the biogeographical history and ecologic affinities. Additionally, their strong morphological variations do not accomplish the unambiguous practical delimitation (i.e., taxonomic predictability) and stability that are required at the genus level [76]. We consider that based on a purely phylogenetic perspective, the proposal of Breslin et al. [38] to include in *Cochemiea* other taxa recognized as *Mammillaria*, *Neolloydia* and *Ortegocactus* is feasible. However, our results identified six lineages in the clade *Cochemiea* sensu Breslin et al. [38], and for us these may represent more than one genus: *Cochemiea* (lineages 7, 8 and 9), *Neolloydia* (5), *Ortegocactus* (6), and *Phellosperma* (4). These two contrasting stances exhibit the degree of subjectivity to establish the supraspecific taxonomic delimitation as has been discussed [76]. We consider that future phylogenetic studies are still necessary, and they must include specimens of the type *M. mammillaris*, and have a higher taxonomic sampling, especially of those taxa that are currently distributed in the west side of Mexican territories along the Pacific Coast. Consequently, we considered that the taxonomic circumscription of *Mammillaria* still remains unresolved. Lastly, our phylogenetic results partially supported the infrageneric classification of *Mammillaria* proposed by Hunt [18]. Accordingly, the taxa of the subgenera *Cochemiea* and *Chilita* were monophyletic. Although all the taxa of the subgenus *Mammillaria* were grouped, the monotypic and small subgenera (*Dolichothele*, *Mammillopsis* and *Oehmea*) as well as some taxa of *Phellosperma* and *Krainzia* were inserted among the species of the *Mammillaria* subgenus. In addition, because we included the raw data attached to Breslin et al. [38], we observed that in our phylogenetic tree, some of their specimens belonging to the same species were placed in discordant positions (*M. grahamii* subsp. *sheldonii* (Britton & Rose) D.R. Hunt (35158, 35161), *M. goodridgii* (35106, 35115, 35167), *M. albicans* Dawson (35107, 35103), *M. armillata* K.Brandegee (35093, 35144, 35089), *M. dioica* K.Brandegee (35170, 35131, 35119), and *M. heyderi* Muehlenpf. (16460)); this indicates the probable wrong taxonomic identification of their specimens, thus we overlooked these for the taxonomic discussion.

Recently, the study of Sánchez et al. [77] obtained a phylogenetic tree based on five chloroplast loci and eight morphologic characters. It showed that *Coryphantha* was a monophyletic genus, when excluding *C. macromeris* (Engelm.) Lem. Moreover, this last

taxon and the taxa of *Escobaria* and *Pelecyphora* were grouped in the same clade, sister to *Coryphantha*, and were reclassified as a single genus (*Pelecyphora*). Our study also showed that *Coryphantha* is a monophyletic taxon, whereas *Escobaria* is paraphyletic with respect to *Coryphantha* and *Pelecyphora*. These discordances may be the result of distinct taxonomic and molecular sampling between the two studies. Nonetheless, it may be necessary to analyze morphological, ecological, and anatomical characters in order to solve these taxonomic issues.

5. Conclusions

We identified that the biogeographical processes, past climate conditions from the Miocene, and the recent emergence of the central portion of the TMVB strongly shaped the evolutionary history of the Mammilloid Clade (*Coryphantha*, *Escobaria*, *Mammillaria*, *Mammilloydia*, *Neolloydia*, *Ortegocactus*, and *Pelecyphora*). The past Mexican arid lands were key to providing ecological suitability for prolific cacti diversification. In these regions, they became abundant and ubiquitous elements of the arid flora. The large Mexican Plateau has been the primary evolutionary scenario for cacti, and this area is key to understand the diversity of cacti in Mexico, southern USA, Caribbean, and South America. Lastly, the Mexican territories harbor most of the world's richness of cacti, and it is urgent to protect these arid lands, particularly the region included in the northern part of Guanajuato, Hidalgo, and Querétaro, and southern of San Luis Potosí. Our findings indicate that the genus *Mammillaria* sensu Hunt [18] is taxonomically composed of distinct evolutionary lineages, whose phylogenetic relationships require more detailed studies to reach a precise taxonomic circumscription. In this light, we consider that it is premature to undertake nomenclatural changes in *Mammillaria*, *Mammilloydia*, *Neolloydia*, and *Ortegocactus* [38,78], and such changes will bring more confusion. Therefore, we recommend maintaining the conventional taxonomic classifications (e.g., [18]) until more robust studies are undertaken. In summary, we conclude that the taxonomic circumscription of the genus *Mammillaria* still needs more work, based on phylogenetic analyses encompassed with robust and detailed ecological studies of the current geographical distribution, past niche modeling, reproductive barriers, and a clear set of diagnostic morphological characters.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/biology12040512/s1. Table S1: List and origin of the studied taxa; File S1: Geographical distribution of the 70 sampled taxa of *Mammillaria*; Table S2: List of 52 loci used for phylogenomic analysis; File S2: Output of the evaluation of six biogeographical models with BioGeoBEARS implemented in RASP4; Figure S1: Full results of ancestral geographical distribution estimated under the DEC model. Figure S2: Results of ancestral geographical distribution estimated under the DEC+J model.

Author Contributions: This article was conceptualized by S.S. as part of the Research Project IN228619 supported by PAPIIT-DGAPA; moreover, she supervised and administered the resources of this research, and got the funding resources. D.A.C. and S.S. planned, discussed, and agreed on the analytical approaches used in this research; D.A.C. carried out in full the statistical formal analysis. A partial support in massive sequencing was provided by F.V.-P.; S.A. contributed with tissue samples for most of de novo sequenced samples. D.A.C., S.S., F.V.-P. and P.D. wrote and prepared the original draft. All authors reviewed, commented, edited the last version of this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by UNAM-PAPIIT-IN228619.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data of the 49 samples de novo sequenced are available in the NCBI database under BioProject PRJNA934337. The SRA accessions for each sample are shown in Table S1.

Acknowledgments: This study was fully granted by UNAM-PAPIIT-IN228619. This article is part of the requirements for obtaining the doctoral degree of D.A.C. at the Posgrado en Ciencias Biológicas (Evolutionary Biology), Universidad Nacional Autónoma de México; she was funded by a grant from the Consejo Nacional de Ciencia y Tecnología CONACyT (855061). The sampling permissions were authorized to S.S. by SEMARNAT SGPA/DGVS/06880/16 for *Mammillaria huitzilopochtli* and SGP/DGVS/01833/08 for *M. napina*; whereas the rest of the studied species were provided from the cacti collection of the Botanical Garden of Instituto de Biología, UNAM. The massive sequencing was carried out by Clara E. Díaz-Velázquez and Aldo Hugo De La Cruz Montoya technicians of the Laboratorio Nacional en Salud: Diagnóstico Molecular y Efecto Ambiental en Enfermedades Crónico-Degenerativas, FES Iztacala, UNAM.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Ronquist, F.; Sanmartín, I. Phylogenetic methods in biogeography. *Annu. Rev. Ecol. Evol. Syst.* **2011**, *42*, 441–464. [CrossRef]
2. Xia, M.; Liu, Y.; Liu, J.; Chen, D.; Shi, Y.; Chen, Z.; Chen, D.; Jin, R.; Chen, H.; Peter, C.H.; et al. Out of the Himalaya-Hengduan Mountains: Phylogenomics, biogeography and diversification of *Polygonatum* Mill. (Asparagaceae) in the Northern Hemisphere. *Mol. Phylogenet. Evol.* **2022**, *169*, 107431. [CrossRef] [PubMed]
3. Rose, J.P.; Xiang, C.L.; Sytsma, K.J.; Drew, B.T. A timeframe for mint evolution: Towards a better understanding of trait evolution and historical biogeography in Lamiaceae. *Bot. J. Linn. Soc.* **2022**, *200*, 15–38. [CrossRef]
4. Liu, S.Y.; Zhu, H.; Yang, J. A phylogenetic perspective on biogeographical divergence of the flora in Yunnan, Southwestern China. *Sci. Rep.* **2017**, *7*, 43032. [CrossRef]
5. Qian, H.; Deng, T.; Ricklefs, R.E. Evolutionary assembly of the Arctic flora. *Glob. Ecol. Biogeogr.* **2022**, *31*, 396–404. [CrossRef]
6. Barros-Souza, Y.; Borges, L.M. Spatial-and lineage-dependent processes underpin floristic assembly in the megadiverse Eastern South American mountains. *J. Biogeogr.* **2023**, *50*, 302–315. [CrossRef]
7. Chen, L.Y.; Zhao, S.Y.; Mao, K.S.; Les, D.H.; Wang, Q.F.; Moody, M.L. Historical biogeography of Haloragaceae: An out-of-Australia hypothesis with multiple intercontinental dispersals. *Mol. Phylogenet. Evol.* **2014**, *78*, 87–95. [CrossRef]
8. Nge, F.J.; Kellermann, J.; Biffin, E.; Waycott, M.; Thiele, K.R. Historical biogeography of *Pomaderris* (Rhamnaceae): Continental vicariance in Australia and repeated independent dispersals to New Zealand. *Mol. Phylogenet. Evol.* **2021**, *158*, 107085. [CrossRef]
9. Liu, X.Q.; Xia, X.M.; Chen, L.; Wang, X.Q. Phylogeny and evolution of Cupressaceae: Updates on intergeneric relationships and new insights on ancient intergeneric hybridization. *Mol. Phylogenet. Evol.* **2022**, *177*, 107606. [CrossRef]
10. Heinrichs, J.; Hentschel, J.; Feldberg, K.; Bombosch, A.; Schneider, H. Phylogenetic biogeography and taxonomy of disjunctly distributed bryophytes. *J. Syst. Evol.* **2009**, *47*, 497–508. [CrossRef]
11. Oyston, J.W.; Wilkinson, M.; Ruta, M.; Wills, M.A. Molecular phylogenies map to biogeography better than morphological ones. *Commun. Biol.* **2022**, *5*, 521. [CrossRef] [PubMed]
12. Svenning, J.-C.; Eiserhardt, W.L.; Normand, S.; Ordonez, A.; Sandel, B. The influence of paleoclimate on present-day patterns in biodiversity and ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **2015**, *46*, 551–572. [CrossRef]
13. Li, Z.Z.; Lehtonen, S.; Martins, K.; Gichira, A.W.; Wu, S.; Li, W.; Hu, G.W.; Liu, Y.; Zou, C.Y.; Wang, Q.F.; et al. Phylogenomics of the aquatic plant genus *Ottelia* (Hydrocharitaceae): Implications for historical biogeography. *Mol. Phylogenet. Evol.* **2020**, *152*, 106939. [CrossRef] [PubMed]
14. Lai, Y.J.; Li, S.J.; Wang, W.M. Evolutionary trends in leaf morphology and biogeography of Altingiaceae based on fossil evidence. *Palaeoworld* **2018**, *27*, 415–422. [CrossRef]
15. Ding, W.N.; Ree, R.H.; Spicer, R.A.; Xing, Y.W. Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science* **2020**, *369*, 578–581. [CrossRef]
16. Antonelli, A.; Nylander, J.A.A.; Persson, C.; Sanmartín, I. Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9749–9754. [CrossRef] [PubMed]
17. Hewitt, G.M. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.* **1996**, *58*, 247–276. [CrossRef]
18. Hunt, D.R. *The New Cactus Lexicon*; DH Books: Milborne Port, UK, 2006.
19. Nyffeler, R.; Eggli, U. Disintegrating Portulacaceae: A new familial classification of the suborder Portulacineae (Caryophyllales) based on molecular and morphological data. *Taxon* **2010**, *59*, 227–240. [CrossRef]
20. Walker, J.F.; Yang, Y.; Feng, T.; Timoneda, A.; Mikenas, J.; Hutchison, V.; Edwards, C.; Wang, N.; Ahluwalia, S.; Olivieri, J.; et al. From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *Am. J. Bot.* **2018**, *105*, 446–462.
21. Magallon, S.; Gomez-Acevedo, S.; Sanchez-Reyes, L.L.; Hernandez-Hernandez, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **2015**, *207*, 437–453. [CrossRef]

22. Hernandez-Hernandez, T.; Brown, J.W.; Schlumpberger, B.O.; Eguiarte, L.E.; Magallon, S. Beyond aridification: Multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytol.* **2014**, *202*, 1382–1397. [CrossRef] [PubMed]
23. Arakaki, M.; Christin, P.A.; Nyffeler, R.; Lendel, A.; Eggli, U.; Ogburn, M.; Spriggs, E.; Moore, M.J.; Edwards, E.J. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8379–8384. [CrossRef]
24. Pearson, P.N.; Palmer, M.R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **2000**, *406*, 695–699. [CrossRef] [PubMed]
25. Tripati, A.K.; Roberts, C.D.; Eagle, R.A. Coupling of CO_2 and Ice Sheet Stability Over Major Climate Transitions of the Last 20 Million Years. *Science* **2009**, *326*, 1394–1397. [CrossRef]
26. Herbert, T.D.; Lawrence, K.T.; Tzanova, A.; Peterson, L.C.; Caballero-Gill, R.; Kelly, C.S. Late Miocene global cooling and the rise of modern ecosystems. *Nat. Geosci.* **2016**, *9*, 843. [CrossRef]
27. Anderson, E.F. *The Cactus Family*; Timber: Portland, OR, USA, 2001.
28. The IUCN Red List of Threatened Species. *Version 2022-1*. Available online: https://www.iucnredlist.org/search?taxonomies=11456&searchType=species (accessed on 1 November 2022).
29. Haworth, A.H. *Synopsis Plantarum Succulentarum*; Richard Taylor: London, UK, 1812.
30. Schumann, K. *Gesamtbeschreibung der Kakteen (Monographia Cactearum)*; Verlag von J. Neumann: Neudamm, Germany, 1898; pp. 472–601.
31. Hunt, D.R. Schumann and Buxbaum reconciled. *Cact. Succ. J. Gr. Brit.* **1971**, *33*, 53–72.
32. Lüthy, J.M. Taxonomische Untersuchung der Gattung *Mammillaria* Haw. Ph.D. Thesis, Universität Bern, Bern, Switzerland, 1995.
33. Hernández, H.M.; Gómez-Hinostrosa, C. *Mapping the Cacti of Mexico Part II: Mammillaria*; DH Books: Milborne Port, UK, 2015.
34. Butterworth, C.A.; Cota-Sanchez, J.H.; Wallace, R.S. Molecular systematics of tribe Cacteae (Cactaceae: Cactoideae): A phylogeny based on rpl16 intron sequence variation. *Syst. Bot.* **2002**, *27*, 257–270.
35. Butterworth, C.A.; Wallace, R.S. Phylogenetic studies of *Mammillaria* (Cactaceae)—Insights from chloroplast sequence variation and hypothesis testing using the parametric bootstrap. *Am. J. Bot.* **2004**, *91*, 1086–1098. [CrossRef] [PubMed]
36. Crozier, B.S. Systematics of Cactaceae Juss: Phylogeny, cpDNA Evolution, and Classification, with Emphasis on the Genus *Mammillaria* Haw. Ph.D. Thesis, University of Texas, Austin, TX, USA, 2005.
37. Hunt, D.R. Revised classified list of the genus *Mammillaria*. *Cact. Succ. J. Gr. Brit.* **1981**, *43*, 41–48.
38. Breslin, P.B.; Wojciechowski, M.F.; Majure, L.C. Molecular phylogeny of the Mammilloid clade (Cactaceae) resolves the monophyly of *Mammillaria*. *Taxon* **2021**, *70*, 308–323. [CrossRef]
39. Solórzano, S.; Chincoya, D.A.; Sanchez-Flores, A.; Estrada, K.; Díaz-Velásquez, C.E.; González-Rodríguez, A.; Vaca-Paniagua, F.; Dávila, P.; Arias, S. De novo assembly discovered novel structures in genome of plastids and revealed divergent inverted repeats in *Mammillaria* (Cactaceae, Caryophyllales). *Plants* **2019**, *8*, 392. [CrossRef] [PubMed]
40. Aiello-Lammens, M.E.; Boria, R.A.; Radosavljevic, A.; Vilela, B.; Anderson, R.P.; Bjornson, R.; Weston, S. spThin: Functions for Spatial Thinning of Species Occurrence Records for Use in Ecological Models. 2019. Available online: https://cran.r-project.org/web/packages/spThin/index.html (accessed on 15 March 2023).
41. Babraham Bioinformatics. Trim Galore. Available online: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed on 21 April 2022).
42. Jin, J.J.; Yu, W.B.; Yang, J.B.; Song, Y.; DePamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **2020**, *21*, 241. [CrossRef] [PubMed]
43. Zhang, Z.; Schwartz, S.; Wagner, L.; Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **2000**, *7*, 203–214. [CrossRef]
44. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef]
45. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.; Von Haeseler, A.; Jermiin, L.S. Model Finder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]
46. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A.; Lanfear, R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **2020**, *7*, 1530–1534. [CrossRef]
47. Bouckaert, R.; Vaughan, T.G.; Barido-Sottani, J.; Duchêne, S.; Fourment, M.; Gavryushkina, A.; Heled, J.; Jones, G.; Kühnert, D.; De Maio, N. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **2019**, *15*, e1006650. [CrossRef]
48. Posada, D.; Crandall, K.A. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **1998**, *14*, 817–818. [CrossRef]
49. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [CrossRef]
50. Drummond, A.J.; Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **2007**, *7*, 214. [CrossRef]
51. Guzmán-Cruz, L.U.; Arias-Montes, S.; Dávila, P. *Catálogo de Cactáceas Mexicanas*; Universidad Nacional Autónoma de México, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad: Ciudad de México, Mexico, 2007.
52. Dicht, R.; Lüthy, A. *Coryphantha: Cacti of Mexico and Southern USA*; Springer: Berlin, Germany, 2005.
53. Rzedowski, J. *Vegetación de México*; Limusa: Ciudad de México, Mexico, 1978.

