UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

BIOLOGIA EVOLUTIVA

VARIACIÓN DEL GENOMA MITOCONDRIAL EN *MAMMILLARIA* (CACTACEAE)

TESIS

(**POR ARTÍCULO CIENTÍFICO**)

STRUCTURAL AND GENE COMPOSITION VARIATION OF THE COMPLETE

MITOCHONDRIAL GENOME OF *MAMMILLARIA HUITZILOPOCHTLI* (CACTACEAE,

CARYOPHYLLALES), REVEALED BY *DE NOVO* ASSEMBLY

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

JOSÉ DAVID CRUZ PLANCARTE

TUTOR(A) PRINCIPAL DE TESIS: DRA. SOFÍA SOLÓRZANO LUJANO
FACULTAD DE ESTUDIOS SUPERIORES IZTACALA, UNAM
COMITÉ TUTOR: DR. DIEGO GONZALEZ HALPHEN
INSTITUTO DE FISIOLOGÍA CELULAR, UNAM
DR. FIDEL ALEJANDRO SÁNCHEZ FLORES
INSTITUTO DE BIOTECNOLOGÍA, UNAM

LOS REYES IZTACALA, TLALNEPANTLA, ESTADO DE MEXICO, 2023

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

BIOLOGIA EVOLUTIVA

VARIACIÓN DEL GENOMA MITOCONDRIAL EN *MAMMILLARIA* (CACTACEAE)

TESIS

(**POR ARTÍCULO CIENTÍFICO**)

STRUCTURAL AND GENE COMPOSITION VARIATION OF THE COMPLETE MITOCHONDRIAL GENOME OF *MAMMILLARIA HUITZILOPOCHTLI* (CACTACEAE, CARYOPHYLLALES), REVEALED BY *DE NOVO* ASSEMBLY

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

JOSÉ DAVID CRUZ PLANCARTE

TUTOR(A) PRINCIPAL DE TESIS: DRA. SOFÍA SOLÓRZANO LUJANO
FACULTAD DE ESTUDIOS SUPERIORES IZTACALA, UNAM
COMITÉ TUTOR: DR. DIEGO GONZALEZ HALPHEN
INSTITUTO DE FISIOLOGÍA CELULAR, UNAM
DR. FIDEL ALEJANDRO SÁNCHEZ FLORES
INSTITUTO DE BIOTECNOLOGÍA, UNAM

LOS REYES IZTACALA, TLALNEPANTLA, ESTADO DE MEXICO, 2023

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

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día **29 de agosto de 2022** se aprobó el siguiente jurado para el examen de grado de **MAESTRO EN CIENCIAS BIOLÓGICAS** en el campo de conocimiento de **Biología Evolutiva** del alumno **CRUZ PLANCARTE JOSÉ DAVID** con número de cuenta **414082920** por la modalidad de graduación de **tesis por artículo científico** titulado: **"Structural and gene composition variation of the complete mitochondrial genoma of *Mammillaria hutzilopochtli* (Cactaceae, Caryophyllales), revealed by de *novo* assembly",** que es producto del proyecto realizado en la maestría que lleva por título: **"Variación del genoma mitocondrial en *Mammillaria* (Cactaceae)",** ambos realizados bajo la dirección de la **DRA. SOFÍA SOLÓRZANO LUJANO,** quedando integrado de la siguiente manera:

Presidente:	**DRA. PATRICIA DOLORES DÁVILA ARANDA**
Vocal:	**DR. ÁNGEL SALVADOR ARIAS MONTES**
Vocal:	**DRA. MARÍA HILDA FLORES OLVERA**
Vocal:	**DRA. HELGA OCHOTERENA BOOTH**
Secretario:	**DR. FIDEL ALEJANDRO SÁNCHEZ FLORES**

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 28 de noviembre de 2022

COORDINADOR DEL PROGRAMA

DR. ADOLFO GERARDO NAVARRO SIGÜENZA

AGRADECIMIENTOS INSTITUCIONALES

AGRADECIMIENTOS A TÍTULO PERSONAL

Quiero agradecer a todas las personas con las que he coincidido a lo largo de mi maestría. A mi tutora, la Dra. Sofía Solórzano por toda la dedicación y el tiempo invertido en mi formación, fue una gran enseñanza en mi vida académica.

A los miembros de mi jurado que se tomaron el tiempo para revisar esta tesis, los Doctores Ángel Salvador Arias Montes y Fidel Alejandro Sánchez Flores; y las Doctoras Helga Ochoterena Booth, María Hilda Flores Olvera y Patricia Dolores Dávila Aranda, sus comentarios fueron importantes.

A la Mtra. Delil Chincoya, por el buen trato que recibí de ti, los consejos y el apoyo, en el laboratorio de molecular y en el laboratorio de cómputo.