54. Matzke, N. Probabilistic historical biogeography: New models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.* **2015**, *5*, 242–248.
55. Yu, Y.; Blair, C.; He, X. RASP 4: Ancestral state reconstruction tool for multiple genes and characters. *Mol. Biol. Evol.* **2020**, *37*, 604–606. [CrossRef] [PubMed]
56. Lawrence, K.T.; Sosdian, S.; White, H.E.; Rosenthal, Y. North Atlantic climate evolution through the Plio-Pleistocene climate transitions. *Earth Planet. Sci. Lett.* **2010**, *300*, 329–342. [CrossRef]
57. Vuilleumier, B.S. Pleistocene changes in the fauna and flora of South America. *Science* **1971**, *173*, 771–780. [CrossRef] [PubMed]
58. Hoag, C.; Svenning, J.C. African environmental change from the Pleistocene to the Anthropocene. *Annu. Rev. Environ. Resour.* **2017**, *42*, 27–54. [CrossRef]
59. Janssens, S.B.; Knox, E.B.; Huysmans, S.; Smets, E.F.; Merckx, V.S. Rapid radiation of *Impatiens* (Balsaminaceae) during pliocene and pleistocene: Result of a global climate change. *Mol. Phylogenet. Evol.* **2009**, *52*, 806–824. [CrossRef]
60. DeChaine, E.G.; Wendling, B.M.; Forester, B.R. Integrating environmental, molecular, and morphological data to unravel an ice-age radiation of arctic-alpine Campanula in western North America. *Ecol. Evol.* **2014**, *4*, 3940–3959. [CrossRef]
61. Dorsey, B.L.; Gregory, T.J.; Sass, C.; Specht, C.D. Pleistocene diversification in an ancient lineage: A role for glacial cycles in the evolutionary history of *Dioon* Lindl. (Zamiaceae). *Am. J. Bot.* **2018**, *105*, 1512–1530. [CrossRef]
62. Nevado, B.; Contreras-Ortiz, N.; Hughes, C.; Filatov, D.A. Pleistocene glacial cycles drive isolation, gene flow and speciation in the high-elevation Andes. *New Phytol.* **2018**, *219*, 779–793. [CrossRef]
63. Cornejo-Romero, A.; Vargas-Mendoza, C.F.; Aguilar-Martínez, G.F.; Medina-Sánchez, J.; Rendón-Aguilar, B.; Valverde, P.L.; Zavala-Hurtado, J.A.; Serrato, A.; Rivas-Arancibia, S.; Pérez-Hernández, M.A.; et al. Alternative glacial-interglacial refugia demographic hypotheses tested on *Cephalocereus columna-trajani* (Cactaceae) in the intertropical Mexican drylands. *PLoS ONE* **2017**, *12*, e0175905. [CrossRef]
64. Ossa, C.G.; Montenegro, P.; Larridon, I.; Pérez, F. Response of xerophytic plants to glacial cycles in southern South America. *Ann. Bot.* **2019**, *124*, 15–26. [CrossRef] [PubMed]
65. Bonatelli, I.A.; Perez, M.F.; Peterson, A.T.; Taylor, N.P.; Zappi, D.C.; Machado, M.C.; Koch, I.; Pires, A.H.C.; Moraes, E.M. Interglacial microrefugia and diversification of a cactus species complex: Phylogeography and palaeodistributional reconstructions for *Pilosocereus aurisetus* and allies. *Mol. Ecol.* **2014**, *23*, 3044–3063. [CrossRef] [PubMed]
66. Ornelas, J.F.; Rodríguez-Gómez, F. Influence of Pleistocene glacial/interglacial cycles on the genetic structure of the Mistletoe Cactus *Rhipsalis baccifera* (Cactaceae) in Mesoamerica. *J. Hered.* **2015**, *106*, 196–210. [CrossRef]
67. Morafka, D.J. A historical biogeography of the Chihuahuan herpetofauna. In *A Biogeographical Analysis of the Chihuahuan Desert through Its Herpetofauna*; Dr. W. Junk: The Hague, The Netherlands, 1977; pp. 157–215.
68. UNESCO. *ATLAS of Arid Zones in Latin America and the Caribbean*; PHI UNESCO: Paris, France, 2010.
69. Guerrero, P.C.; Durán, A.P.; Walter, H.E. Latitudinal and altitudinal patterns of the endemic cacti from the Atacama Desert to Mediterranean Chile. *J. Environ.* **2011**, *75*, 991–997. [CrossRef]
70. Medellín-Leal, F. The Chihuahuan Desert. In *Reference Handbook on the Deserts of North America*; Bender, G.L., Ed.; Greenwood Press: Westfield, CT, USA, 1982; pp. 321–381.
71. Jürgens, N.; Oncken, I.; Oldeland, J.; Gunter, F.; Rudolph, B. *Welwitschia*: Phylogeography of a living fossil, diversified within a desert refuge. *Sci. Rep.* **2021**, *11*, 114. [CrossRef]
72. Puente, M.E.; Li, C.Y.; Bashan, Y. Rock-degrading endophytic bacteria in cacti. *Environ. Exp. Bot.* **2009**, *66*, 389–401. [CrossRef]
73. Bregman, R. Forms of seed dispersal in Cactaceae. *Acta Bot. Neerl.* **1988**, *37*, 395–402. [CrossRef]
74. Gómez-Tuena, A.; Orozco-Esquivel, M.T.; Ferrari, L. Igneous petrogenesis of the Trans-Mexican volcanic belt. *Geo. Soc. Amer. Spec.* **2007**, *422*, 129–181.
75. Ledesma-Vázquez, J.; Johnson, M.E.; Gonzalez-Yajimovich, O.; Santamaría-del-Angel, E. Gulf of California geography, geological origins, oceanography, and sedimentation patterns. In *Atlas of Coastal Ecosystems in the Western Gulf of California*; Johnson, M.E., Ledesma-Vázquez, J., Eds.; University Arizona Press: Tucson, Arizona, 2009; pp. 1–10.
76. Humphreys, A.M.; Linder, H.P. Concept versus data in delimitation of plant genera. *Taxon* **2009**, *58*, 1054–1074. [CrossRef]
77. Sánchez, D.; Vázquez-Benítez, B.; Vázquez-Sánchez, M.; Aquino, D.; Arias, S. Phylogenetic relationships in *Coryphantha* and implications on *Pelecyphora* and *Escobaria* (Cacteae, Cactoideae, Cactaceae). *PhytoKeys* **2022**, *188*, 115–165. [CrossRef]
78. Korotkova, N.; Aquino, D.; Arias, S.; Eggli, U.; Franck, A.; Gómez-Hinostrosa, C.; Guerrero, P.C.; Hernández, H.M.; Kohlbecker, A.; Köhler, M.; et al. Cactaceae at Caryophyllales.Org- A Dynamic Online Species-Level Taxonomic Backbone for the Family. *Willdenowia* **2021**, *51*, 251–270. [CrossRef]

CAPÍTULO 2. Effectiveness of Two Universal Angiosperm Probe Sets Tested In Silico for Caryophyllids Taxa with Emphasis on Cacti Species

(Artículo publicado)

genes

MDPI

Communication

Effectiveness of Two Universal Angiosperm Probe Sets Tested In Silico for Caryophyllids Taxa with Emphasis on Cacti Species

Delil A. Chincoya [1,2,*] and Sofía Solórzano [1,*]

[1] Laboratorio de Ecología Molecular y Evolución, UBIPRO, FES Iztacala, Universidad Nacional Autónoma de México, Avenida de los Barrios 1, Los Reyes Iztacala, Tlalnepantla de Baz, Estado de México 54090, Mexico
[2] Posgrado en Ciencias Biológicas, UNAM, FES Iztacala, Circuito de Posgrados, Ciudad Universitaria, Coyoacán, Ciudad de México 04510, Mexico
* Correspondence: dela@comunidad.unam.mx (D.A.C.); solorzanols@unam.mx (S.S.)

Abstract: In angiosperms, huge advances in massive DNA sequencing technologies have impacted phylogenetic studies. Probe sets have been developed with the purpose of recovering hundreds of orthologous loci of targeted DNA sequences (TDS) across different plant lineages. We tested in silico the effectiveness of two universal probe sets in the whole available genomes of Caryophyllids, emphasizing phylogenetic issues in cacti species. A total of 870 TDS (517 TDS from Angiosperm v.1 and 353 from Angiosperms353) were individually tested in nine cacti species and *Amaranthus hypochondriacus* (external group) with ≥17 Gbp of available DNA data. The effectiveness was measured by the total number of orthologous loci recovered and their length, the percentage of loci discarded by paralogy, and the proportion of informative sites (PIS) in the alignments. The results showed that, on average, Angiosperms353 was better than Angiosperm v.1 for cacti species, since the former obtained an average of 275.6 loci that represent 123,687 bp, 2.48% of paralogous loci, and 4.32% of PIS in alignments, whereas the latter recovered 148.4 loci (37,683 bp), 10.38% of paralogous loci, and 3.49% of PIS. We recommend the use of predesigned universal probe sets for Caryophyllids, since these recover a high number of orthologous loci that resolve phylogenetic relationships.

Keywords: cactaceae; caryophyllids; phylogenomics; target enrichment; universal probe sets

Citation: Chincoya, D.A.; Solórzano, S. Effectiveness of Two Universal Angiosperm Probe Sets Tested In Silico for Caryophyllids Taxa with Emphasis on Cacti Species. *Genes* **2022**, *13*, 570. https://doi.org/10.3390/genes13040570

Academic Editor: Pablo C. Guerrero

Received: 23 February 2022
Accepted: 21 March 2022
Published: 24 March 2022

1. Introduction

High-throughput DNA sequencing technologies (HTDNAs) significantly improve denser molecular sampling of DNA than Sanger-based capillarity technology. This increment of whole genomes published either as fully assembled and annotated genomes, or as raw data with poor or null informatics processing, has positively impacted advances in examining fundamental questions in flowering land plants. Fortunately, most of these massive DNA data are being deposited in free-access digital reservoirs, allowing the world's scientific community to use them freely. In the particular case of angiosperms, most problems studied with HTDNAs are those of a phylogenetic nature. Presently, the term phylogenomics is used when a high number of loci (i.e., dozens and even hundreds) from DNA/RNA are sampled in a single or in various genomes (i.e., chloroplast, mitochondria, and nucleus) contained in plants. In land-flowering plants, phylogenomic analysis based on chloroplast genomes [1,2] and transcriptomes [3] has allowed scientists to resolve the relationships of early divergent lineages of angiosperms. The different genome-scale analyses converged in a topology of the species tree in which three orders—Amborellales (Amborellaceae, *Amborella trichopoda*), Nymphaeales (Cabombaceae, Hydatellaceae, and Nymphaeaceae), and Austrobaileyales (Austrobaileyaceae, Schisandraceae and Trimeneaceae)—were, successively, the sister lineages to the whole-monophyletic speciose group of angiosperms [1–3]. Recently, the old problem of determining the origin of angiosperms was reexamined in the broader phylogenetic study of Li et al. [4]. This study integrated previously published data obtained using HTDNAs, as well as other

taxa de novo sequenced by the authors, in order to obtain 80 loci of the genome of the chloroplasts (>82,286 bp) of 2351 ingroup flowering land species. This study concluded that angiosperms emerged in the early Jurassic, which conflicts with the pollen fossil records data that indicate that angiosperms arose in the Cretaceous period. The results of this study revamped an old discussion with those fossil records defenders [5], and it is a fact that massive sequencing data will alter our views of the evolution of the biodiversity of our planet; they could even break scientific paradigms.

In addition, HTDNAs are also fertile tools for those scientific questions focused on smaller scales, either focused on the evolutionary divergence processes that occurred in particular flowering groups or in the biota evolution of local areas. Moreover, phylogenomic sampling may drastically increase the phylogenetic resolution yet at lower taxonomic hierarchical levels and in those plant groups that recently diverged [6–8]. The benefits of having a fully resolved phylogenetic tree species impact many areas: the inherent value is that phylogenetic relationships are clarified—often, in a phylogenetic tree, the diverging process may be inferred, which eventually may help to understand patterns of richness around the planet at different taxonomic hierarchical levels (i.e., orders, families, and genera). Furthermore, a fully resolved phylogeny may support a solution to those pendant taxonomic problems and the species boundaries can be clearly delimited.

Today, it is a fact that HTDNAs are a strategy that may provide denser molecular sampling, and for many research groups around the world, these technologies are a real available methodological alternative. Currently, to access this denser molecular sampling, there are different techniques that include: transcriptomic sequencing, low coverage sequencing (i.e., genome skimming), restriction-site-associated DNA sequencing (e.g., RAD-Seq), and target enrichment (studies cited in McKain et al. [9] and Yu et al. [10]). For the target-enrichment approach, one way is for the scientific team to start from zero in order to design the specific targeted probes (i.e., probe sets, probe kits, enrichment probes, and enrichment kits) for the studied plant taxa [7,11,12]; however, to begin, they will need to generate the primary data, which must be represented by at least a single whole-genome sequence or transcriptome from a taxon. If such primary data were already sequenced with free access, the scientific team would save money and time. If not, researchers would need to generate this primary set of DNA data and, consequently, would require more resources to experimentally produce the primary DNA/RNA data. In this scenario, the scientific team would have to design, essay, and fully complete experimental tests for a set of taxa; however, they would be dealing with complicated bioinformatics analyses. An alternative strategy is to use some of the available predesigned universal targeted probe sets and experimentally test the effectiveness of these on the studied taxa. These probe sets are presumed to be of universal efficiency and are expected to find dozens or hundreds of nuclear orthologous loci across different angiosperm taxa; however, these results are unknown until the experimental and bioinformatics processes are completed. Lastly, we propose here an alternative strategy that consists of testing in silico the available universal angiosperm probe sets, without preliminary experimental costs. This in silico test requires the availability of primary massive DNA sequences. It is expected that in this quantity of DNA, the nuclear genomes of the angiosperms of interest are represented, and thus it is expected that a high number of variable and orthologous loci will be recovered. In this in silico test, the research team must uniquely establish a detailed and appropriate bioinformatic protocol that can be implemented with free-access software, as in this study.

The objective of the present study was to compare the effectiveness in silico of two probe sets that promise universal results for angiosperms. The probe set Angiosperm v.1 was designed from the genomes of 25 species and contains 56,862 probes, expected to recover up to 517 individual exons with a maximum total length of 150 kbp [13]. In contrast, the probe set Angiosperms353 was designed from the transcriptomes of over 600 species and contains 75,151 probes, which are expected to recover 353 coding sequences (CDS) composed of one or more exons with a total length of 260 kbp [14]. These two universal probe sets have been applied in specific studies in various lineages of angiosperms for phy-

logenetic purposes. For example, the Angiosperm v.1 probe set was successfully used in the species of the genera *Aristolochia* [15] and *Acer* [16]. Recently, the probe set Angiosperms353 resolved phylogenetic issues in the genera *Nepenthes* [17], in the species of the genus *Cyperus* [18], and in the species of the ten families that comprise the order Cornales [19]. The results of these studies showed that these two probe sets were appropriate alternatives for obtaining loci that were used to construct resolved phylogenetic trees.

Accordingly, in this study, we tested in silico these two probe sets (Angiosperm v.1 and Angiosperms353) in ten Caryophyllids species: nine species of cacti (Cactaceae) and the species *Amaranthus hypochondriacus* (Amaranthaceae), used in the phylogenetic analysis as an external group. We focused on cacti species because phylogenetic studies carried out with these taxa repeatedly result in unresolved phylogenetic trees [20–22]. This lack of phylogenetic resolution may be due to the fact that most of these studies used poor molecular sampling (1–12 loci), or may be due to other factors, such as the recent postulation of the origin of Cactaceae (30 to 35 Ma) based on the molecular-clock model [23,24]. Moreover, it has been hypothesized that a global aridification that occurred nearly ∼12 million years ago (the late Miocene epoch) promoted the main evolutionary divergence in the Cactaceae family [23]. Consequently, the relatively recent origin of the Cactaceae family and the rapid and recent divergence of cacti species may cause a weak and shallow molecular separation, which has been reflected in those unresolved phylogenetic tree species. In addition, as HTDNAs have shown a significant increase in the number of loci-resolved phylogenetic tree species even in the taxa that have recently diverged [8], we are expecting that these two probe sets have a similar effectiveness in recovering nuclear orthologous loci that serve the cacti species analyzed in the phylogenetic studies.

2. Materials and Methods

2.1. Genomes of the Species Analyzed

At the beginning of our study, we carried out preliminary tests with the genomes <12 Gbp; however, the bioinformatic protocol from the pipeline used here did not work. For this, we e-searched all cacti species represented by >15 Gb of massive DNA sequencing in the NCBI digital database. A total of eight genomes of different cacti species were downloaded from the SRA repository. These DNA data ranged from 17 Gb (*Pereskia humboldtii*) to 57 Gb (*Selenicereus undatus*) (Table S1). In addition, the complete genome of the small globose cactus *Mammillaria huitzilopochtli* was de novo sequenced (raw data are available in S.S.) based on the PE genomic library sequenced in NovaSeq of Illumina. These nine cacti species are grouped into three cacti subfamilies and five tribes. The subfamily Cactoideae was represented by the tribes Cacteae (*M. huitzilopochtli*), Cereeae (*Cereus fernambucensis*), Hylocereeae (*Selenecereus undatus*), and Pachycereeae (*Carnegiea gigantea*, *Lophocereus schottii*, *Pachycereus pringlei*, and *Stenocereus thurberi*). The subfamily Opuntioideae was represented by the tribe Opuntieae (*Opuntia sulphurea*), and the subfamily Pereskioideae by the species *P. humboldtii*. In addition, *Amaranthus hypochondriacus* (Amaranthaceae, Caryophyllales; accession number in genebank, SRR2106212) was used as an external group for phylogenetic analyses of the studied species.

2.2. Bioinformatic Process to Test In Silico the Two Probe Sets

The massive amount of available raw data of each of the ten species (i.e., ten genomes) were firstly filtered by reads using Trim Galore version 0.6.7 [25] for subsequent bioinformatic analyses. For this, the total of the reads of each genome was trimmed, and the adapters were removed; the only reads used in the analysis were those that had a PHRED quality score ≥15 and a length ≥80 bp. The reads per genome that passed this filtering were in silico tested with the original predesigned targeted DNA sequences (TDS) from the two probe sets that were compared. Since the probe set Angiosperm v.1 includes 517 TDS, and Angiosperms353 includes 353 TDS, a total of 870 TDS were tested individually per genome following the pipeline available in HybPiper version 1.3.1 [26]. Accordingly, each of the 870 TDS were individually mapped and assembled with the main script (read_first.py)

of this pipeline. The target files used in HybPiper correspond to the sequences from which the probe sets were designed and were downloaded from https://github.com/mossmatters/Angiosperms353/blob/master/Angiosperms353_targetSequences.fasta (Angiosperms353) (accessed on 15 March 2022) and https://www.biorxiv.org/content/10.1101/086298v2.supplementary-material (Angiosperm v.1) (accessed on 15 March 2022). Since the sequences of Angiosperm v.1 were 517 alignments in Phylip format, they were converted to fasta format, the gaps were removed, and the sequences were merged. Two scripts in HybPiper (scripts get_seq_lengths.py and hybpiper_stats.py) were used to obtain the length of the coding sequences (i.e., exons) and the statistical values that estimate the recovery efficiency of each of the two probe sets per genome; lastly, the R script (gene_recovery_heatmap.R) was used to plot the heatmaps and to visualize the efficiency of the recovering loci per genome. Finally, the script retrieve_sequences.py from the pipeline in HybPiper was used to generate the multifasta files to recover the total number of orthologous nuclear loci per genome. To avoid a possible bias in paralogous identification, we extracted the largest exon from each gene of the Angiospems353 dataset, and we generated the target file (Supplementary File S1) that contains the individual exons of 3829 of the 4781 CDS (target instances) contained in the Angiosperms353 dataset [14]. This target file was used to run an independent HybPiper pipeline, and the paralogous genes were added to the paralogs identified in the analysis of the complete CDS from Angiosperms353. All the sequences identified as paralogous loci in Angiosperm v.1 and Angiosperms353 (CDS + exon) were discarded for the phylogenetic analysis. The relative effectiveness of each of the two probe sets was estimated per species based on the following parameters: (1) the total number of TDS recovered (i.e., the number of loci) by each probe set; (2) the length (bp) of each locus was summed to have the total length (bp) of all the recovered loci; (3) the total number of paralogous loci was discarded for phylogenetic analysis; and (4) the proportion of parsimoniously informative sites (PIS) was obtained from the alignments of the DNA sequences of the orthologous loci.

2.3. Phylogenomics Analysis

The phylogenetic trees were based on the total nuclear orthologous loci recovered by Angiosperms353 (240 loci) and Angiosperm v.1 (71). The alignment of the species of each locus was carried out with MAFFT version v7.310 [27]. We observed a large disparity in the distribution of the orthologous loci among the ten studied species, and in order to diminish the missing data for phylogenetic analysis, we chose those loci that were recorded in the alignments of at least seven species. In addition, for each orthologous locus, its respective proportion of PIS was estimated with the program AMAS [28]. The application ModelFinder [29], implemented in IQ-TREE2 version 2.1.4-β [30], was used to select the partitioning scheme and to estimate the substitution models for each partition and for each locus. The DNA sequences of the orthologous loci obtained with each probe set were concatenated to construct the phylogenetic species-trees with IQ-TREE2 that ran 1000 replicates of an ultrafast bootstrap (UFBoot). In order to obtain the gene concordance factors (gCF) for each node of each of the two phylogenetic trees, we constructed an individual gene tree for each locus with IQ-TREE2.

3. Results

3.1. Effectiveness of the Two Probe Sets

The results showed that, among the ten studied species, the two probe sets tested had clear differences in the global efficiency of the TDS captured (Figure 1, Table 1). For the nine cacti species, Angiosperms353 showed a higher efficiency than Angiosperm v.1, which is supported by the total number of loci recovered (Figure 1), the length of the DNA sequences recovered by the locus, and the number of paralogous loci that were discarded (Table 1). These differences in global efficiency are visualized by the abundant dark lines seen in the heatmap for each species (Figure 1). Each vertical line indicates that an individual TDS was found in the tested genome, and its darkness indicates the proportion of the length of

this sequence that was captured (Figure 1). For example, for *A. hypochondriacus*, the probe set Angiosperms353 showed a higher efficiency in capturing loci (316; Table 1); however, the length of the 216 loci captured with Angiosperm v.1 was larger (68,205 bp; Table 1); therefore, the heatmap is darker for this species (Figure 1).

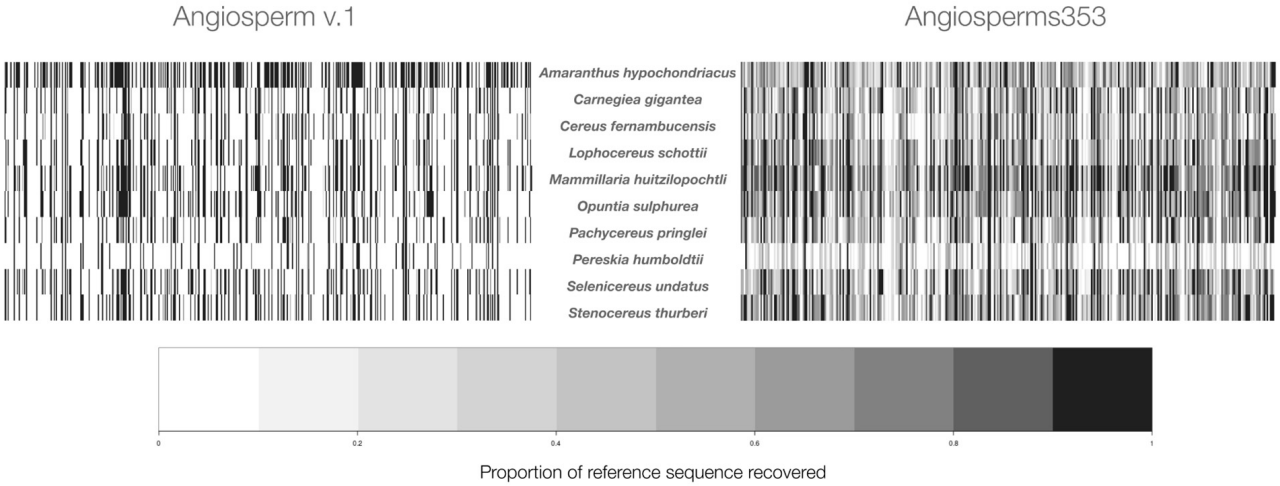

Figure 1. Heatmap per species is presented by raw data. To the left of the species names are the results obtained with Angiosperm v.1, and to the right, the results obtained with Angiosperms353. The gaps (white color) are caused by the absence of vertical lines, which indicates that the targeted DNA sequences did not find those loci in the genome tested.

Table 1. Comparison of the four parameters used to estimate the effectiveness between the two probe sets in silico tested. The results outside parentheses correspond to results obtained with Angiosperm v.1, and those inside parentheses with Angiosperms353. Total number of loci recovered with each probe set; number of loci that found a proportion ≥50% of the length of the targeted DNA sequence; the total length (bp) of sequences identified as exons and the quotient calculated; total length of exons from Angiosperms353/length of exons from Angiosperm v.1; and the absolute number of paralogous loci identified per species.

Species	Number of Loci Recovered	Number of Loci with >50% Sequence Length	Total Length of Exons (bp): Quotient	Number of Paralogous Loci
A. hypochondriacus	264 (316)	256 (167)	68,205 (65,658):0.96	24 (8)
C. gigantea	129 (270)	123 (165)	32,133 (124,773):3.88	12 (8)
C. fernambucensis	132 (269)	125 (145)	30,822 (116,277):3.77	13 (1)
L. schottii	137 (287)	135 (194)	35,127 (140,259):3.99	18 (6)
M. huitzilopochtli	177 (319)	168 (245)	43,875 (174,285):3.97	36 (21)
O. sulphurea	155 (292)	149 (204)	40,611 (150,720):3.71	18 (8)
P. pringlei	123 (253)	121 (158)	31,683 (122,844):3.88	6 (4)
P. humboldtii	68 (173)	67 (69)	18,885 (65,658):3.48	2 (2)
S. undatus	154 (287)	149 (170)	37,866 (133,071):3.51	16 (6)
S. thurberi	145 (290)	143 (204)	37,656 (143,322):3.81	16 (8)

The respective heatmaps of the cacti species show abundant gaps with Angiosperm v.1, because a large number of its 517 targeted DNA sequences did not match (Figure 1) in these genomes. This poorer capture of Angiosperm v.1 was documented in the four parameters used to compare its effectiveness to Angiosperms353. The results showed that, in the ten studied species, the probe set Angiosperms353 captured the highest number of nuclear orthologous loci (Table 1). In the ten studied species, with Angiosperms353, the average number of the orthologous loci was 275.6 ± 41.3 SD, and with Angiosperm v.1, it was 148.4 ± 49.6 SD. Moreover, in the nine cacti species studied, the length of the exons with Angiosperms353 was approximately four times larger than with Angiosperm v.1 (Table 1). In the ten studied species, the number of orthologous loci identified with Angiosperms353

was higher than with Angiosperm v.1 (Table 1). Additionally, in the ten studied species, Angiosperm v.1 recovered a total of 70 paralogous loci, and the percentage of these loci varied from 2.9% (*P. humboldtii*) to 20.3% (*M. huitzilopochtli*). In contrast, paralogous loci in Angiosperms353 varied from 0.4% (*C. fernambucensis*) to 6.7% (*M. huitzilopochtli*).

The orthologous loci identified in 70% of the analyzed species totaled 240 alignments with Angiosperms353, and only 71 alignments were obtained with Angiosperm v.1. The lowest number of alignments was obtained for the species *P. humboldtii*: 141 and 22 alignments obtained with Angiosperms353 and Angiosperm v.1, respectively. In contrast, the highest number of alignments was obtained for *A. hypochondriacus* (62) and *M. huitzilopochtli* (239) with Angiosperm v.1 and Angiosperms353, respectively (Figure 2).

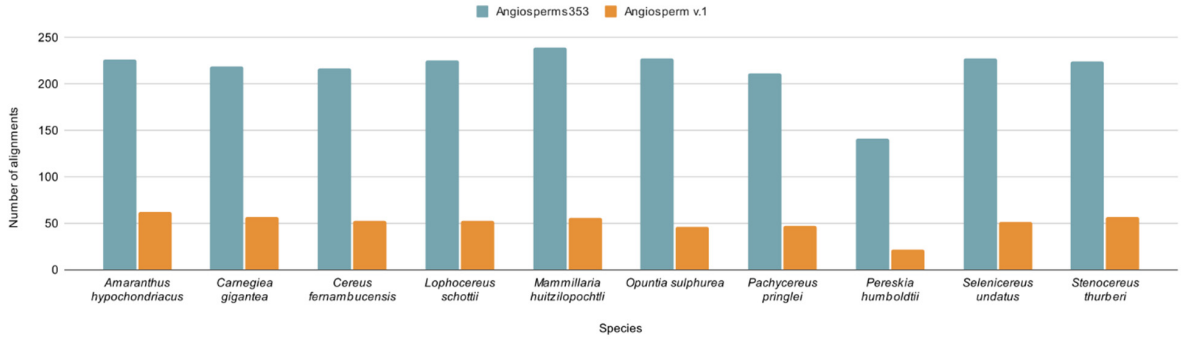

Figure 2. Total number of alignments obtained with Angiosperms353 and Angiosperm v.1 per species.

With respect to the number of parsimoniously informative sites (PIS), Angiosperms353 recovered 4.32% of PIS in the sequences of the total number of orthologous loci aligned, whereas Angiosperm v.1 recovered 3.49% of PIS in such alignments.

3.2. Phylogenomics Analysis

The phylogenetic trees constructed with the DNA datasets provided by each of the two probe sets showed a fully resolved topology (Figure 3). Interestingly, in these two trees, the seven species of the subfamily Cactoideae were clearly grouped in a single monophyletic node. However, these trees showed subtle differences in the topology of the position occupied by *O. sulfurea*, as well as in the statistical values of the UFBoot and in the gCF. In the tree constructed from the data of Angiosperms353, the nodes were 100% supported by UFBoot values, but in the tree of Angiosperm v.1, these values varied from 88 to 100%. However, these two trees showed low values of gene concordance (<53%) for all the nodes.

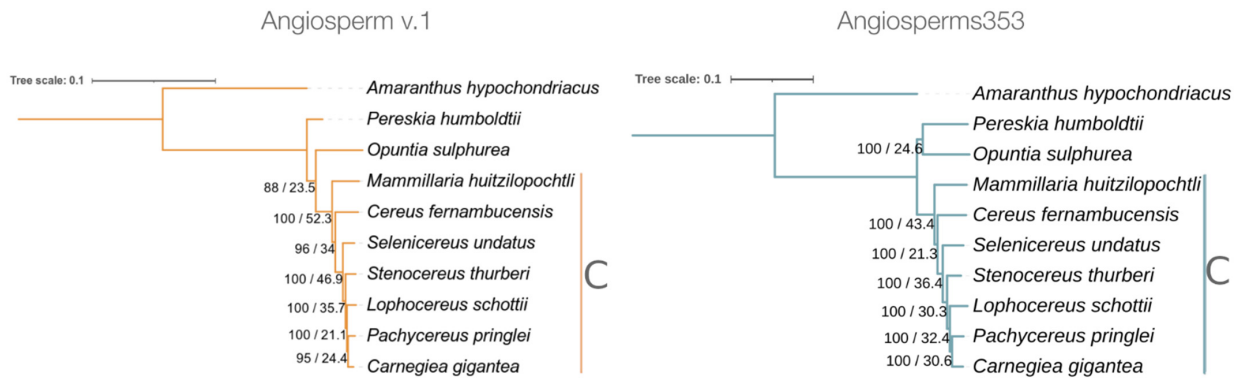

Figure 3. Phylogenetic trees obtained with ML for the nine studied cacti species; the prince's feather (*A. hypochondriacus*) was the outgroup. The tree constructed with data from Angiosperm v.1 was based on 71 loci (20,082 bp), and that constructed with data obtained from Angiosperms353 was based on 262 loci (186,973 bp). The numbers to the side of the branches correspond to UFBoot (above slash) and gCF (below slash) values. The species grouped in the subfamily Cactoideae are indicated by the letter C.

4. Discussion

Our results showed that the two probe sets tested here were useful for Caryophyllids. However, Angiosperms353 had notably better results than Angiosperm v.1 for the nine cacti species, irrespective of the tribe or the subfamily in which these species are classified. It is evident that this dissimilar effectiveness was not related to the larger number of targeted DNA sequences (517) included in Angiosperms v.1 vs. 353 of Angiosperms353. We consider that these superior results obtained with Angiosperms353 for the cacti species could be explained by factors involved in the design process of the targeted DNA sequences for these probe sets: (1) the relatively wider taxonomic sampling employed to design Angiosperms353 (over 600 taxa) vs. Angiosperm v.1. (25 taxa), which could enhance the probability of sampling a larger number of targeted loci dispersed across different lineages, and that (2) Angiosperms353 was designed to recover up to 260 kbp of exons, whereas Angiosperm v.1 was designed to recover only 150 kbp.

Although these two probe sets were designed to recover only single copy genes, they failed because paralog sequences were recovered, but a substantially lower percentage was discarded with Angiosperms353. In addition, the species *P. humboldtii* showed the lowest global effectiveness with the two probe sets. We consider that these results may be caused by the relatively small quantity of the massive DNA dataset (17 Gb) available for this species, and it is probable that the nuclear genome was not completely sampled. Thus, it is important to consider the size of the nuclear genome in order to choose the appropriate platform for massive sequencing of the interested species.

With respect to *A. hypochondriacus*, the better results were obtained with Angiosperm v.1, except for the parameter of the number of paralogous loci. We consider that this was caused by the fact that, during the design of the targeted sequences of Angiosperm v.1, the authors sampled *β vulgaris* (Chenopodiaceae), a species that is phylogenetically close to the Amaranthaceae family, according to previous studies [31,32]. Consequently, we believe that the taxonomic identity of the taxa sampled during the design of the targeted sequences may impact the success of posterior studies of the predesigned probe sets.

Although, in this study, the primary objective was not to resolve the phylogenetic relationships of cacti species, the phylogenetic trees obtained allowed us to make some considerations. The two probe sets provided data used to obtain trees with similar topologies accompanied by low values of gCF. Consequently, these values of gCF indicated a high gene-tree discordance, and we believe that this was not caused by the number of loci used, since the number of loci obtained with the Angiosperms353 dataset (240) was higher than that obtained with Angiosperm v.1 (71). It is probable that these gCF values are related to the poor taxonomic sampling included in our study; however, previous phylogenomic studies with a broader taxonomic sampling also obtained high levels of gene-tree discordance [32,33]. Among the two trees, there was a punctual difference in the position occupied by *O. sulphurea* with respect to the species grouped in the Cactoideae subfamily, similar to the tree obtained with Angiosperm v.1 in previous phylogenetic studies [32–34]. However, based on the global parameters used to measure effectiveness, we consider that Angiosperms353 provided superior results for the cacti phylogenetic studies to Angiosperm v.1.

5. Conclusions

We concluded that the universal probe sets tested here are a confident strategy for studies that require a high number of single copy orthologous loci. We recommend that researchers compare previous in silico analyses of the selected probe sets in order to estimate their effectiveness. We should mention that, in this study, we focused on exons; thus, the molecular variation of introns was not explored. For introns, we would expect larger variation levels. It is likely that the molecular variation contained in introns may separate the more recent evolutionary processes that occurred at intraspecific levels. Lastly, our results confirmed that non-specific probe sets should not necessarily diminish the molecular sampling across different angiosperm groups. In fact, in a previous study

carried out with the taxa of the genus *Cyperus* (Cyperaceae), the Angiosperms353 probe set provided a similar proportion of PIS compared to that obtained with a family-specific probe set, producing a similar resolution power at infrageneric taxonomic levels [18].

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/genes13040570/s1, Table S1: List of the studied cacti species and the subfamily and tribes in which they are classified. Supplementary File S1: Target file with 3829 individual largest exons that were recovered from the total of the 4781 CDS (target instances) contained in Angiosperms353 dataset.