A toda la gente de la UBIPRO, FES Iztacala. En particular a todas las personas del Banco de Semillas que siempre me recibieron con un cálido saludo y con quienes compartimos más de un pastel.

Entrando en el ámbito familiar. Primeramente, quisiera expresar el profundo agradecimiento hacia mis padres, Daniel Cruz y Rebeca Plancarte. Su amor, su apoyo y sus consejos fueron vitales, gracias por siempre estar ahí para mí y por haberme instruido en todos los aspectos de mi vida. Gracias a ustedes he cumplido las metas que me he propuesto hasta la fecha.

A mis hermanos. Van por orden alfabético, pero ustedes saben que son igualmente especiales para mí, Daniel, Indira y Jenner. Gracias por permanecer unidos, sin duda alguna compartimos una hermandad y una amistad como pocos hermanos logran tener. Estoy orgulloso por sus logros y me motivan a seguir mejorando. Las gracias también son infinitas, por el apoyo, las charlas y las experiencias que hemos vivido juntos.

Para concluir, y sin poder explicar con palabras la importancia de su compañía, quiero expresarle todo mi agradecimiento, mi cariño y mi amor a Larisa. Empezamos y finalizamos este proyecto juntos, es un logro de los dos.

ÍNDICE

ÍNDICE DE FIGURAS Y TABLAS

RESUMEN

El genoma mitocondrial de plantas terrestres es conocido por características particulares como su gran tamaño y el elevado número de rearreglos estructurales en comparación con el genoma mitocondrial en animales o el genoma de cloroplasto. A pesar de ser poco estudiados, los genomas mitocondriales de plantas ya caracterizados y disponibles públicamente muestran una gran diversidad de tamaño, estructura, secuencias repetidas y secuencias de origen de cloroplasto, núcleo y transferidas horizontalmente desde otras especies. La familia Cactaceae (Caryophyllales) no tiene ninguna especie con el genoma mitocondrial caracterizado, por lo tanto, en este estudio ensamblamos y caracterizamos el genoma mitocondrial de *Mammillaria huitzilopochtli*, el cual tiene una longitud de 2,050,004 pb y un contenido de GC de 42.9%. El análisis del genoma permitió identificar 34 genes codificantes de proteínas, 28 ARNs de transferencia y 3 ARNs ribosomales típicos de plantas terrestres, de estos, cinco genes eran copias de genes presentes en el cloroplasto. La comparación con otras plantas terrestres y especies del orden Caryophyllales reveló que el genoma mitocondrial aquí caracterizado se encuentra entre los más grandes del orden. El análisis filogenético mostró relaciones concordantes del orden Caryophyllales con respecto a otras filogenias reconstruidas con loci de cloroplasto. Finalmente, se propone al genoma mitocondrial como una importante fuente de variación que debería incluirse en futuros estudios que busquen elucidar cuestiones evolutivas en grupos de plantas terrestres.

ABSTRACT

The mitochondrial genomes of land plants are known for some particular features like their large size and the high number of structural rearrangements. Despite being poorly studied, the mitochondrial genomes of land plants already characterized and publicly available show a high diversity in size, structure, repeated sequences and sequences of chloroplast and nuclear origin as well as sequences horizontally transferred from other species. The family Cactaceae (Caryophyllales) does not have any species with their mitochondrial genome characterized, therefore, in this study, we assembled, annotated, and characterized the mitochondrial genome of the globose cactus *Mammillaria huitzilopochtli*, which has a length of 2,050,004 bp and a GC content of 42.9%. The genome characterization identified 34 protein coding genes, 28 tRNAs and three typical rRNAs. Five of the 65 genes annotated are copies of genes already present in the chloroplast. The comparison with other land plants and species of the Caryophyllales order revealed that the mitochondrial genome here characterized is one of the largest genomes of the order. The phylogenetic analysis showed concordant relationships among species of Caryophyllales when compared to other phylogenetic trees reconstructed with chloroplast loci. Finally, we propose the mitochondrial genome of land plants as an important source of variation that should be included in further studies involving evolutionary questions of land plant groups.

INTRODUCCIÓN GENERAL

Genoma mitocondrial: evolución, composición genética y variación estructural

Las plantas terrestres contienen en sus células tres fuentes distintas de información genómica: los cloroplastos (ADNcp), la mitocondria (ADNmt) y el núcleo (ADNn). Estos tres genomas difieren en su origen, funciones, longitud, composición genética y en su estructura o modo de organización (Saxena *et al.*, 2014). El ADNcp en las plantas terrestres es un genoma haploide con una longitud que va de 107 a 218 kpb, con un promedio de 120 genes que se encargan principalmente de la fotosíntesis (Röschenbleck *et al.*, 2016; Solórzano *et al.*, 2019). Por su parte, el ADNmt en plantas también es un genoma haploide pero tiene una gran variación en su tamaño, de 66 kpb (*Viscum scurruloideum*, (Skippington *et al.*, 2015) a 11.7 Mpb (*Larix sibirica* (Putintseva *et al.*, 2020), con un promedio de 60 genes que participan en los procesos de fosforilación oxidativa y transporte de electrones para la producción energética de la célula (Burger *et al.*, 2003; Dong *et al.*, 2020). Con respecto a su estructura, el genoma mitocondrial puede encontrar organizado en una sola molécula circular o en varias unidades llamadas cromosomas (Wu *et al.*, 2015). Finalmente, el ADNn puede ser diploide o poliploide, por lo tanto, la longitud de estos genomas es la más variable, desde 63 Mpb en la carnívora *Genlisea aurea* (Leushkin *et al.*, 2013), hasta 150 Gpb en *Paris japónica*, el más grande reportado hasta la fecha (Pellicer *et al.*, 2010).

El origen de la mitocondria está asociado a la aparición de las primeras células eucariotas. La hipótesis endosimbiótica de la mitocondria se sustenta con información filogenética y paleontológica, ésta, plantea que el origen de este orgánulo se remonta a ~2,000 millones de años, por medio de la fagocitación de una α-proteobacteria por parte de un arqueón (Knoop *et al.*, 2011; Mower *et al.*, 2012). Desde su origen y a lo largo de su historia evolutiva en las células eucariontes, sólo quedan remanentes del endosimbionte, con un genoma reducido que se mantiene en lo que hoy conocemos como mitocondria, en donde miles de genes no esenciales fueron perdidos y muchos otros fueron transferidos al genoma del hospedero (ADNn) (Lang *et al.*, 1999).

A diferencia de su contraparte en animales, el ADNmt de plantas terrestres es más extenso y complejo. Más allá de la diferencia en la variación de su longitud, el arreglo estructural suele ser muy diferente, ya que presenta un alto contenido de regiones repetidas sujetas a recombinación homóloga, así como una gran cantidad de espaciadores intergénicos no codificantes y regiones de ADN adquirido del núcleo, cloroplasto o lateralmente de otras especies (Ward *et al.*, 1981; Roger *et al.*, 2017). Sin embargo, las tasas

de mutación en las regiones codificantes son muy bajas, lo que implica que son altamente conservadas, explicado con ello, el importante papel fisiológico que cumple la mitocondria en la célula, en donde participa en roles relacionados con la respiración celular (Palmer *et al.*, 2000), el metabolismo (Araújo *et al.*, 2014), la muerte celular programada (Van Aken y Van Breusegem, 2015) y en la esterilidad citoplásmica del macho en ciertas especies (Hu *et al.*, 2014; Li *et al.*, 2018).

La variación estructural del ADNmt de angiospermas es altamente complejo y heterogéneo tanto en su longitud como en sus arreglos estructurales, esta variación ha sido reportada incluso en especies pertenecientes a un mismo género, tal es el caso de *Silene latifolia* (Caryophyllaceae) con un genoma constituido por una sola molécula circular de 253 kpb (Sloan *et al.*, 2010), *Silene noctiflora* con 7.3 Mpb organizados en 65 cromosomas (Wu *et al.*, 2015) o *Silene conica,* con 11.3 Mpb y 128 cromosomas (Sloan *et al.*, 2012) es el genoma mitocondrial más grande reportado a la fecha para un especie de angiosperma.

El ADNmt de angiospermas cuenta con un grupo básico (*core*) de genes que incluyen una unidad de maturasa R (matR), seis subunidades de la ATP sintasa, nueve subunidades del complejo mitocondrial I de NADH deshidrogenasa, citocromo b, citocromo c y las 3 subunidades del citocromo oxidasa. También tiene 17 genes variables, entre los que se encuentran las subunidades ribosomales y las subunidades de succinato deshidrogenasa, cuyos genes han migrado al núcleo de manera independiente en varios grupos de plantas (Adams *et al.*, 2002; Adams y Palmer, 2003). La transferencia de genes de la mitocondria al cloroplasto podría ser de origen reciente debido a la homología de las regiones para muchas de las especies en las que se ha documentado, sin embargo, la mayoría de estas regiones son pseudogenes no codificantes (Bensasson *et al.*, 2001). Asimismo, la transferencia de genes del cloroplasto a la mitocondria parece ser un proceso que se remonta a eventos ancestrales en muchos grupos de plantas y que sigue ocurriendo, siendo los genes del ARNs de transferencia los que más comúnmente migran y que parecen cumplir un rol activo en la síntesis de proteínas en los ribosomas mitocondriales (Warren *et al.*, 2021). La transferencia intergénica de mitocondria al núcleo y viceversa también es común y aporta sustancialmente a la extensión del tamaño del ADNmt, por ejemplo, el ADNmt del melón (*Cucumis melo*, Cucurbitaceae) tiene una longitud de 2.7 Mpb, en donde cerca del 47% es de origen nuclear (Rodríguez-Moreno *et al.*, 2011). Los procesos de migración intergénica no están del todo claros, pero se ha propuesto que están mediados por factores genéticos asociados a la reparación cromosomal y a la acción de agentes mutagénicos sobre la membrana mitocondrial (Blanchard y Schmidt, 1995)