Author Contributions: Conceptualization, D.A.C.; methodology, D.A.C.; formal analysis, D.A.C.; writing—original draft preparation, D.A.C. and S.S.; writing—review and editing, D.A.C. and S.S.; supervision, S.S.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by UNAM-PAPIIT IN228619.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: SEMARNAT (SGPA/DGVS/06880/16) supplied collecting permission to sampling *Mammillaria huitzilopochtli* specimen. D.A.C. (310003715) is a student at Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México; she is granted by Consejo Nacional de Ciencia y Tecnología CONACyT (855061). We thank comments of two anonymous reviewers that improved the quality of this article. Macrogen Inc., Seoul South Korea provided the sequencing service of the whole genome of *M. huitzilopochtli*.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Ruhfel, B.; A Gitzendanner, M.; Soltis, P.S.; E Soltis, D.; Burleigh, J.G. From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* **2014**, *14*, 23. [CrossRef] [PubMed]
2. Gitzendanner, M.A.; Soltis, P.S.; Wong, G.K.-S.; Ruhfel, B.R.; Soltis, D.E. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *Am. J. Bot.* **2018**, *105*, 291–301. [CrossRef] [PubMed]
3. Wickett, N.J.; Mirarab, S.; Nguyen, N.; Warnow, T.; Carpenter, E.; Matasci, N.; Ayyampalayam, S.; Barker, M.S.; Burleigh, J.G.; Gitzendanner, M.A.; et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E4859–E4868. [CrossRef]
4. Li, H.T.; Yi, T.S.; Gao, L.M.; Ma, P.F.; Zhang, T.; Yang, J.B.; Gitzendanner, M.A.; Fritsch, P.W.; Cai, J.; Luo, Y.; et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat. Plants* **2019**, *5*, 461–470. [CrossRef]
5. Coiro, M.; Doyle, J.A.; Hilton, J. How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *New Phytol.* **2019**, *223*, 83–99. [CrossRef] [PubMed]
6. Parks, M.; Cronn, R.; Liston, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* **2009**, *7*, 84. [CrossRef]
7. Nicholls, J.A.; Pennington, R.T.; Koenen, E.J.M.; Hughes, C.E.; Hearn, J.; Bunnefeld, L.; Dexter, K.G.; Stone, G.N.; Kidner, C.A. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Front. Plant Sci.* **2015**, *6*, 710. [CrossRef]
8. Bagley, J.C.; Uribe-Convers, S.; Carlsen, M.M.; Muchhala, N. Utility of targeted sequence capture for phylogenomics in rapid, recent angiosperm radiations: Neotropical *Burmeistera* bellflowers as a case study. *Mol. Phylogenet. Evol.* **2020**, *152*, 106769. [CrossRef]
9. McKain, M.R.; Johnson, M.G.; Uribe-Convers, S.; Eaton, D.; Yang, Y. Practical considerations for plant phylogenomics. *Appl. Plant Sci.* **2018**, *6*, e1038. [CrossRef]
10. Yu, X.; Yang, D.; Guo, G.; Gao, L. Plant phylogenomics based on genome-partitioning strategies: Progress and prospects. *Plant Divers.* **2018**, *40*, 158–164. [CrossRef]
11. Mandel, J.R.; Dikow, R.B.; Funk, V.A.; Masalia, R.R.; Staton, S.E.; Kozik, A.; Michelmore, R.W.; Rieseberg, L.H.; Burke, J.M. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Appl. Plant Sci.* **2014**, *2*, 1300085. [CrossRef] [PubMed]

12. Stephens, J.D.; Rogers, W.L.; Heyduk, K.; Cruse-Sanders, J.M.; Determann, R.O.; Glenn, T.C.; Malmberg, R.L. Resolving phylogenetic relationships of the recently radiated carnivorous plant genus *Sarracenia* using target enrichment. *Mol. Phylogenet. Evol.* **2015**, *85*, 76–87. [CrossRef] [PubMed]
13. Buddenhagen, C.; Lemmon, A.R.; Lemmon, E.M.; Bruhl, J.; Cappa, J.; Clement, W.L.; Donoghue, M.; Edwards, E.J.; Hipp, A.L.; Kortyna, M.; et al. Anchored phylogenomics of angiosperms I: Assessing the robustness of phylogenetic estimates. *BioRxiv* **2016**, *2016*, 086298. [CrossRef]
14. Johnson, M.G.; Pokorny, L.; Dodsworth, S.; Botigué, L.R.; Cowan, R.S.; Devault, A.; Eiserhardt, W.L.; Epitawalage, N.; Forest, F.; Kim, J.T.; et al. A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-Medoids Clustering. *Syst. Biol.* **2019**, *68*, 594–606. [CrossRef] [PubMed]
15. Wanke, S.; Mendoza, C.G.; Müller, S.; Paizanni Guillén, A.; Neinhuis, C.; Lemmon, A.R.; Lemmon, E.M.; Samain, M.-S. Recalcitrant deep and shallow nodes in *Aristolochia* (Aristolochiaceae) illuminated using anchored hybrid enrichment. *Mol. Phylogenet. Evol.* **2017**, *117*, 111–123. [CrossRef]
16. Li, J.; Stukel, M.; Bussies, P.; Skinner, K.; Lemmon, A.R.; Lemmon, E.M.; Brown, K.; Bekmetjev, A.; Swenson, N.G. Maple phylogeny and biogeography inferred from phylogenomic data. *J. Syst. Evol.* **2019**, *57*, 594–606. [CrossRef]
17. Murphy, B.; Forest, F.; Barraclough, T.; Rosindell, J.; Bellot, S.; Cowan, R.; Golos, M.; Jebb, M.; Cheek, M. A phylogenomic analysis of *Nepenthes* (Nepenthaceae). *Mol. Phylogenet. Evol.* **2020**, *144*, 106668. [CrossRef]
18. Larridon, I.; Villaverde, T.; Zuntini, A.R.; Pokorny, L.; Brewer, G.E.; Epitawalage, N.; Fairlie, I.; Hahn, M.; Kim, J.; Maguilla, E. Tackling rapid radiations with targeted sequencing. *Front. Plant Sci.* **2020**, *10*, 1655. [CrossRef]
19. Thomas, S.K.; Liu, X.; Du, Z.-Y.; Dong, Y.; Cummings, A.; Pokorny, L.; Xiang, Q.-Y.; Leebens-Mack, J.H. Comprehending Cornales: Phylogenetic reconstruction of the order using the Angiosperms353 probe set. *Am. J. Bot.* **2021**, *108*, 1112–1121. [CrossRef]
20. Butterworth, C.A.; Wallace, R.S. Phylogenetic studies of *Mammillaria* (Cactaceae)—Insights from chloroplast sequence variation and hypothesis testing using the parametric bootstrap. *Am. J. Bot.* **2004**, *91*, 1086–1098. [CrossRef]
21. Griffith, M.P.; Porter, J.M. Phylogeny of Opuntioideae (Cactaceae). *Int. J. Plant Sci.* **2009**, *170*, 107–116. [CrossRef]
22. Bárcenas, R.T.; Yesson, C.; Hawkins, J.A. Molecular systematics of the Cactaceae. *Cladistics* **2011**, *27*, 470–489. [CrossRef] [PubMed]
23. Arakaki, M.; Christin, P.-A.; Nyffeler, R.; Lendel, A.; Eggli, U.; Ogburn, R.M.; Spriggs, E.; Moore, M.J.; Edwards, E.J. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8379–8384. [CrossRef] [PubMed]
24. Magallon, S.; Gomez-Acevedo, S.; Sanchez-Reyes, L.L.; Hernandez-Hernandez, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **2015**, *207*, 437–453. [CrossRef] [PubMed]
25. Babraham Bioinformatics. Trim Galore. Available online: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed on 21 December 2021).
26. Johnson, M.G.; Gardner, E.M.; Liu, Y.; Medina, R.; Goffinet, B.; Shaw, A.J.; Zerega, N.J.C.; Wickett, N.J. HybPiper: Extracting Coding Sequence and Introns for Phylogenetics from High-Throughput Sequencing Reads Using Target Enrichment. *Appl. Plant Sci.* **2016**, *4*, 1600016. [CrossRef] [PubMed]
27. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef]
28. Borowiec, M.L. AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **2016**, *4*, e1660. [CrossRef]
29. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.; Von Haeseler, A.; Jermiin, L.S. Model Finder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef]
30. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A.; Lanfear, R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **2020**, *7*, 1530–1534. [CrossRef]
31. Cuénoud, P.; Savolainen, V.; Chatrou, L.W.; Powell, M.; Grayer, R.J.; Chase, M.W. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *Am. J. Bot.* **2002**, *89*, 132–144. [CrossRef]
32. Walker, J.F.; Yang, Y.; Feng, T.; Timoneda, A.; Mikenas, J.; Hutchison, V.; Edwards, C.; Wang, N.; Ahluwalia, S.; Olivieri, J.; et al. From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *Am. J. Bot.* **2018**, *105*, 446–462. [CrossRef] [PubMed]
33. Wang, N.; Yang, Y.; Moore, M.J.; Brockington, S.F.; Walker, J.F.; Brown, J.W.; Liang, B.; Feng, T.; Edwards, C.; Mikenas, J.; et al. Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. *Mol. Biol. Evol.* **2019**, *36*, 112–126. [CrossRef]
34. Edwards, E.J.; Nyffeler, R.; Donoghue, M.J. Basal Cactus Phylogeny: Implications of *Pereskia* (Cactaceae) Paraphyly for the Transition to the Cactus Life Form. *Am. J. Bot.* **2005**, *92*, 1177–1188. [CrossRef] [PubMed]

CAPÍTULO 3. Incomplete Lineage Sorting And Deep Reticulate Evolution Explain Phylogenetic Relationships In Cacti

(Manuscrito en preparación)

Incomplete lineage sorting and deep reticulate evolution explain phylogenetic relationships in cacti

Delil A. Chincoya[1,2] and Sofía Solórzano[1]

[1] Laboratorio de Ecología Molecular y Evolución, UBIPRO, FES Iztacala, Universidad Nacional Autónoma de México, Avenida de los Barrios 1, Los Reyes Iztacala, Tlalnepantla de Baz 54090, Estado de México, Mexico

[2] Posgrado en Ciencias Biológicas, Unidad de Posgrado, Edificio D, 1° Piso, Circuito de Posgrados, Ciudad Universitaria, Coyoacán 04510, Ciudad de México, Mexico

1. Introduction

The absence of robust hypotheses for branching patterns of biological diversity has been historically a bottleneck for the advance of evolutionary studies. Resolved phylogenies provide a necessary historical context for studying the evolution of traits (e.g. Carrive et al., 2020; Zhang, 2020, Gamisch, 2021), spatiotemporal patterns (e.g. Deng et al., 2018; Song et al. 2020) and community assemblages (e.g. Liu, 2017, Gong et al., 2019). Additionally, resolved phylogenies play a crucial role in conservation efforts, providing critical information for identifying conservation priorities between populations (Moritz, 1994), species (Byrne, 2001) and geographic areas (Tucker, 2017). Phylogenetic studies have also been shown to refine taxonomic works (Peruzzi, 2023), with many examples of taxonomic decisions having a significant impact on conservation efforts (Morrison et al., 2009). Moreover, resolved phylogenies enable predictions about the potential vulnerability of particular taxa to environmental change (Leao, 2014, Molina-Venegas, 2020). Thus, by providing a historical context, phylogenies can help us to understand the complex processes that shape the biological diversity and develop effective strategies to preserve it.

The high-throughput sequencing technologies have revolutionized the field of phylogenetics. The vast amount of molecular data generated simultaneously from multiple loci lead to a significant improvement in the accuracy and resolution of phylogenetic hypotheses (Lemmon et al., 2013). Today, there are numerous well resolved phylogenies for various lineages of plants, many of them estimated from the chloroplast genome (e.g. Serna-Sánchez et al., 2021, Yao, 2019). The extended

use of the chloroplast genome in evolutionary studies is due to its conserved nature across plant lineages, maternal inheritance, and the ease of sequencing and analysis (Gao, 2010). Nevertheless, it is common to find discordances between phylogenies estimated from chloroplast and nuclear genomes. This discordance may be due to differences in inheritance patterns, evolutionary rates, and selection pressures. In angiosperms, the chloroplast genome typically is maternally inherited and present substitution rates ~5x lower than nuclear genome (Drouin, 2008). Analyzing both, organellar and nuclear genome we can reach a more complete understanding of the evolutionary relationships of organisms and the various factors that have shaped their history over time. The assessment of phylogenetic discordances between genomic compartments (i. e. chloroplast and nuclei) and even between loci of nuclear genome can yield information about biological processes that occurred in the past (e.g. Rose et al., 2021, Wang et al., 2021, Zhang et al., 2022). This includes punctual events of reticulate evolution where interspecific gene flow occurred (i.e Hybridization and introgression) or those related to stochastic transmission and sort of genetic information along generations (i.e. incomplete lineage sorting).

Phylogenetic discordance is not uniformly distributed across angiosperm lineages. Within the order Caryophyllales, the family Cactaceae stands out as having particularly high levels of discordance compared to other families. This discordance is highlighted by a recent phylotranscriptomic analysis that includes 36 of the 1440 species in Cactaceae, which found that some relationships had levels of gene-tree discordance exceeding 75% (Walker et al., 2009). Despite the prevalence of discordance in Cactaceae, only a few studies have specifically assessed the phylogenetic discordance between genomic compartments (e.g, Köhler, 2021) or between loci of nuclear genome (e.g. Copetti et al., 2017) within particular groups of Cactaceae. Hence, due to the limited number of these studies, there is still much to be learned about the extent and nature of phylogenetic discordance in Cactaceae.

The first molecular study of the Cactaceae that sampled *Mammillaria* Haw., which with 164 species (Hunt, 2006 cf. Korotkova et al.,2021) is the larger genus of the family, was conducted two decades ago (Butterworth et al., 2002). This study analyzed 15 *Mammillaria* species based in one chloroplast locus (rpl16 intron) and identified that this genus grouped more than one evolutionary entity. The study also introduced the term "Mammilloid clade" to refer to the clade that includes *Mammillaria* and six other genera (*Coryphantha* (Engelm.) Lem.*, Escobaria* Britton & Rose*, Mammilloydia* Buxb.,

Neolloydia Britton & Rose*, Ortegocactus* Alexander, and *Pelecyphora* Ehrenb.). The seven genera of Mammilloid clade share morphological similarities (i.e. short globular or cylindrical shape, tuberculated steam and dimorphic areoles) and diverged recently, in the last ~8 My (Arakaki et al., 2011, Hernandez-Hernandez et al., 2014, Chincoya et al., 2023). Subsequent studies have replicated the finding of non-monophyly of *Mammillaria* (Butterworth y Wallace, 2004, Crozier, 2005, Barcenas, 2011) with a larger taxonomic representation for this genus (24 to 141 species) but still a poor molecular sampling (one to ten chloroplast loci). Unfortunately, all these studies yielded unresolved phylogenies with low support for many relationships. Recently, two phylogenies were published from chloroplast genomic data for 52 (Breslin et al.,2021) and 70 (Chincoya et al., 2023) taxa of Mammillaria genus. These studies include representatives for five (*Coryphantha, Escobaria, Mammillaria*, *Neolloydia,* and *Ortegocactus*) (Breslin et al., 2021) and all the seven (Chincoya et al., 2023) genera that comprise Mammilloid clade. These studies not only allowed to obtain fully resolved phylogenies but also allowed to identify those taxa assumed as *Mammillaria* that are an independent evolutionary lineage and cause the non-monophyly of the genus. Furthermore, both studies recovered three main clades, with high support, within the Mammilloid clade. The first clade was composed by species of *Mammillaria* and *Mammilloydia*, the second by *Coryphantha*, *Escobaria* and *Pelecyphora* and the third composed by members of *Mammillaria*, *Neolloydia* and *Ortegocactus* (Chincoya et al., 2023). Hence, after two decades of phylogenetic uncertainty, availability of resolved phylogenies for Mammilloid clade open possibilities for testing new evolutionary questions.

In this study we analyze the levels of phylogenetic discordance in taxa of Mammilloid clade with emphasis in the *Mammillaria* genus in order to better understand the patterns of diversification of this group of cacti. Due to the rapid diversification of taxa of the Mammilloid clade, we hypothesize high levels of phylogenetic discordance between distinct genomic compartments and between gene-trees of the nuclear genome. By examining the extent and nature of phylogenetic discordance, we expect to shed light on the historical processes that have shaped the evolutionary relationships of these cacti. Hence, the objectives of this study were to compare the phylogenetic relationships estimated from chloroplast and nuclear genomes, to assess the nuclear gene tree conflict and to evaluate the impact of reticulate events in phylogenetic discordance.

2. Material and methods

2.1. Taxon sampling

A total of 47 taxa of Mammilloid clade were sampled (Table S1). Since our study emphasis in *Mammillaria*, ~80% of sampled taxa (37 taxa) belong to this genus. The remaining sampled taxa belong to genus *Coryphantha* (3 taxa), *Escobaria* (2 taxa), *Mammilloydia* (1 taxa), *Neolloydia* (2 taxa), *Ortegocactus* (1 taxa) and *Pelecyphora* (1 taxa). We choose a member of the tribe Cacteae, *Stenocactus multicostatus* (Hildm.) A. Berger ex A.W. Hill*,* as an outgroup. The specimens for 46 of these taxa come from the collection of the Botanical Garden of the Universidad Nacional Autónoma de México. The tissues of *M. napina* J.A.Purpus and *M. huitzilopochtli* D.R.Hunt were obtained from previous research projects (S.S.).

2.2. DNA Extraction, target enrichment and High-Throughput Sequencing

In order to isolate > 1 ug of gDNA with A260/280 ratio ≥ 1.7 per sample, 30-100 mg of tissue was processed with the DNeasy plant mini kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. gDNA was used to prepare pair-end libraries that were enriched with putatively single-copy genes (Johnson et al., 2019) using myBaits Expert Angiosperms 353 v1 (Daicel Arbor Biosciences, Michigan, USA). The enriched libraries were sequenced in NovaSeq S4 platform (Ilumina, CA, USA). The preparation, enrichment and sequencing of the libraries was carried out by Daicel Arbor Biosciences, Michigan, USA.