Respecto a la transferencia horizontal de genes, el caso mejor caracterizado es el del género *Amborella* (Amborellaceae, Amborellales), en el cual en algunas especies se han identificado genes adquiridos por transferencia horizontal provenientes de otros grupos taxonómicos distantes como los musgos. De los 31 genes codificantes con los que cuenta, al menos 20, han adquirido copias adicionales que cuentan con total funcionalidad (Bergthorsson *et al.*, 2004). Adicionalmente, eventos de transferencia horizontal de genes han sido documentados en especies de plantas parásitas de la familia Rafflesiaceae y sus hospederos de la familia Vitaceae (Rice *et al.*, 2013). Esta transferencia no se limita a especies de plantas, puesto que también se ha reportado que ocurre con otros grupos como virus (Goremykin *et al.*, 2008), bacterias (Alverson *et al.*, 2011) y hongos (Rice *et al.*, 2013). Esta transferencia puede ocurrir por varios mecanismos, el más reportado es la facilitada por contacto directo entre plantas parásitas y sus hospederos (Won y Renner, 2003). Además se han propuesto otros mecanismos de transferencia, tales como la polinización, herbivoría, patógenos o simbiontes (bacterias, hongos y virus) o la adquisición de ADN directamente del suelo (Bergthorsson *et al.*, 2004; Richardson y Palmer, 2007).

El ADN mitocondrial en Caryophyllales y Cactaceae

El orden Caryophyllales es uno de los más diversos en las angiospermas al incluir cerca de 9,000 especies en sus diferentes familias (Brockington *et al.*, 2009). En este orden podemos encontrar plantas suculentas adaptadas a ambientes áridos como las de la familia de los cactus (Cactaceae), hierbas y arbustos de la familia Caryophyllaceae e incluso especies de plantas insectívoras pertenecientes a Nepenthaceae. Los tiempos de divergencia estimado de 71 millones de años (Brockington *et al.*, 2009), lo que sitúa al orden como uno de los linajes con mayor diversificación en angiospermas (Mering *et al.*, 2019).

La falta de genomas mitocondriales caracterizados para especies de este orden, ilustra lo poco que se ha estudiado al genoma mitocondrial, el cuál es regularmente descartado en estudios dedicados a responder cuestiones evolutivas en plantas (Rydin *et al.*, 2017). En el orden Caryophyllales, de acuerdo con la base de datos de GenBank (NCBI, 6 de junio, 2022), sólo se ha caracterizado el genoma mitocondrial para 22 especies, de ellas, 5 tienen su ADNmt seccionado en cromosomas, como *Fallopia multiflora* (Polygonaceae) con 2 cromosomas (Kim y Kim, 2018) y el ADNmt gigante de *Silene conica* (Caryophyllaceae), con 128 cromosomas (Sloan *et al.*, 2012). Estas gran variación nos permite entender la complejidad del genoma mitocondrial y la potencial razón de la falta de estudios existente

(Kozik *et al.*, 2019). Esta situación empeora cuando vemos el panorama para todas las angiospermas, para las cuales sólo se han caracterizado ~800 ADNmt para un total de 350,000 especies que se estiman para este grupo de plantas (Joppa *et al.*, 2011). La poca información disponible y el escaso conocimiento del ADNmt en plantas, aún no permite establecer las implicaciones de la variación estructural y el contenido de genes en la historia evolutiva de estos organismos, ni conocer en detalle los procesos evolutivos a los que ha estado sometido el genoma de este orgánulo (Xiong *et al.*, 2008; Xi *et al.*, 2013).

Dentro de las Caryophyllales, no encontramos ningún genoma mitocondrial descrito para la familia Cactaceae, icónico grupo de plantas, con ~1,500 especies descritas que se caracterizan por sus adaptaciones a ambientes áridos y semiáridos del continente americano (Hunt, 2006), y solo una especie distribuida fuera del continente americano, en África (*Rhipsalis baccifera*) (Cota-Sánchez y Bomfim-Patrício, 2010). De acuerdo con análisis filogenéticos moleculares calibrados, se estima que el origen de la familia Cactaceae se originó hace 32 millones de años, similar a otras familias del orden con adaptaciones a la aridez (Arakaki et al., 2011; Hernández-Hernández *et al.*, 2014). Esta estimación sugiere que los procesos evolutivos en Cactaceae, entre los que destaca la especiación por radiación adaptativa, han ocurrido en lapsos relativamente cortos en la familia (Arakaki *et al.*, 2011; Hernández-Hernández *et al.*, 2014).

El género con mayor riqueza de especies en la familia Cactaceae es *Mammillaria.* Con ~150 especies (Hunt, 2006), el cuál apareció aproximadamente hace solo 7.3 millones de años (Hernández-Hernández *et al.*, 2014). Una alta proporción de sus especies se distribuyen en México y previamente se ha propuesto a la región centro-norte del país como su centro de diversificación evolutiva (Hernández-Hernández *et al.*, 2014). Las relaciones filogenéticas de *Mammillaria* han resultado ser difíciles de elucidar y no han permitido resolver conflictos taxonómicos con otros géneros de cactus: *Coryphantha*, *Escobaria*, *Neolloydia*, *Ortegocactus* y *Pelecyphora.* Al igual que en todos los grupos de plantas, las relaciones filogenéticas del género se han reconstruido principalmente usando una cantidad pequeña de marcadores moleculares de cloroplasto, tal es el caso de lo reportado por Crozier (2005), quien usando 10 regiones de cloroplasto concluyó que sólo el subgénero *Mammillaria,* propuesto por Hunt (2006), se recupera como un grupo monofilético. Por su parte, Butterworth y Wallace (2004) no lograron separar a *Mammillaria* de otros géneros cercanos usando el intrón *rpl16* y el espaciador intergénico *psbA-trnH*. Resultado similares se han obtenido en otros estudios usando pocos marcadores moleculares que no llegan a ser

suficientemente informativos (Hernández-Hernández *et al.*, 2011; Bárcenas *et al.*, 2011; López-Ortiz, 2017).

En angiospermas, se han propuesto loci de cloroplasto como marcadores moleculares de alta variación (*trnK-matK*, *psbA-trnH*, *trnL-trnF*), sin embargo, no han resultado ser útiles para resolver filogenias en Cactaceae (Nyffeler, 2002). Actualmente, los avances tecnológicos en los métodos de secuenciación han permitido un acercamiento genómico al problema. Así, recientemente Chincoya y colaboradores (2020) compararon genomas completos de cloroplasto y lograron identificar 20 loci ortólogos como fuentes potenciales de alta variación molecular en el género *Mammillaria*. Además, la genómica comparada ha permitido la caracterización del genoma de cloroplasto en el género. Se encontró que en siete especies se identificaron tres estructuras distintas en su ADNcp que difieren en longitud y en composición de sus inversos repetidos, entre otros rearreglos genómicos (Solórzano et al., 2019). En contraste, para los cactus columnares *Carnegia gigantea* (Sanderson et al., 2015) y *Lophocereus schotii* (GenBank, KY886917) se reportó la ausencia de inversos repetidos. Estos resultados, implican la existencia de una alta variación estructural del ADNcp en Cactaceae.

En este trabajo, tenemos como objetivo ensamblar, caracterizar y comparar con otras plantas terrestres el genoma mitocondrial de *Mammillaria huitzilopochtli*. Esta caracterización es la primera que se realiza en la familia Cactaceae y nos permite explorar la utilidad potencial del ADNmt para aportar información sobre la historia evolutiva del género y la familia. Para la especie, se conoce la estructura de su genoma de cloroplasto (Solórzano *et al.*, 2019), la cual mantiene una estructura cuatripartita de 115,886 pb, con una región larga de copia única (71,997 pb), así como una región corta de copia única (29,401 pb) y los inversos repetidos (14,488 pb cada uno) con un total de 120 genes, el genoma de cloroplast de *M. huitzilopochtli* mantiene sintenia con el de *M. crucígera, M. solisioides* y *M. supertexta*. El genoma nuclear de *M. huitzilopochtli* es diploide y está estructurado en 11 pares de cromosomas. Por su parte, el contenido nuclear 2C que fue estimado por citometría de flujo es de 3.121 picogramos o 1C de 1.529 Mbp (del Ángel-Piña, 2005).

A partir de la información disponible para otros géneros de plantas, se espera que la mitocondria presente altos niveles de variación estructural respecto a otras especies de Caryophyllales y que la variación molecular sea suficiente para reconstruir las relaciones filogenéticas de las especies evaluadas.