2.3. Sequencing data processing

Trimming and removal of adapters from the raw reads was performed with Trim Galore version 0.6.7. and reads with PHRED ≥20 and length ≥80 bp were conserved. To recover the nuclear loci, the file with the 4,781 target regions on which the Angiosperms353 probes (Johnson et al., 2019) were designed was used as a starting point. Additionally, this file was enriched with regions from four Caryophyllales species (*Lophophora williamsii* (Lem. ex Salm-Dyck) J.M. Coult*, Opuntia polyacantha* Haw.*, Pereskia aculeata* Mill. and *Portulaca amilis* Speg.) available in the mega353 file (McLay et al., 2021), resulting in a final target file with 5,209 target regions for the 353 target genes. The nuclear loci recovery was performed with the HybPiper pipeline version 1.3.1 (Johnson et al., 2016). With the

main HybPiper script (read_first.py) the mapping, assembly and recovery of the nuclear loci was performed. The intronate.py script was used to recover the intron sequences flanking the exons. The efficiency of recovery of nuclear loci was evaluated with the get_seq_lengths.py and hybpiper_stats.py scripts. Putative paralogs were identified with paralog_retriever.py script; for these loci were built their respective gene trees with FastTree version 2.1.10 (Price et al., 2010) to determine if they were alleles or paralogs. Those genes identified as paralogous in > 20% of the species were excluded from the dataset. For downstream analysis were used the sequences composed by introns and exons (i.e. supercontig sequences), which were aligned with MAFFT v7.310 (Katoh, 2009). In order to obtain high-quality alignments, free of gappy rich or poorly aligned regions, the alignments were trimmed with the automated option of TrimAl v1.4.rev15 (Capella-Gutiérrez et al., 2009). Respect to chloroplast sequences, coding and non-coding loci were identified and recovered from complete chloroplast genomes as was described in a previous study (Chincoya et al., 2023). High-quality alignments of chloroplast loci were obtained in the same way that those of nucleus were obtained.

2.4. Estimation of phylogenetics relationships

Species trees were estimated using supermatrix (for chloroplast and nuclear sequences) and coalescent based methods (for nuclear sequences only). In order to limit the amount of missing data, only those alignments with sequences for > 85 % of taxa were analyzed. Under previous criteria, a total of 322 nuclear loci (655,929 bp) and 127 chloroplast loci (76,668 bp) were used in phylogenetic analysis. The chloroplast and nuclear supermatrix species trees and gene trees of nuclear genome were estimated with RAxML (Stamatakis 2014) with *Stenocactus multicostatus* as outgroup. Phylogenetic relationships were estimated under the GTR + I + Γ substitution model with 1000 replicates of rapid bootstrap. For the nuclear coalescent species tree, the gene trees nodes with support values < 50% were collapsed and the resulting gene trees were summarized with ASTRAL-III (Zhang et al., 2018).

2.5. Assessment of gene tree conflict and reticulation events in nuclear dataset

The concordance analysis between the 322 nuclear gene trees and the coalescent based species tree was performed with Phyparts v0.0.1 (Smith et al., 2015). The Phyparts output was visualized with

phypartspiecharts.py (Johnson, 2017). To assess the effect of reticulation events in gene tree conflict, were inferred phylogenetic networks with PhyloNet v.3.8.2 (Wen et al., 2018) using the maximum pseudolikelihood method. Two particular instances were addressed: 1) the backbone relationships between the four main clades recovered in species trees of the nuclear dataset (Cochemiea, Coryphantha, Mammillaria and Pelecyphora); for this, were subsampled a total of 11 species of these clades and *S. multicostatus*. 2) The shallow relationships between five *Mammillaria* taxa with high levels of discordance: *M. heyderi* subsp*. heyderi* Muehlenpf*., M. magnimamma* Haw*., M. perbella* Hildm. ex K. Schum*, M. scrippsiana* (Britton & Rose) Orcutt and *M. uncinata* Zucc. ex Pfeiff., using *Coryphantha clavata* as outgroup. Phylogenetic networks were estimated for zero to four reticulations running 20 independent searches. For each reticulation number, was selected the network with higher logL. Comparisons between networks with different number of reticulations were performed using the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC). Best phylogenetic networks for each number of reticulations were visualized in IcyTree (Vaughan, 2017).

3. Results

3.1. Phylogenetic analysis of nuclear genome, topological comparison with chloroplast phylogeny and assessment of gene tree conflict

Most of the relationships estimated with the nuclear dataset were recovered with a high support, with only two of them recovered with values < 95% of rapid bootstrap (BS) (Fig. 1, left). Phylogenetic relationships estimated with the chloroplast dataset also showed high support values; only three of them being < 95 % of rapid BS. Nuclear phylogenetic analysis recovered four main clades between the 47 species of Mammilloid clade (*Coryphantha, Escobaria, Mammillaria, Mammilloydia, Neolloydia, Pelecyphora* and *Ortegocactus*). These four clades (Cochemiea, Coryphantha, Escobaria and Mammillaria) were recovered with 100 % of rapid BS, nevertheless the position of Coryphantha clade respect to Cochemiea + Mammillaria clade was recovered with 76 % of rapid BS. In contrast, chloroplast phylogenetic analysis recovered only three main clades (Cochemiea, Coryphantha and Mammillaria) for species grouped in Mammilloid clade (Fig. 1, right). These three clades, and the relationships between them, were recovered with high support values (rapid BS > 95 %).

Phylogenies estimated from nuclear and plastid genomes were also discordant in shallow relationships inside main clades, particularly between species of Mammillaria clade. For example, one of these discordances were found in the identification of the sister lineage to the rest of Mammillaria clade. In nuclear phylogeny *Mammilloydia candida* (Scheidw.) Buxb. was recovered with low support (44 % of rapid BS) as the sister, whereas in plastid phylogeny a small clade composed by *Mammillaria beneckei* Ehrenb., *M. napina* and *M. sphacelata* Mart. was sister to the rest of Mammillaria clade (98 % of rapid BS).

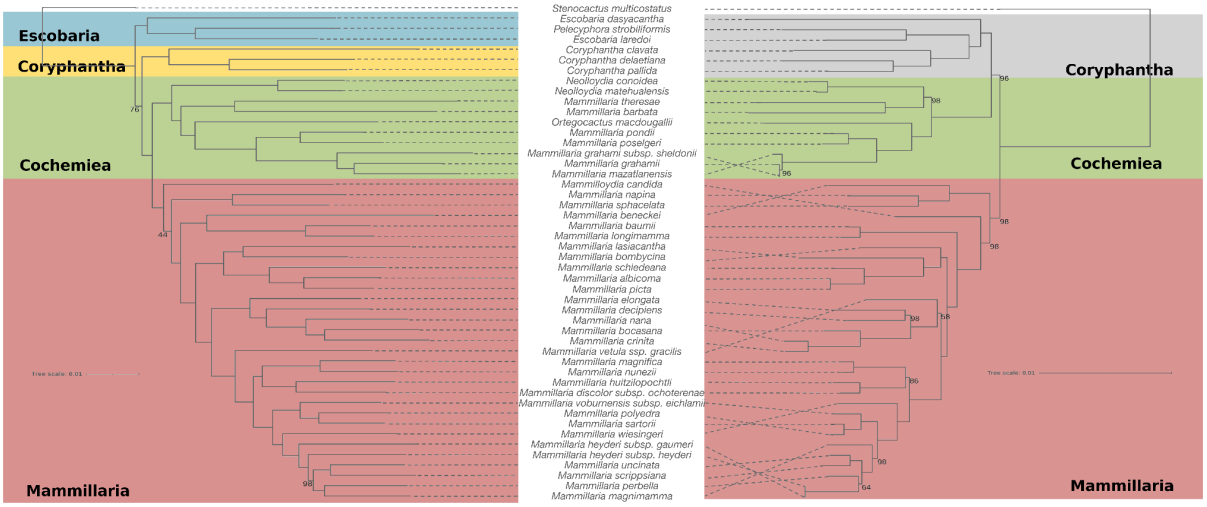

Fig. 1. Phylogenetic relationships estimated from nuclear (left) and plastid (right) loci. Numbers below the nodes indicate values of rapid BS < 100%. The main clades identified for each genomic compartment are indicated with different colors.

Species trees estimated from the nuclear genome with supermatrix (Fig. 1, left) and coalescent methods (Fig. 2) provided the same topology. Nuclear topology was highly supported according to rapid BS, nevertheless was poorly supported by the 322 gene trees analyzed. Only 30% of the nodes were supported by > 50% of gene trees (Fig. 2). This high proportion of gene tree discordance was obtained in both, backbone (between the four main clades) and shallow (into the clades) relationships. This widespread discordance is due to many topologies with low frequency estimated from gene-trees rather than a strong support for an alternative topology with high frequency.

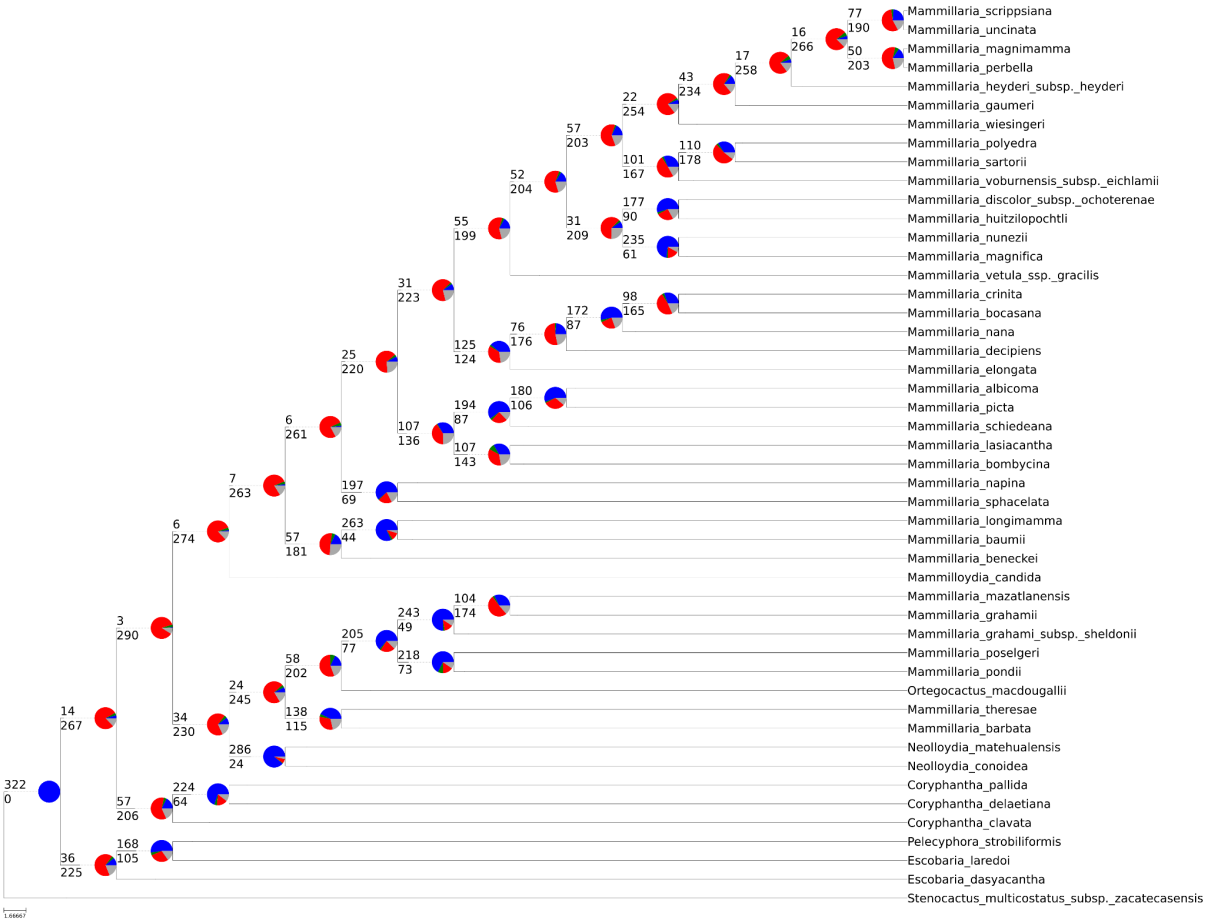

Fig. 2. Phylogenetic relationships estimated with coalescent based methods and gene tree conflict. Pie charts in each node show the proportion of concordant and conflicting gene trees with this bipartition. Blue slice indicates the proportion of gene trees that support the showed topology, green slice indicates the proportion of conflicting gene trees that support another majority topology, red slice indicates the proportion of conflicting gene trees that support multiple minoritary topologies and gray slice indicates the proportion of gene trees non informative for this bipartition.

3.2. Reticulation events as a source of phylogenetic discordance

The phylogenetic networks for backbone relationships of Mammilloid clade inferred a maximum of three reticulation events between the four main clades recovered in nuclear dataset (Escobaria, Coryphantha, Cochemiea and Mammillaria) (Fig. 3). Neither of the reticulate events inferred was present in all the phylogenetic networks, however the hybrid origin of Coryphantha clade was estimated by two networks (1 and 2 reticulations). According to AIC (72360.72) and BIC (72365.73) (Table S1), the best phylogenetic network for backbone relationships of Mammilloid clade is that with

two reticulations. The first reticulate event of the best phylogenetic network inferred that Coryphantha clade was related to Mammillaria and Escobaria clades by an inheritance probability (I. P.) of 0.67 and 0.33, respectively. The second reticulate event estimates that the ancestral lineage of Coryphantha-Escobaria was related to Cochemieae (I. P. 0.63) and some taxa, probably extinct or not sampled, phylogenetically close to the whole Mammilloid clade (I. P. 0.37).

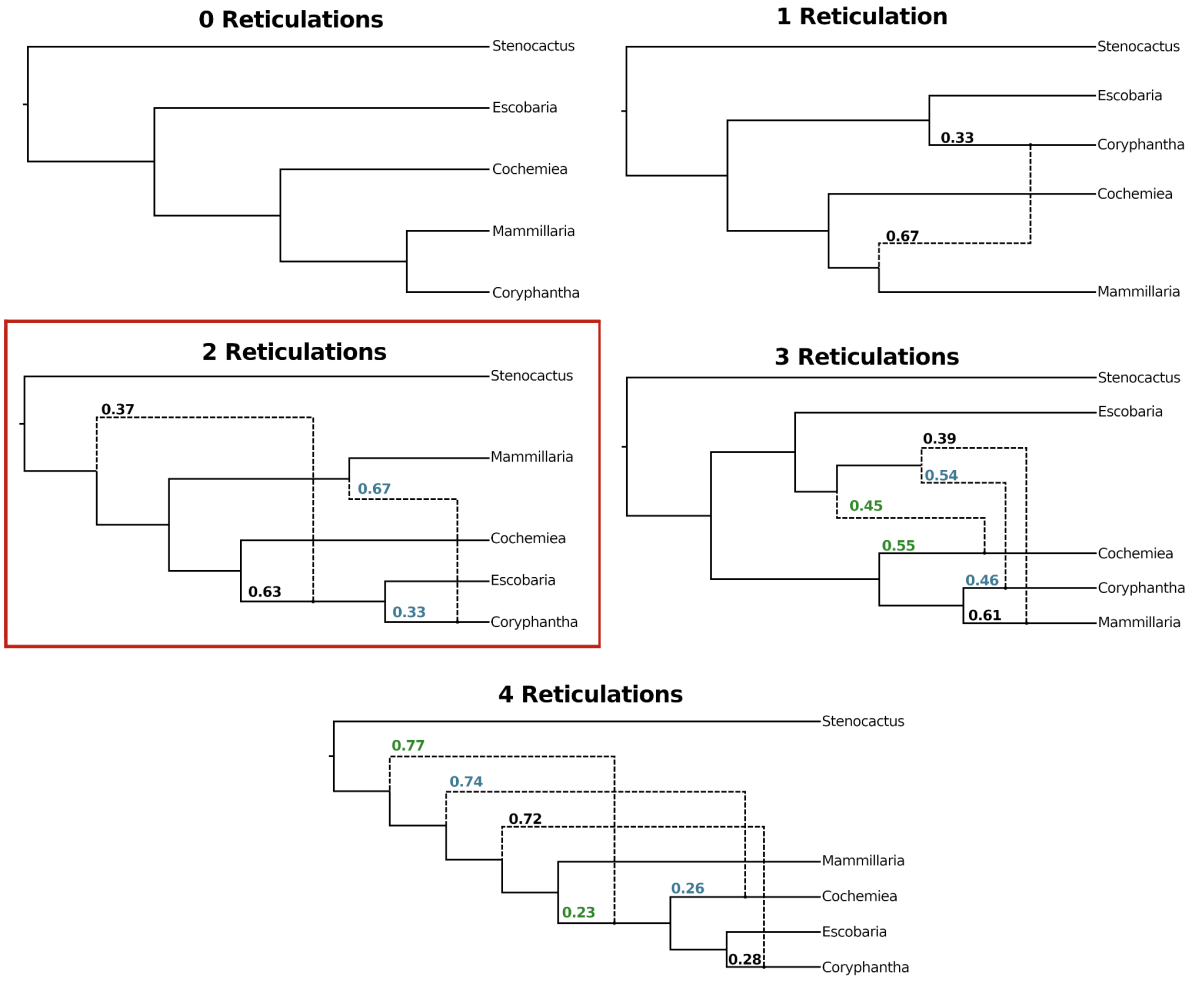

Fig. 3. Best phylogenetic networks inferred up to four reticulations for backbone relationships between the four main clades recovered in the nuclear species tree. Optimal network according to information criteria is marked inside the red square. Numbers above the lines indicate the inheritance probabilities for each hybrid node.