1   TITLE PAGE

2   Title: Structural and gene composition variation of the complete mitochondrial genome

3   of *Mammillaria huitzilopochtli* (Cactaceae, Caryophyllales), revealed by *de novo*

4   assembly

5

6   Authors names: David Cruz Plancarte[1,2], Sofía Solórzano[1*]

7

8   Affiliations: [1*]Laboratorio de Ecología Molecular y Evolución, Unidad de Biotecnología y Prototipos

9   FES Iztacala, Universidad Nacional Autónoma de México, Avenida de los Barrios 1, Los Reyes

10   Iztacala, Tlalnepantla de Baz 54090, Mexico

11   [2]Posgrado en Ciencias Biológicas, UNAM, Circuito de Posgrados, Ciudad Universitaria, Coyoacán,

12   Ciudad de México 04510, Mexico.

13

14   *Corresponding author E-mail: solorzanols@unam.mx

15

ABSTRACT

Background: Structural descriptions of complete genomes have elucidated evolutionary processes in angiosperms. In Cactaceae (Caryophyllales) a high structural diversity of chloroplast genome has been identified among and within genera. In this study we assembled the first mitochondrial genome (mtDNA) for the short-globose cactus *Mammillaria huitzilopochtli*. For comparative issues we used the published genomes of 19 distinct angiosperms; and the gymnosperm *Cycas taitungensis* was the external group for phylogenetic issues.

Results: The mtDNA of *M. huitzilopochtli* was assembled in one linear chromosome of 2,052,004 bp, in which 65 genes were annotated. These genes add 57,606 bp and include 34 protein coding genes, 27 tRNAs and three rRNAs. In the non-coding sequences the repeats were abundant with a total number of 4,550 (179,215 bp). In addition, five complete genes (*psaC* and four tRNAs) of chloroplast origin were documented. For most of the protein coding genes was estimated a negative selection. The phylogenetic tree showed concordant topology with previous analysis based on chloroplast genome.

Conclusions: The type and number of genes contained in mtDNA of *M. huitzilopochtli* was similar to those reported for other 19 angiosperms irrespective of their phylogenetic relationships. Although other Caryophyllids exhibit strong differences in structural arrangement an in total size of mtDNA, these do not increase the typical number and types of genes to those identified for *M. huitzilopochtli*. We concluded that mtDNA in angiosperms increases its size by non-coding sequences, and not by a significant gain of coding genes.

Keywords: Cactaceae, *Mammillaria huitzilopochtli*; mitochondrial genome; Caryophyllids.

Background

In plants, mitochondria are essential organelles to provide cell energy through respiration [1, 2], besides other important roles in metabolism [3], such as, stress tolerance [4], programmed cell death [5], and in ~150 species, particularly in major crop species, such as, *Beta vulgaris*, *Capsicum annuum*, *Daucus carota* and *Zea mays*; were identified mitochondrial mutations that are associated to male sterility [6].

Recently, we searched in the NCBI site (April 20, 2022) complete organelle genomes for angiosperm taxa; and ~450 of mitochondria (mtDNA) and ~8000 of chloroplast (cpDNA) were documented. This unequal number of sequenced genomes has impacted on the relatively poor biological and evolutionary knowledge that exist for plant mtDNA. The genomic comparisons between mtDNA and cpDNA, indicate that the former shows larger size and stronger structural complexity than the latter [7]. Accordingly, it has been identified that mtDNA is organized either in a single or in various molecules, which are named as chromosomes, and these can be organized in linear or circular forms [8]. Currently, the factors and processes that drive the structural organization of plant mtDNA have not been entirely elucidated. The available data suggest that in flowering plants the number and length of mitochondrial chromosomes is not necessarily determined only by the total size of mtDNA. For example, the shortest mitochondrial genome (66 kbp) was documented in the parasitic mistletoe *Viscum scurruloideum* (Santalaceae), and it is organized in two chromosomes [9]. In contrast, larger mtDNAs that vary of 154 kbp (*Zelkova schneideriana*, Ulmaceae, MW717907) to 2 Mbp (*Corchorus capsularis*, Malvaceae, KT894204) are organized in a single chromosome. In addition, currently the largest mtDNA (11.3 Mbp) was documented in *Silene conica* (Caryophyllaceae), which exhibits a complex organization in the huge number of 128 circular chromosomes [10].

In spite of this strong variation in size and structural organization, the mtDNA of angiosperms contains a relatively low number of genes that varies of 28 in *Viscum scurruloideum* (Santalaceae) [9] to

61 69 in *Sesuvium portulacastrum* (Aizoaceae) [11]. In mtDNA of plants, the typical functional genes con-

62 tained are of the three main types: protein-coding genes, tRNAs and rRNAs. As occurs in other ge-

63 nomes, these genes are separated among them by noncoding DNA sequences named as intergenic

64 spacers [12]. It has been proposed that the relatively low number of genes contained in mtDNA was

65 caused by a large-scale gene migration from mitochondria to nuclear genome occurred along the evolu-

66 tionary history of plants [13]. In fact, most of the ~2,000 functional mitochondrial proteins currently

67 identified, are encoded in the nuclear genome; and only ~1% of them are encoded in DNA sequences

68 of mtDNA [1, 14]. In addition, gene transference between cytoplasmic genomes is also common; such

69 gene migration may include complete genes or fragments of non-coding sequences. Thus, this dynamic

70 intergenomic gene transfer was documented in many land plant taxa [15]. For example, the mtDNA of

71 the cantaloupe *Cucumis melo* (Cucurbitaceae) has a total size of 2.7 Mbp, and nearly 46.77% and

72 1.41% were identified from nuclear and plastidic origin, respectively [16]. Consequently, the interge-

73 nomic gene transfer increases may promote the total size of mtDNA [15, 17]. In addition, in the

74 mtDNA of angiosperms also has been documented horizontal gene transfer among diverse taxonomic

75 groups, such as, virus [18], bacteria [19], fungi [20]; as well as, in higher vascular plants [21, 22].In

76 addition, mtDNA contains abundant repeated DNA sequences (repeats) that are located predominantly

77 at the non-coding sequences properly recognized as intergenic spacers. These abundant repeats also

78 cause considerable increments in the total size of mtDNA [23]. Moreover, it has been suggested that

79 these repeats may be involved in the homologous recombination and regulation of the complete replica-

80 tion of mtDNA [7].

81   Currently, the mutation process and its underlying factors have not been evaluated in detail for

82 plants. However, preliminary comparisons of distinct mitochondrial coding genes showed lower muta-

83 tion rates than those recorded for plastidic (3X) and nuclear (16X) genomes [24, 25]. Thus, in coding

84 sequences of these mtDNA´s the mutations are restricted, and consequently these low levels of molecu-

lar variation do no represent an adequate source for phylogenetic studies [26]. It is important to mention that the abundant and large continuous sequences of noncoding regions (i.e., introns and intergenic spacers) have not been explored as potential sources of molecular variation. Thus, the non-coding regions of mtDNAs would be sources of molecular markers to resolve many types of biological issues. Lastly, probably in the mtDNA of plants the evolutionary story was impressed, and it may help to elucidate the enigmatic and the not fully resolved evolutionary history of angiosperms.

Presently, for angiosperms most of the phylogenetic studies have been carried out with plastidic loci (e.g. [27], [28]), however, this genome did not effectively work in whole flowering groups; as is the case of cacti species. The nearly 1,500 members of Cactaceae [29] are recognized as a monophyletic group [30], however, their internal phylogenetic relationships have not been fully resolved (e.g. [31, 32]). In this study we sequenced and assembled de novo the mitochondrial genome for *Mammillaria huitzilopochtli* D. R. Hunt. (Cactaceae, Caryophyllales). Recently, the whole cpDNA of this short-globose cactus *M. huitzilopochtli* was described [33], and its relative plastidic molecular variation was assessed in comparison to other *Mammillaria* species [34]. The objectives here studied were, 1) to describe the structural organization of the whole mitochondrial genome in this cactus, 2) to estimate the mutation rates of coding regions among 21 species; 3) to compare our results to those mtDNA reported for other 20 land plants, emphasizing Caryophyllids.