Respect to shallow relationships between *M. heyderi* subsp. *heyderi*, *M. magnimamma, M. perbella, M. scripssiana* and *M. uncinata*, according to AIC (12729.29) and BIC (12731.27), a bifurcating tree is better than any of the networks inferred by Phylonet with up to four reticulations (Fig. 4, Table S1).

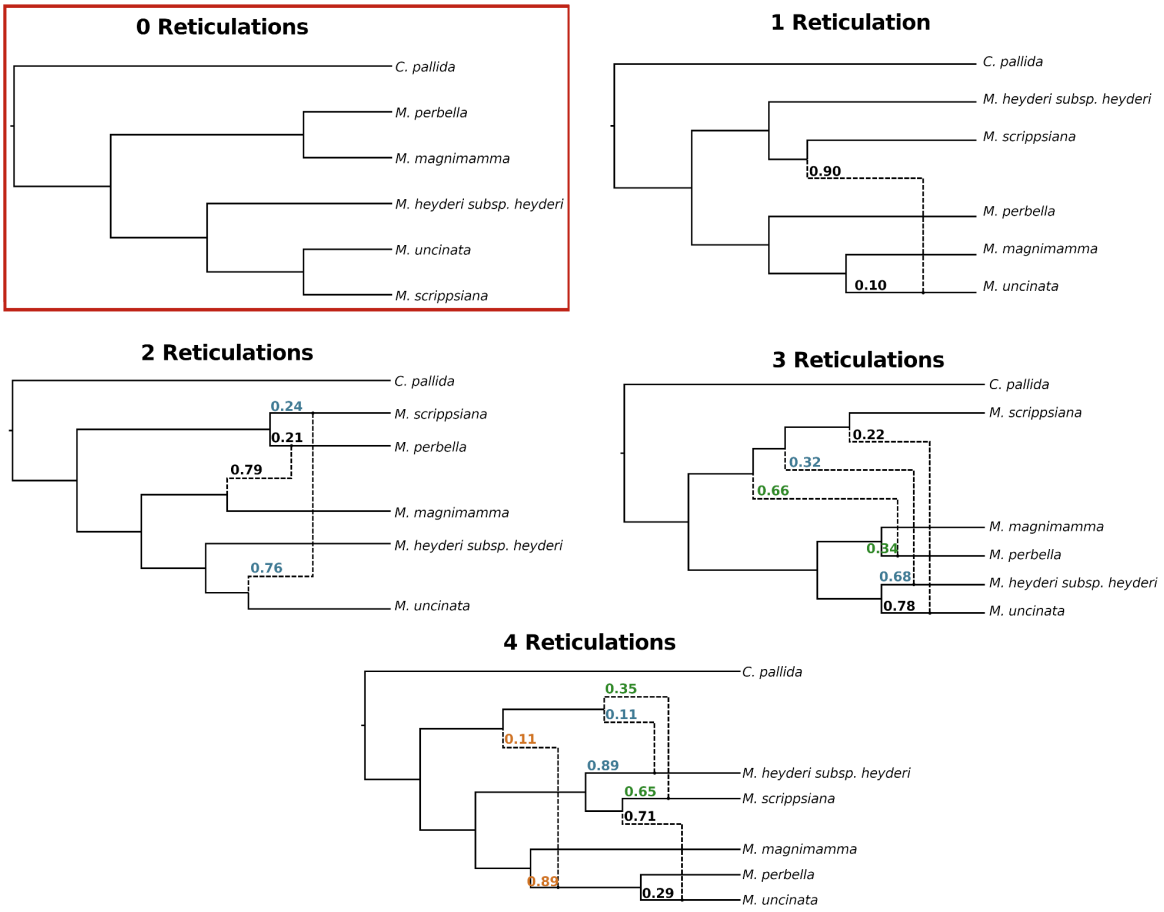

Fig. 4. Best phylogenetic networks inferred up to four reticulations for shallow relationships between 5 taxa with high levels of phylogenetic discordance. Optimal network according to information criteria is marked inside the red square. Numbers above the lines indicate the inheritance probabilities for each hybrid node.

4. Discussion

Our results indicate that the process that generate phylogenetic discordance shaped the genomes of Mammilloid clade both at backbone (i.e number of main clades recovered and relationships between them) and shallow relationships. This is concordant with previous studies in the Cactaceae family that analyzed cytonuclear discordance (Köhler, 2021) and discordance between nuclear loci (Copetti et al., 2017).

4.1. Concordance and discordance in the Mammilloid clade and its taxonomic consequences

It is remarkable that species trees estimated from nuclear dataset based on supermatrix and coalescent methods yielded the same topology. Nevertheless, highlight the degree of informativeness of the distinct support values for each tree. The significant gene tree conflict, particularly in certain nodes where strong bootstrap values are also observed, lead to a cautious approach when interpreting some inferred relationships. This becomes especially important when results from molecular studies are applied in the circumscription of previous taxonomic classification systems (e.g. Korotkova et al, 2021 and studies here cited). It is undeniable the urgency of refinement of taxonomic work, particularly of those groups that face conservation threats, as is the case of nearly 30 % of the members of the Cactaceae family (IUCN). Nevertheless, hurried decisions can lead to instability and even some taxonomic changes can be detrimental to conservation of taxa (Morrison et al., 2009).

Respect to cytonuclear discordance highlights the differences in backbone relationships. Nuclear topology was similar to the previously documented by Sánchez et al, 2022 from five plastid loci and eight morphological characters. The phylogeny of Sánchez et al, 2022, although partially unresolved, recovered to *Coryphantha* (excluding *C. macromeris* (Engelm.) Britton & Rose) as monophyletic and recovered in a sister clade those members of *Pelecyphora* and *Escobaria*. As a consequence of their taxonomic findings, the authors proposed reclassifying the species of the second clade as a single genus (*Pelecyphora*). In contrast, plastid topology recovered in our study did not support the monophyly of *Coryphantha*. This result has already been documented in studies also performed with genome scale plastid data (Breslin et al., 2021, Chincoya et al., 2023), but with a denser taxonomic representation. Due to limited taxonomic representation of taxa of these three genera, their taxonomic circumscription exceeds the limits of our study. Nevertheless, our findings emphasize the importance of contrasting the results of multiple datasets in order to achieve a stable taxonomic delimitation.

Our results provide new insights to phylogenetic relationships of Mammilloid clade estimated from markers inherited biparentally, especially with regard to the delimitation of *Mammillaria s.s.* Plastid and nuclear genomes have different evolutionary histories because they are subject to different modes of inheritance and have different rates of mutation (Drouin, 2008). However, the concordance of nuclear and plastid phylogenies indicate that there is a consistent pattern of evolution across different parts of the genome which would support the circumscription of genera. This finding is particularly noteworthy for the delimitation of *Mammillaria*, which according to both genomes is

non-monophyletic. *Mammillaria s.s.*, which our results suggest is equivalent to the Mammillaria clade recovered in both phylogenies, had previously been identified in phylogenetic studies of chloroplast genome with a broad taxonomic sampling (Breslin et al., 2021, Chincoya et al., 2023) and also in those studies with lower molecular and taxonomic sampling (Vázquez-Sánchez et al., 2013). This delimitation of *Mammillaria s. s*. includes *Mammilloydia candida*, that previously has been considered as *Mammillaria candida* (Scheidw.), but due to characteristics of testa seed was considered as independent genus (Buxbaum, 1951, Hunt, 1981). Notably, the Cochemieae clade is also concordant between plastid and nuclear phylogenies; however, our taxonomic sampling is insufficient to draw firm conclusions on this clade.

4.2. Evolutionary processes behind phylogenetic discordance

Discordance in shallow relationships was marked in the Mammillaria clade, which is the clade better taxonomically represented. The most notorious difference is the position of *Mammilloydia candida,* that only in nuclear topology was sister to the rest of the clade. This difference is not due to distinct heritability patterns between plastid and nucleus, since previously the position of *M. candida* was obtained from plastid genome analysis (Chincoya et al., 2023). Therefore, this may be explained by differences in datasets, as taxonomic sampling of filtering criteria. In contrast, distinct heritability patterns may be the source of differences in the relationships of the two subspecies of *M. heyderi.* Since plastid loci have a smaller effective population size than nuclear loci (Hudson and Coyne, 2002), and therefore requires less time to complete it coalescence process, possibly this result indicates that the 0.16 Mya since the divergence of these taxa (Chincoya et al., 2023) was enough to allow the chloroplast genome complete its lineage sorting, whereas in nuclear loci, this process is still in progress.

It is common that studies that search for causes of discordance find that various processes are operating in the studied taxa (e.g. Dong et al., 2022, Rose et al., 2021, Wang et al., 2021). Reticulate events (i.e. hybridization and introgression) are commonly invoked as causes of discordances in plants. In Cactaceae, it is common to assume that these processes are ubiquitous, based on specimens with intermediate morphologies to putative parents or in their capacity to form artificial hybrids (extensively discussed in Machado, 2008, Granados-Aguilar et al, 2022). In addition, there is evidence based on molecular studies of such processes (Granados-Aguilar et al, 2022, Arakaki et al.,

2021, Khan et al., 2020). In our study we only found support for reticulate evolution at deep nodes, but not at shallow relationships in a clade with high cytonuclear and gene-tree discordance. Our results are concordant with those obtained in a recent study that rejects that the reticulate events explain the high levels of phylogenetic discordance estimated in a species complex (*M. haageana* complex) with morphological similarity (Cervantes *et al.*, 2023). Since the Pleistocene epoch is considered an epoch that facilitate interespecific gene flow due to geographic range shifts (Vuilleumier, 1971, Žerdoner et al, 2021, Folk et al., 2023) and Mammilloid clade experience even large scale geographic expansions in this epoch (Chincoya et al., 2023), we did not discard hybridization or introgression in other taxa of Mammilloid clade. Nevertheless, our results suggest that reticulate events are not the main cause of the widespread phylogenetic discordance observed in the taxa of the Mammilloid clade.

In addition to events of interspecific gene flow, the stochasticity in the transmission and sort of alleles can be the cause of the observed discordance. This is particularly true in cases where ancestral polymorphisms persist during speciation events, resulting in incomplete lineage sorting (ILS) and gene trees that are inconsistent with the species tree. Some lineages are more susceptible to ILS due to specific patterns and modes of speciation. This process is particularly common in phylogenetic patterns characterized by short internodes typical of rapid diversifications; but also in “young” lineages with shallow relationships (Degnan, 2009). In Cactaceae in general and in the Mammilloid clade in particular are present these two factors that favouring the presence of ILS. The Mammilloid clade is younger than 8 My (Arakaki et al., 2011, Hernandez-Hernandez et al., 2014, Chincoya et al., 2023) with most of the divergence events occurred in Pliocene and Pleistocene (Chincoya et al., 2023). Added to the recent origin of Mammilloid clade, its rapid diversification makes it one of the lineages with the highest diversification rates in the family (Hernandez-Hernandez et al., 2014). This suggests that ILS is the main evolutionary process that leads to the high levels of phylogenetic discordance observed in the studied cacti and likely in the whole family.

5. Conclusions

Our study provides valuable insights into the evolutionary processes underlying phylogenetic discordance in the Mammilloid clade. Our findings suggest that ILS is the main source of phylogenetic discordance observed, but highlights the potential role of reticulate evolution in shaping the deep nodes of the Mammilloid clade. We emphasize the need for caution in interpreting results from

molecular studies and applying them in the refinement of previous taxonomic classification systems. Nevertheless we consider that the concordance of nuclear and plastid phylogenies can support the circumscription of genera, especially with regard to the delimitation of *Mammillaria s.s.* Future studies should aim to increase taxonomic sampling of taxa from *Mammillaria* and allied genera to provide a more comprehensive understanding of their relationships. Additionally, we recommend specific phylogeographic studies in order to identify punctual events of introgression or hybridization.

6. References

1. Arakaki, M.; Christin, P.A.; Nyffeler, R.; Lendel, A.; Eggli, U.; Ogburn, M.; Spriggs, E.; Moore, M.J.; Edwards, E.J. Contemporaneous and recent radiations of the world's major succulent plant lineages. Proc. Natl. Acad. Sci. USA 2011, 108, 8379–8384.
2. Arakaki, M., Speranza, P., Soltis, P. S., & Soltis, D. E. (2021). Examination of reticulate evolution involving *Haageocereus* and *Espostoa*. *Haseltonia*, *27*(1), 102-112.
3. Breslin, P.B.; Wojciechowski, M.F.; Majure, L.C. Molecular phylogeny of the Mammilloid clade (Cactaceae) resolves the monophyly of *Mammillaria*. Taxon 2021, 70, 308–323.
4. Butterworth, C. A., & Wallace, R. S. (2004). Phylogenetic studies of *Mammillaria* (Cactaceae)—insights from chloroplast sequence variation and hypothesis testing using the parametric bootstrap. *American Journal of Botany*, *91*(7), 1086-1098.
5. Butterworth, C. A., Cota-Sanchez, J. H., & Wallace, R. S. (2002). Molecular systematics of tribe Cacteae (Cactaceae: Cactoideae): a phylogeny based on rpl16 intron sequence variation. *Systematic Botany*, *27*(2), 257-270.
6. Buxbaum, F. 1951. Die phylogenie der nordamerikanischen Echinocacteen.Trib. Euechinocactineae F. Buxb. Österreichische botanische Zeitschrift 98: 44–104.
7. Byrne, M., Tischler, G., Macdonald, B., Coates, D. J., & McComb, J. (2001). Phylogenetic relationships between two rare acacias and their common, widespread relatives in south-western Australia. *Conservation Genetics*, *2*, 157-166.
8. Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972-1973.

9. Carrive, L., Domenech, B., Sauquet, H., Jabbour, F., Damerval, C., & Nadot, S. (2020). Insights into the ancestral flowers of Ranunculales. *Botanical Journal of the Linnean Society*, *194*(1), 23-46.
10. Cervantes, C. R., Montes, J. R., Rosas, U., & Arias, S. (2023). Phylogenetic discordance and integrative species delimitation in the *Mammillaria haageana* species complex (Cactaceae). *Molecular Phylogenetics and Evolution*, 107891.
11. Chincoya, D. A., Arias, S., Vaca-Paniagua, F., Dávila, P., & Solórzano, S. (2023). Phylogenomics and biogeography of the Mammilloid clade revealed an intricate evolutionary history arose in the Mexican Plateau. *Biology*, *12*(4), 512.
12. Copetti, D., Búrquez, A., Bustamante, E., Charboneau, J. L., Childs, K. L., Eguiarte, L. E., ... & Sanderson, M. J. (2017). Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences*, *114*(45), 12003-12008.
13. Deng, M., Jiang, X. L., Hipp, A. L., Manos, P. S., & Hahn, M. (2018). Phylogeny and biogeography of East Asian evergreen oaks (Quercus section Cyclobalanopsis; Fagaceae): insights into the Cenozoic history of evergreen broad-leaved forests in subtropical Asia. *Molecular Phylogenetics and Evolution*, *119*, 170-181.
14. Dong, W., Li, E., Liu, Y., Xu, C., Wang, Y., Liu, K., ... & Zhou, S. (2022). Phylogenomic approaches untangle early divergences and complex diversifications of the olive plant family. *BMC biology*, *20*(1), 1-25.
15. Drouin, G., Daoud, H., & Xia, J. (2008). Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular phylogenetics and evolution*, *49*(3), 827-831.
16. Folk, R. A., Gaynor, M. L., Engle-Wrye, N. J., O'Meara, B. C., Soltis, P. S., Soltis, D. E., ... & Okuyama, Y. (2023). Identifying climatic drivers of hybridization with a new ancestral niche reconstruction method. *Systematic Biology*, syad018.
17. Gamisch, A., Winter, K., Fischer, G. A., & Comes, H. P. (2021). Evolution of crassulacean acid metabolism (CAM) as an escape from ecological niche conservatism in Malagasy Bulbophyllum (Orchidaceae). *New Phytologist*, *231*(3), 1236-1248.

18. Gao, L., SU, Y. J., & Wang, T. (2010). Plastid genome sequencing, comparative genomics, and phylogenomics: current status and prospects. *Journal of Systematics and Evolution*, *48*(2), 77-93.
19. Gong, Y., Ling, H., Lv, G., Chen, Y., Guo, Z., & Cao, J. (2019). Disentangling the influence of aridity and salinity on community functional and phylogenetic diversity in local dryland vegetation. *Science of the Total Environment*, *653*, 409-422.
20. Granados-Aguilar, X., Rosas, U., González-Rodríguez, A., & Arias, S. (2022). The prickly problem of interwoven lineages: hybridization processes in Cactaceae. *Botanical Sciences*, *100*(4), 797-813.
21. Hernández-Hernández, T., Brown, J. W., Schlumpberger, B. O., Eguiarte, L. E., & Magallón, S. (2014). Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New phytologist*, *202*(4), 1382-1397.
22. Hudson, R. R., & Coyne, J. A. (2002). Mathematical consequences of the genealogical species concept. *Evolution*, *56*(8), 1557-1565.
23. Johnson, M. G., Pokorny, L., Dodsworth, S., Botigue, L. R., Cowan, R. S., Devault, A., ... & Wickett, N. J. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic biology*, *68*(4), 594-606.
24. Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., ... & Wickett, N. J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in plant sciences*, *4*(7), 1600016.
25. Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Bioinformatics for DNA sequence analysis*, 39-64.
26. Khan, G., Franco, F. F., Silva, G. A., Bombonato, J. R., Machado, M., Alonso, D. P., ... & Moraes, E. M. (2020). Maintaining genetic integrity with high promiscuity: Frequent hybridization with low introgression in multiple hybrid zones of Melocactus (Cactaceae). *Molecular Phylogenetics and Evolution*, *142*, 106642.
27. Köhler, M., Oakley, L. J., Font, F., Peñas, M. L. L., & Majure, L. C. (2021). On the continuum of evolution: a putative new hybrid speciation event in Opuntia (Cactaceae) between a native

and an introduced species in southern South America. *Systematics and Biodiversity*, *19*(8), 1026-1039.

28. Korotkova, N., Aquino, D., Arias, S., Eggli, U., Franck, A., Gómez-Hinostrosa, C., ... & Berendsohn, W. G. (2021). Cactaceae at Caryophyllales.org–a dynamic online species-level taxonomic backbone for the family. *Willdenowia*, *51*(2), 251-270.
29. Leao, T. C., Fonseca, C. R., Peres, C. A., & Tabarelli, M. (2014). Predicting extinction risk of Brazilian Atlantic Forest angiosperms. *Conservation Biology*, *28*(5), 1349-1359.
30. Lemmon, E. M., & Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 99-121.
31. Liu, S., Zhu, H., & Yang, J. (2017). A phylogenetic perspective on biogeographical divergence of the flora in Yunnan, Southwestern China. *Scientific Reports*, *7*(1), 1-10.
32. Machado, M. C. (2008). What is the role of hybridization in the evolution of the Cactaceae?. *Bradleya*, *2008*(26), 1-18.
33. McLay, T. G., Birch, J. L., Gunn, B. F., Ning, W., Tate, J. A., Nauheimer, L., ... & Jackson, C. J. (2021). New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences*, *9*(7).
34. Molina-Venegas, R., Ramos-Gutiérrez, I., & Moreno-Saiz, J. C. (2020). Phylogenetic patterns of extinction risk in the endemic flora of a Mediterranean hotspot as a guiding tool for preemptive conservation actions. *Frontiers in Ecology and Evolution*, *8*, 571587.
35. Moritz, C. (1994). Defining 'evolutionarily significant units' for conservation. *Trends in ecology & evolution*, *9*(10), 373-375.
36. Morrison III, W. R., Lohr, J. L., Duchen, P., Wilches, R., Trujillo, D., Mair, M., & Renner, S. S. (2009). The impact of taxonomic change on conservation: Does it kill, can it save, or is it just irrelevant?. *Biological conservation*, *142*(12), 3201-3206.
37. Peruzzi, L. (2023). Advances in Plant Taxonomy and Systematics. *Biology*, *12*(4), 570.
38. Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one*, *5*(3), e9490.
39. Rose, J. P., Toledo, C. A., Lemmon, E. M., Lemmon, A. R., & Sytsma, K. J. (2021). Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant

genus Polemonium (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Systematic Biology*, *70*(1), 162-180.

40. Sánchez, D., Vázquez-Benítez, B., Vázquez-Sánchez, M., Aquino, D., & Arias, S. (2022). Phylogenetic relationships in Coryphantha and implications on Pelecyphora and Escobaria (Cacteae, Cactoideae, Cactaceae). *PhytoKeys*, *188*, 115.
41. Serna-Sánchez, M. A., Pérez-Escobar, O. A., Bogarín, D., Torres-Jimenez, M. F., Alvarez-Yela, A. C., Arcila-Galvis, J. E., ... & Arias, T. (2021). Plastid phylogenomics resolves ambiguous relationships within the orchid family and provides a solid timeframe for biogeography and macroevolution. *Scientific Reports*, *11*(1), 6858.
42. Smith, S. A., Moore, M. J., Brown, J. W., & Yang, Y. (2015). Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC evolutionary biology*, *15*, 1-15.
43. Song, Y. G., Fragnière, Y., Meng, H. H., Li, Y., Bétrisey, S., Corrales, A., ... & Kozlowski, G. (2020). Global biogeographic synthesis and priority conservation regions of the relict tree family Juglandaceae. *Journal of Biogeography*, *47*(3), 643-657.
44. Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., ... & Mazel, F. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, *92*(2), 698-715.
45. Vaughan, T. G. (2017). IcyTree: rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics*, *33*(15), 2392-2394.
46. Vázquez-Sánchez, M., Terrazas, T., Arias, S., & Ochoterena, H. (2013). Molecular phylogeny, origin and taxonomic implications of the tribe Cacteae (Cactaceae). *Systematics and Biodiversity*, *11*(1), 103-116.
47. Vuilleumier, B. S. (1971). Pleistocene Changes in the Fauna and Flora of South America: Present speciation patterns of the South American biota resulted from Pleistocene climatic changes. *Science*, *173*(3999), 771-780.
48. Wang, H. X., Morales‐Briones, D. F., Moore, M. J., Wen, J., & Wang, H. F. (2021). A phylogenomic perspective on gene tree conflict and character evolution in Caprifoliaceae using target enrichment data, with Zabelioideae recognized as a new subfamily. *Journal of Systematics and Evolution*, *59*(5), 897-914.

49. Wen, D., Yu, Y., Zhu, J., & Nakhleh, L. (2018). Inferring phylogenetic networks using PhyloNet. *Systematic biology*, *67*(4), 735-740.
50. Yao, G., Jin, J. J., Li, H. T., Yang, J. B., Mandala, V. S., Croley, M., ... & Li, D. Z. (2019). Plastid phylogenomic insights into the evolution of Caryophyllales. *Molecular Phylogenetics and Evolution*, *134*, 74-86.
51. Žerdoner Čalasan, A., Hurka, H., German, D. A., Pfanzelt, S., Blattner, F. R., Seidl, A., & Neuffer, B. (2021). Pleistocene dynamics of the Eurasian steppe as a driving force of evolution: Phylogenetic history of the genus Capsella (Brassicaceae). *Ecology and evolution*, *11*(18), 12697-12713.
52. Zhang, Q., Zhao, L., Folk, R. A., Zhao, J. L., Zamora, N. A., Yang, S. X., ... & Yu, X. Q. (2022). Phylotranscriptomics of Theaceae: generic-level relationships, reticulation and whole-genome duplication. *Annals of botany*, *129*(4), 457-471.
53. Zhang, Z., Wang, X. M., Liao, S., Zhang, J. H., & Li, H. Q. (2020). Phylogenetic reconstruction of Ficus subg. Synoecia and its allies (Moraceae), with implications on the origin of the climbing habit. *Taxon*, *69*(5), 927-945.

DISCUSIÓN GENERAL

La investigación derivada de esta tesis representa un aporte significativo al conocimiento de la historia evolutiva del género *Mammillaria* y de la familia Cactaceae en general. Particularmente, el origen reciente y la rápida diversificación (Arakaki *et al*., 2011, Hernández-Hernández *et al*., 2014) del género *Mammillaria* representan un reto para resolver sus relaciones filogenéticas. A pesar de estas circunstancias, en esta investigación se obtuvo una filogenia completamente resuelta con la mayor representación taxonómica del género *Mammillaria* hasta la fecha.

En el capítulo que integra filogenómica con biogeografía histórica (Capítulo 1) identificamos tres grandes clados en los que se agrupan las especies del clado Mammilloide (*Coryphantha, Escobaria, Mammillaria, Mammilloydia, Neolloydia, Pelecyphora* y *Ortegocactus*). Estos tres grandes clados, a su vez, pueden dividirse en nueve linajes distintos de acuerdo a sus patrones biogeográficos. Las especies del género *Mammillaria* se segregaron en seis de estos linajes, los cuales se originaron a partir de cuatro ancestros distintos, confirmando el origen polifilético del género identificado en estudios previos (Butterworth *et al*., 2002, Butterworth y Wallace, 2004, Crozier, 2005, Bárcenas *et al*., 2011, Breslin *et al*., 2021). En el artículo proponemos que uno de los linajes que identificamos, el linaje 2, podría ser equivalente a *Mammillaria sensu stricto*. El 85% de las especies que constituyen a este linaje pertenecen al subgénero *Mammillaria*, el resto provienen de los subgéneros *Dolichothele, Krainzia, Mammillopsis, Oehmea* y *Phellosperma.* Lo anterior contrasta con la propuesta de Crozier (2005), quien sugería una circunscripción del género limitada solo a una parte del subgénero *Mammillaria.* Personalmente, considero más apropiado extender a *Mammillaria sensu stricto* mediante la inclusión del linaje hermano al linaje 2, compuesto por *Mammillaria albiflora* y *Mammilloydia candida.* De lo contrario, la exclusión de estas dos especies tendría la inevitable consecuencia de unificarlas bajo el género *Mammilloydia*, sin embargo, no comparten características morfológicas que, a su vez, las distingan de su grupo hermano. Además, priorizar el establecimiento de géneros

más grandes suele favorecer a la practicidad y a la estabilidad del sistema de clasificación (Humphreys y Linder, 2009).

La consideración de los patrones biogeográficos conduce a tomar con cautela la propuesta de Breslin *et al*. (2021) de unificar bajo el género *Cochemiea* a los miembros de *Neolloydia*, *Ortegocactus* y parte de *Mammillaria*. Los taxa agrupados bajo el género *Cochemiea* muestran una gran disparidad en cuanto a morfología e historias biogeográficas, por lo que argumentamos que se trata de una decisión precipitada. Si bien, nosotros no rechazamos esta propuesta, hacemos un llamado a recibirla de forma crítica y esperar a incorporar nuevas líneas de evidencia (*i. e.* filogenias estimadas a partir de genoma nuclear y caracteres morfológicos, ecológicos o anatómicos) que permitan respaldar o rechazar este planteamiento.

El análisis de biogeografía histórica permitió identificar a la Altiplanicie Mexicana como el área en donde se originó el clado Mammilloide y donde ocurrieron los eventos de divergencia tempranos. Fue hasta los últimos 4.5 Ma, bajo condiciones climáticas similares a las actuales (Herbert *et al*., 2016), que los miembros del clado Mammilloide diversificaron rápidamente y salieron de la Altiplanicie Mexicana para colonizar otras regiones áridas de Norteamérica. La colonización de algunas de las regiones más distantes a su centro de origen (*e. g.* California, Península de Yucatán, Valle de Tehuacán-Cuicatlán) ocurrió en los últimos 2.5 Ma, durante el Pleistoceno. De acuerdo a los resultados, el clado Mammilloide colonizó las regiones áridas del sur de México antes de que ocurriera la elevación de la parte central de la Faja Volcánica Transmexicana. Esto sugiere que la Faja Volcánica Transmexicana representó una barrera biogeográfica para los cactus al interrumpir la continuidad de las regiones áridas del Norte (Altiplanicie Mexicana) y Sur de México (Valle de Tehuacán-Cuicatlán y Depresión del Balsas). De manera que la historia evolutiva y la distribución geográfica del clado Mammilloide ha sido configurada tanto por factores climáticos globales como por eventos orográficos locales.

En el Capítulo 2 realizamos un evaluación *in silico* de la eficiencia de dos conjuntos de sondas universales para angiospermas: Angiosperm v.1 (Buddenhagen *et al*., 2016) y

Angiosperms353 (Johnson *et al*., 2019), para recuperar loci nucleares en Cactaceae. Esta evaluación, que nos permitió elegir el método de muestreo de genoma nuclear presentado en el Capítulo 3, fue pensada como una herramienta para la comunidad científica interesada en resolver relaciones filogenéticas en la familia Cactaceae y el orden Caryophyllales. Con base en los resultados, recomendamos al conjunto de sondas Angiosperms353 para la realización de estudios filogenómicos en Cactaceae. En promedio, este conjunto de sondas permite recuperar 123,687 pb en las especies de Cactaceae, las cuales corresponden a ~80% de las regiones objetivo (276 loci) y <3% son afectadas por paralogía.

En contraste al enfoque "universal" en el diseño de sondas que busca recuperar loci en una gran amplitud taxonómica, existe un enfoque "específico", en el que las sondas se diseñan para un grupo taxonómico en particular. Hasta la fecha se han publicado dos conjuntos de sondas específicas para la familia Cactaceae (Acha y Majure, 2022, Romeiro-Brito *et al.,* 2022). Las sondas específicas no necesariamente tienen un mejor desempeño que las sondas universales en el grupo para el que fueron diseñadas (*e. g.* Larridon *et al.*, 2020, Romeiro-Brito *et al.,* 2022). Estudios *in silico* similares al que realizamos permitirán decidir entre ambas aproximaciones, o incluso combinarlas (Hendriks e*t al*., 2021), antes de invertir recursos en alguna de ellas. Sin embargo, la gran ventaja de las sondas universales frente a las específicas es que facilitan la reutilización de los datos de secuenciación. El uso del conjunto de sondas Angiosperms353 está ampliamente extendido (Baker *et al*., 2021, McDonnell *et al.*, 2021); hasta mayo de 2023 se cuenta con los datos de secuenciación de las bibliotecas enriquecidas con este conjunto para más de 10,500 plantas (Sequence Read Archive, 2009). Por lo tanto, al utilizar sondas universales se contribuye a la formación de un banco de cientos de loci ortólogos que hacen factible abordar preguntas evolutivas con una gran amplitud taxonómica.

En el Capítulo 3 hicimos una evaluación de la discordancia filogenética entre compartimentos genómicos (*i. e.* cloroplasto y núcleo) y entre loci del genoma nuclear. Las discordancias entre las filogenias estimadas con distintos compartimentos genómicos eran esperables debido a las diferencias en los patrones de heredabilidad y tasas de mutación

(Drouin, 2008). Sin embargo, ambas filogenias fueron concordantes respecto al clado compuesto por *Mammilloydia candida* y especies del género *Mammillaria,* también identificado con un muestreo taxonómico mayor (Capitulo 1), proporcionando nueva evidencia que apoya la circunscripción de *Mammillaria sensu stricto*.

La evaluación de la discordancia filogenética permitió identificar que a pesar del alto valor de soporte del árbol de especies de la matriz concatenada existen múltiples historias evolutivas discordantes en los árboles de genes. La estimación de redes filogenéticas permitió identificar que la discordancia entre las relaciones profundas del clado Mammilloide probablemente tiene su origen en eventos de evolución reticulada (*i. e.* hibridación e introgresión). En particular, el clado conformado por *Escobaria* y *Pelecyphora* probablemente tiene un origen híbrido. Históricamente se ha considerado que en las especies que constituyen el clado Mammilloide los procesos de evolución reticulada han sido frecuentes en su historia evolutiva (*e.g.* Remski *et al.*, 1954, Hunt, 2006). Hasta la fecha no existen estudios que confirmen el origen híbrido de especies pertenecientes al clado Mammilloide; los eventos de hibridación en poblaciones naturales han sido asumidos principalmente por la presencia de especímenes con características intermedias (Hunt, 1987, Hunt, 2006). En contraste, un estudio reciente que evaluó la hibridación en el complejo de especies de *M. haageana* rechazó la hipótesis de que hayan ocurrido eventos de hibridación entre las especies analizadas (Cervantes *et al.*, 2023). Nuestros resultados de redes filogenéticas tampoco apoyan a los eventos de evolución reticulada como procesos que hayan configurado la historia evolutiva de las especies a niveles taxonómicos someros. Lo anterior sugiere que el principal proceso evolutivo causante de discordancia filogenética en el clado Mammilloide es el sorteo incompleto de linajes, lo cual es compatible con su rápida diversificación.

La alta discordancia filogenética documentada en esta tesis, y en otros estudios realizados en la familia Cactaceae (Walker *et al*., 2018, Copetti *et al.*, 2017), podría presentarse también en genes responsables de la expresión fenotípica. El sorteo incompleto de linajes podría afectar también a aquellos genes asociados a la expresión de carácteres

morfológicos. Si los genes asociados a caracteres morfológicos fuesen discordantes, entonces la distribución de los estados de carácter en los taxa no reflejaría los procesos de divergencia. Recientemente, se demostró que los clados con alta discordancia entre árboles de genes coinciden con mayor variación morfológica en aves, mamíferos y plantas (Parins-Fukuchi *et al.*, 2021). La familia Cactaceae destaca por su diversidad morfológica, pero también por la frecuencia con que se presenta evolución convergente en muchos rasgos (Larridon *et al*., 2015, Hernández-Hernández *et al*., 2011). Por lo tanto, es posible que la convergencia morfológica que se ha documentado en Cactaceae, y que ha dificultado su taxonomía, sea causada en parte por el sorteo incompleto de linajes.

CONCLUSIONES

Esta investigación demuestra la utilidad de los datos genómicos para dilucidar los patrones de diversificación en uno de los escenarios biológicos más desafiantes: el que se presenta en aquellos linajes de origen reciente y rápida diversificación. Más allá de obtener una hipótesis robusta sobre las relaciones filogenéticas del clado Mammilloide, el nivel de resolución filogenética obtenido abrió la posibilidad de abordar otras preguntas evolutivas y revisitar los pendientes taxonómicos añejos. Los resultados de esta tesis permitieron comprender cómo los cambios climáticos globales y los eventos orográficos locales influyeron en la diversificación y la expansión geográfica de las cactáceas estudiadas. Lejos de visualizar a la discordancia filogenética como un obstáculo metodológico, en esta tesis se contempla como una ventana al pasado que permite inferir los procesos biológicos que han moldeado los genomas y la historia evolutiva de las especies. Si bien, resultaría ambicioso y reduccionista asegurar que esta investigación logra zanjar la controversia de más de 200 años que ha generado la delimitación taxonómica de *Mammillaria*, es indudable que permite identificar a un clado que putativamente corresponde a *Mammillaria sensu stricto.*

Para continuar construyendo el conocimiento del clado Mammilloide, se recomienda mantener la perspectiva filogenómica y aumentar el muestreo taxonómico, particularmente de los géneros *Coryphantha* y *Escobaria.* Esta investigación demostró la relevancia de la Altiplanicie Mexicana para el origen y diversificación del clado Mammilloide, sin embargo, debido a su amplitud geográfica, se requiere de estudios detallados dentro de la Altiplanicie. La realización de estudios biogeográficos dentro de la Altiplanicie permitiría identificar la región específica en donde se originó el clado Mammilloide, así como aquellas en las que han ocurrido los eventos de divergencia. Además, debido a que el sorteo incompleto de linajes, un proceso microevolutivo, se invoca como el principal causante de discordancia filogenética, se recomienda la realización de estudios con muestreo poblacional, para evaluar la discordancia entre especies cercanas. El muestreo a nivel poblacional también

permitiría abordar otra controversia común en el clado Mammilloide: la delimitación taxonómica a nivel interespecífico. Por último, lejos de demeritar el uso de caracteres morfológicos, anatómicos y ecológicos para establecer límites taxonómicos, esta tesis pugna por adoptar una visión integral que favorezca la estabilidad de los sistemas de clasificación.

REFERENCIAS BIBLIOGRÁFICAS

1. Anderson, E. F. (1986). A revision of the genus Neolloydia B. & R. (Cactaceae). *Bradleya*, *1986*(4), 1-28.
2. Anderson, E. F. (2001). *The Cactus Family*. Oregon: Timber Press.
3. Arakaki, M., Christin, P. A., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R. M., ... y Edwards, E. J. (2011). Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proceedings of the National Academy of Sciences*, *108*, 8379-8384.
4. Acha, S. y Majure, L. C. (2022). A new approach using targeted sequence capture for phylogenomic studies across Cactaceae. *Genes*, *13*(2), 350.
5. Arias, S., Gama S., Guzmán-Cruz, L. U. (1997). Flora del valle de Tehuacán-Cuicatlán. Fascículo 14: Cactaceae A. L. Juss. Instituto de Biología, Universidad Nacional Autónoma de México, México, DF, Mexico.
6. Baker, W. J., Dodsworth, S., Forest, F., Graham, S. W., Johnson, M. G., McDonnell, A., … y Wickett, N. J. (2021). Exploring Angiosperms353: An open, community toolkit for collaborative phylogenomic research on flowering plants.
7. Bárcenas, R. T., Yesson, C. y Hawkins, J. A. (2011). Molecular systematics of the Cactaceae. *Cladistics*, *27*, 470-489.
8. Bog, M., Xu, S., Himmelbach, A., Brandt, R., Wagner, F., Appenroth, K. J., Sree, K. S. (2020). Genotyping-by-sequencing for species delimitation in Lemna section Uninerves Hegelm. (Lemnaceae). En: X. H. Cao, P. Fourounjian, W. Wang (Eds.), *The Duckweed Genomes* (pp. 115–123). Berlín: Springer.
9. Breslin, P. B., Wojciechowski, M. F., y Majure, L. C. (2021). Molecular phylogeny of the Mammilloid clade (Cactaceae) resolves the monophyly of *Mammillaria*. *Taxon*, *70*(2), 308-323.

10. Britton, N.L. y Rose. J.N. (1923). *The Cactaceae: descriptions and illustrations of plants of the cactus family.* Volúmen IV. Washington: Carnegie Institution of Washington.
11. Buddenhagen, C., Lemmon, A. R., Lemmon, E. M., Bruhl, J., Cappa, J., Clement, W. L., ... y Mast, A. (2016). Anchored phylogenomics of angiosperms I: assessing the robustness of phylogenetic estimates. *BioRxiv*, 086298.
12. Butterworth, C. A., Cota, J. H. y Wallace, R. S. (2002). Molecular phylogenetics of the tribe Cacteae (Cactoideae: Cactaceae) using rpl16 intron sequence data. *Systematic Botany*, *27*, 257-270.
13. Butterworth, C. A. y Wallace, R. S. (2004). Phylogenetic studies of *Mammillaria* (Cactaceae)—insights from chloroplast sequence variation and hypothesis testing using the parametric bootstrap. *American Journal of Botany*, *91*, 1086-1098.
14. Buxbaum, F. (1951). Die Phylogenie der nordamerikanischen Echinocacteen. Trib. Euechinocactineae F. Buxb. *Oesterreichische Botanische Zeitschrift*, *98*, 44-104.
15. Carrive, L., Domenech, B., Sauquet, H., Jabbour, F., Damerval, C., & Nadot, S. (2020). Insights into the ancestral flowers of Ranunculales. *Botanical Journal of the Linnean Society*, 194(1), 23-46.
16. Cervantes, C. R., Montes, J. R., Rosas, U., & Arias, S. (2023). Phylogenetic discordance and integrative species delimitation in the *Mammillaria haageana* species complex (Cactaceae). *Molecular Phylogenetics and Evolution*, 107891.
17. Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., … y Stevens, P. F. (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. J *Botanical Journal of the Linnean Society, 181*, 1–20.
18. Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., ... y Albert, V. A. (1993). Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Annals of the Missouri Botanical Garden*, 528-580.

19. Copetti, D., Búrquez, A., Bustamante, E., Charboneau, J. L., Childs, K. L., Eguiarte, L. E., ... y Sanderson, M. J. (2017). Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences*, *114*(45), 12003-12008.
20. Crozier, B.S. (2005). Systematics of Cactaceae Juss.: phylogeny, cpDNA evolution, and classification, with emphasis on the genus *Mammillaria* Haw. Tesis de Doctorado. Universidad de Texas. Estados Unidos de América.
21. Delsuc, F., Brinkmann, H. y Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, *6*, 361-375.
22. Dicht, R. y Lüthy, A. (2005). *Coryphantha. Cacti of Mexico and Southern USA*. Berlin: Springer-Verlag.
23. Drouin, G., Daoud, H., y Xia, J. (2008). Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular phylogenetics and evolution*, *49*(3), 827-831.
24. Edwards, E. J., Nyffeler, R. y Donoghue, M. J. (2005). Basal cactus phylogeny: implications of *Pereskia* (Cactaceae) paraphyly for the transition to the cactus life form. *American Journal of Botany*, *92*(7), 1177-1188.
25. Eisen, J. A. y Fraser, C. M. (2003). Phylogenomics: intersection of evolution and genomics. *Science*, *300*, 1706-1707.
26. Engelmann, G. (1856). Synopsis of the Cactaceae of the territory of the United States and adjacent regions. *Proceedings of the American Academy. 3*: 259–346.
27. Gamisch, A., Winter, K., Fischer, G. A., y Comes, H. P. (2021). Evolution of crassulacean acid metabolism (CAM) as an escape from ecological niche conservatism in Malagasy *Bulbophyllum* (Orchidaceae). *New Phytologist*, *231*(3), 1236-1248.
28. García, N., Folk, R. A., Meerow, A. W., Chamala, S., Gitzendanner, M. A., de Oliveira, R. S., … y Soltis, P. S. (2017). Deep reticulation and incomplete lineage

sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Molecular Phylogenetics and Evolution*, *111*, 231-247.

29. Gitzendanner, M. A., Soltis, P. S., Wong, G. K. S., Ruhfel, B. R. y Soltis, D. E. (2018). Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *American Journal of Botany*, *105*(3), 291-301.
30. Griffith, M. P. y Porter, J. M. (2009). Phylogeny of Opuntioideae (Cactaceae). *International Journal of Plant Sciences*, *170*, 107-116.
31. Harpke, D., Peterson, A., Hoffmann, M. H., y Röser, M. (2006). Phylogenetic evaluation of chloroplast trnL–trnF DNA sequence variation in the genus *Mammillaria* (Cactaceae). *Schlechtendalia*, *14*, 7-16.
32. Haworth, A.H. (1812) *Synopsis Plantarum Succulentarum: cum descriptionibus, synonymis, locis, observationibus anglicanis, culturaque.* Londrés: Richard Taylor.
33. Herbert, T. D., Lawrence, K. T., Tzanova, A., Peterson, L. C., Caballero-Gill, R. y Kelly, C. S. (2016). Late Miocene global cooling and the rise of modern ecosystems. *Nature Geoscience*, *9*(11), 843-847.
34. Hendriks, K. P., Mandáková, T., Hay, N. M., Ly, E., Hooft van Huysduynen, A., Tamrakar, R., … y Bailey, C. D. (2021). The best of both worlds: Combining lineage‐specific and universal bait sets in target‐enrichment hybridization reactions. *Applications in Plant Sciences*, *9*(7).
35. Hernandez, H. M., y Godinez, H. (1994). Contribución al conocimiento de las cactáceas mexicanas amenazadas. *Acta Botánica Mexicana*, (26), 33-52.
36. Hernández, H. M. y Gómez-Hinostrosa C. (2015). *Mapping the Cacti of Mexico, their geographical distribution based on referenced records. Part II Mammillaria*. Milbourne Port: Dh books.
37. Hernández‐Hernández, T., Brown, J. W., Schlumpberger, B. O., Eguiarte, L. E. y Magallón, S. (2014). Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New phytologist*, *202*, 1382-1397.

38. Hernández-Hernández, T., Hernández, H. M., De-Nova, J. A., Puente, R., Eguiarte, L. E. y Magallón, S. (2011). Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *American journal of botany*, *98*(1), 44-61.
39. Humphreys, A. M. y Linder, H. P. (2009). Concept versus data in delimitation of plant genera. *Taxon*, *58*(4), 1054-1074.
40. Hunt, D. R. (1971). Schumann and Buxbaum reconciled. *The Cactus and Succulent Journal of Great Britain*, 53-72.
41. Hunt, D. R. (1981). Revised classified list of the genus Mammillaria. *The Cactus and Succulent Journal of Great Britain*, *43*(2/3), 41-48.
42. Hunt, D. R, (1987). A new review of *Mammillaria* names SZ. *Bradleya*, *1987*(5), 17-48.
43. Hunt, D. R. (2006). *The new cactus lexicon, text and atlas*. Milbourne Port: Dh books.
44. Jiao, B., Chen, C., Wei, M., Niu, G., Zheng, J., Zhang, G., … y Gao, T. (2023). Phylogenomics and morphological evolution of the mega-diverse genus *Artemisia* (Asteraceae: Anthemideae): implications for its circumscription and infrageneric taxonomy. *Annals of Botany*, *131*, 867-883.
45. Johnson, M. G., Pokorny, L., Dodsworth, S., Botigue, L. R., Cowan, R. S., Devault, A., ... y Wickett, N. J. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic biology*, *68*(4), 594-606.
46. Kadereit, J. W., Albach, D. C., Ehrendorfer, F., Galbany-Casals, M., Garcia-Jacas, N., Gehrke, B., … y Dillenberger, M. S. (2016). Which changes are needed to render all genera of the German flora monophyletic?. *Willdenowia*, *46*, 39-91.
47. Korotkova, N., Aquino, D., Arias, S., Eggli, U., Franck, A., Gómez-Hinostrosa, C., ... y Berendsohn, W. G. (2021). Cactaceae at Caryophyllales. org–a dynamic online species-level taxonomic backbone for the family. *Willdenowia*, *51*, 251-270.

48. Larridon, I., Villaverde, T., Zuntini, A. R., Pokorny, L., Brewer, G. E., Epitawalage, N., ... y Baker, W. J. (2020). Tackling rapid radiations with targeted sequencing. *Frontiers in plant science*, *10*, 1655.
49. Larridon, I., Walter, H. E., Guerrero, P. C., Duarte, M., Cisternas, M. A., Hernández, C. P., … y Samain, M. S. (2015). An integrative approach to understanding the evolution and diversity of *Copiapoa* (Cactaceae), a threatened endemic Chilean genus from the Atacama Desert. *American Journal of Botany*, *102*, 1506-1520.
50. Lemaire, C. (1868). *Les cacteés: histoire, patrie, organes de vegetation inflorescence, culture, etc*. Paris: Libraire Agricole de la Maison Rustique.
51. Low, S. L., Yu, C. C., Ooi, I. H., Eiadthong, W., Galloway, A., Zhou, Z. K. y Xing, Y. W. (2021). Extensive Miocene speciation in and out of Indochina: The biogeographic history of *Typhonium sensu stricto* (Araceae) and its implication for the assembly of Indochina flora. *Journal of Systematics and Evolution*, *59*(3), 419-428.
52. Lüthy, J. M. (1995). Taxonomische untersuchung der gattung *Mammillaria* Haw. Tesis de doctorado. Universität Bern. Suiza.
53. Machado, M. C. (2008). What is the role of hybridization in the evolution of the Cactaceae?. *Bradleya*, *2008*, 1-18.
54. Magallón, S., Gómez‐Acevedo, S., Sánchez‐Reyes, L. L. y Hernández‐Hernández, T. (2015). A metacalibrated time‐tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, *207*(2), 437-453.
55. Mandujano, M. C., Carrillo-Angeles, I., Martínez-Peralta, C. y Golubov, J. (2010). Reproductive Biology of Cactaceae. En: Ramawat K. (Ed.), *Desert Plants* (pp. 197-230). Berlín: Springer.
56. McDonnell, A. J., Baker, W. J., Dodsworth, S., Forest, F., Graham, S. W., Johnson, M. G., ... y Wickett, N. J. (2021). Exploring Angiosperms353: Developing and applying a universal toolkit for flowering plant phylogenomics. *Applications in plant sciences*, *9*(7).

57. Mutke, J. (2015). Cactus ecology and biogeography. En: Barthlott, W. (Ed.), *Biogeography and biodiversity of cacti* (pp. 13-23). Oldenburgo: Schumannia.
58. Paetzold, C., Wood, K. R., Eaton, D. A., Wagner, W. L. y Appelhans, M. S. (2019). Phylogeny of Hawaiian *Melicope* (Rutaceae): RAD-seq resolves species relationships and reveals ancient introgression. *Frontiers in plant science*, *10*, 1074.
59. Pamilo, P. y Nei, M. (1988). Relationships between gene trees and species trees. *Molecular biology and evolution*, *5*, 568-583.
60. Parins-Fukuchi, C., Stull, G. W. y Smith, S. A. (2021). Phylogenomic conflict coincides with rapid morphological innovation. *Proceedings of the National Academy of Sciences*, *118*, e2023058118.
61. Parks, M., Cronn, R. y Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC biology*, *7*, 1-17.
62. Peters, E. M., Martorell, C., y Ezcurra, E. (2008). Nurse rocks are more important than nurse plants in determining the distribution and establishment of globose cacti (*Mammillaria*) in the Tehuacán Valley, Mexico. *Journal of arid environments*, *72*(5), 593-601.
63. Pilbeam, J. (1999). *Mammillaria. The Cactus File Handbook 6.* Southampton: Cirio Publishing Services Ltd.
64. Ramirez-Barahona, S., Sauquet, H. y Magallon, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology & Evolution*, *4*(9), 1232-1238.
65. Remski, M. F. (1954). Cytological investigations in *Mammillaria* and some associated genera. *Botanical Gazette*, *116*(2), 163-171.
66. Ritz, C. M., Martins, L., Mecklenburg, R., Goremykin, V. y Hellwig, F. H. (2007). The molecular phylogeny of *Rebutia* (Cactaceae) and its allies demonstrates the influence of paleogeography on the evolution of South American mountain cacti. *American Journal of Botany*, *94*(8), 1321-1332.

67. Romeiro-Brito, M., Telhe, M. C., Amaral, D. T., Franco, F. F. y Moraes, E. M. (2022). A target Capture Probe Set Useful for Deep-and Shallow-Level Phylogenetic Studies in Cactaceae. *Genes*, *13*(4), 707.

68. Rose, J. P., Kriebel, R., Kahan, L., DiNicola, A., González-Gallegos, J. G., Celep, F., ... y Drew, B. T. (2021). Sage insights into the phylogeny of *Salvia*: dealing with sources of discordance within and across genomes. *Frontiers in Plant Science*, 2606.

69. Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E. y Burleigh, J. G. (2014). From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC evolutionary biology*, *14*(1), 1-27.

70. Sánchez, D., Vázquez-Benítez, B., Vázquez-Sánchez, M., Aquino, D. y Arias, S. (2022). Phylogenetic relationships in *Coryphantha* and implications on *Pelecyphora* and *Escobaria* (Cacteae, Cactoideae, Cactaceae). *PhytoKeys*, *188*, 115.

71. Schlumpberger, B. O. y Renner, S. S. (2012). Molecular phylogenetics of *Echinopsis* (Cactaceae): Polyphyly at all levels and convergent evolution of pollination modes and growth forms. *American Journal of Botany*, *99*, 1335-1349.

72. Schumann, K.M. (1899). *Gesamtbeschreibung der Kakteen (Monographia cactacearum)*. Berlín: Neudamm.

73. Sequence Read Archive (SRA) [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2009 - [18 de mayo de 2023]. Disponible en: https://www.ncbi.nlm.nih.gov/sra/?term=angiosperms353

74. Tamashiro, R. A., White, N. D., Braun, M. J., Faircloth, B. C., Braun, E. L. y Kimball, R. T. (2019). What are the roles of taxon sampling and model fit in tests of cyto-nuclear discordance using avian mitogenomic data?. *Molecular phylogenetics and evolution*, *130*, 132-142.

75. Villaverde, T., Pokorny, L., Olsson, S., Rincón-Barrado, M., Johnson, M. G., Gardner, E. M., … y Sanmartín, I. (2018). Bridging the micro-and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist*, *220*(2), 636-650.

76. Walker, J. F., Yang, Y., Feng, T., Timoneda, A., Mikenas, J., Hutchison, V., ... & Smith, S. A. (2018). From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *American Journal of Botany*, *105*(3), 446-462.
77. Whitfield, J. B., Lockhart, P. J. (2007). Deciphering ancient rapid radiations. *Trends in ecology & evolution*, *22*, 258-265.
78. Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., ... y Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences*, *111*(45), E4859-E4868.
79. Xia, M., Liu, Y., Liu, J., Chen, D., Shi, Y., Chen, Z., ... y Qiu, Y. (2022). Out of the Himalaya-Hengduan Mountains: phylogenomics, biogeography and diversification of *Polygonatum* Mill. (Asparagaceae) in the Northern Hemisphere. *Molecular Phylogenetics and Evolution*, *169*, 107431.
80. Xiang, Q. Y., Soltis, D. E. y Soltis, P. S. (1998). Phylogenetic relationships of Cornaceae and close relatives inferred from matK and rbcL sequences. *American Journal of Botany*, *85*, 285-297.
81. Yang, L. H., Shi, X. Z., Wen, F. y Kang, M. (2023). Phylogenomics reveals widespread hybridization and polyploidization in *Henckelia* (Gesneriaceae). *Annals of Botany*, mcad047.
82. Ye, X. Y., Ma, P. F., Yang, G. Q., Guo, C., Zhang, Y. X., Chen, Y. M., ... y Li, D. Z. (2019). Rapid diversification of alpine bamboos associated with the uplift of the Hengduan Mountains. *Journal of Biogeography*, *46*(12), 2678-2689.
83. Zhang, Z., Wang, X. M., Liao, S., Zhang, J. H., y Li, H. Q. (2020). Phylogenetic reconstruction of *Ficus* subg. *Synoecia* and its allies (Moraceae), with implications on the origin of the climbing habit. *Taxon*, 69(5), 927-945.