Results

Characterization of the mitochondrial genome of *Mammillaria huitzilopochtli*

The newly assembled mitochondrial genome of *M. huitzilopochtli* had 2.052 Mbp in total size, and it is organized in a single linear molecule. This mtDNA had a higher proportion of A′s (28.6%) and T′s (28.4%), followed by G′s and C′s (each one, 21.5%). This genome was composed of 12 different types of genes, 10 of them corresponded to different types of protein coding genes (Fig. 1).
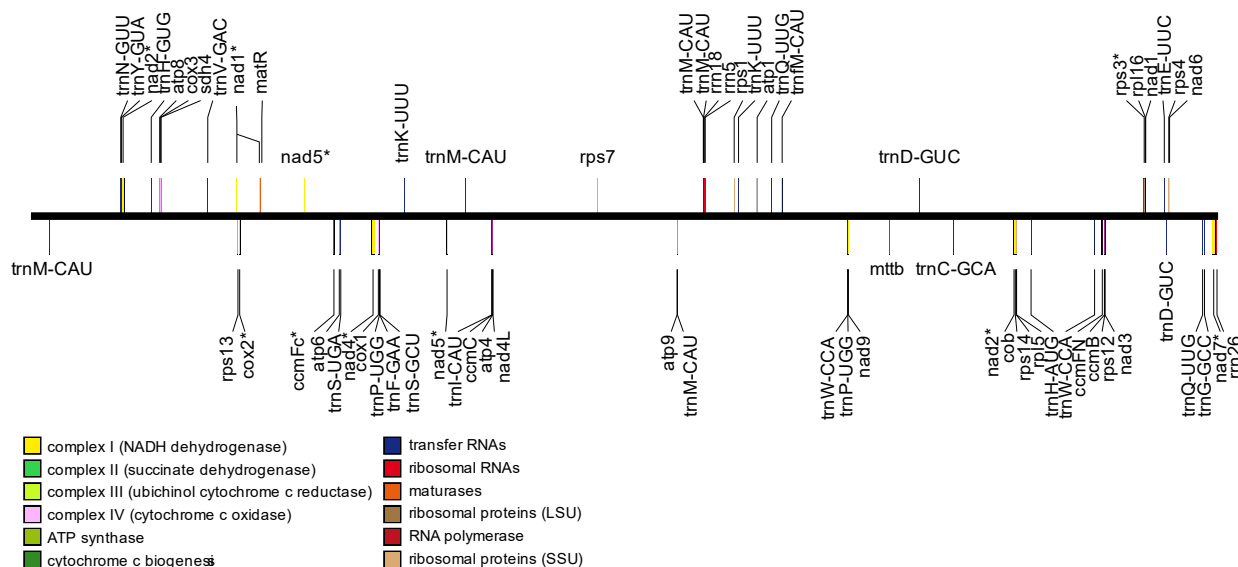
Fig. 1 Gene composition and total size of the mitochondrial genome of *Mammillaria huitzilopochtli*. The color of the square and the label indicate the type of the protein-coding gene, excepting those indicated for tRNA and rRNAs.

In the mtDNA of *M. huitzilopochtli*, a total of 65 genes were annotated, six of them had of one to four additional copies (Table 1). Of these 65 genes, 34 (33 of mitochondrial origin and the gene *psaC* was of plastidic origin) corresponded to some protein coding genes (PCGs). These PCGs were of ten different families of genes, eight of them with introns included (Table 1). In addition, a total of 28 subunits of tRNAs, four of them were of plastidic origin; and three subunits of rRNAs were documented (Fig. 1). These 65 genes represented only 2.8% (57,606 bp) of the DNA sequence of the total genome size, consequently, a 97.2% of DNA sequences corresponded to non-coding sequences of the types of intergenic spacers (Fig. 1).

Table 1 (included at the end of the document)

Of the total of 33 mitochondrial PCGs, 29 of them (87.8%) had the typical ATG start codon. In addition, four PCGs had other alternative codons: ACG (*nad1*), TTG (*rps4*), ATA (*mttb*) and GTG (*rpl16*). With respect to stop codons, TAA was recorded for 13 PCGs, and codon TGA was identified for other 13 different PCGs, and six genes had TAG as stop codons. The gene *atp9* had the alternative stop codon CGA. In addition, in eight of the 33 PCGs were identified of one to five introns (Table 1).

126 Of these, *nad7* had four introns; *nad2* had three, *nad4* and *nad5* had two; and for *ccmFc*, *cox2*, *nad1*

127 and *rps3* one intron was identified (Table 1). These introns varied in length of 838 bp (*nad5*) to 2,350

128 bp (*nad2*); three of these introns were trans-spliced (*nad1*, *nad2* and *nad5*), and five (*ccmFc, cox2,*

129 *nad4*, *nad7* and *rps3*) were cis-spliced.

130   With respect to repeated sequences of microsatellite type, a total of 1,219 were recorded along the

131 mtDNA of *M. huitzilopochtli*. These microsatellites were composed by one type of nucleotide (396

132 repeats), dinucleotide (462), trinucleotide (59), tetranucleotide (170); and 109 microsatellites showed a

133 composed motif (i.e., two types of repeated motifs separated by a non-microsatellite sequence). Lastly,

134 only a total of 23 complex microsatellites were composed by five to six nucleotides, and these were

135 distributed along the non-coding IGS (Table 2). Twenty of these microsatellites were composed by

136 three repeats, and these microsatellites were abundant on the IGS of *trnD-GUC – cox2* (5 repeats); and

137 *nad1 - rps3* (4 repeats) (Table 2).

138 **Table 2** Distribution and location of the microsatellites composed by five to six nucleotides. The coordinates of start and
139 end of the microsatellite sequences refer to the assembled mitochondrial genome of *Mammillaria huitzilopochtli*

Number	(Motif) number of repeats	Start	End	Location
1	(AAGAGT)3	34,066	34,084	*trnQ^{UUG} - trnD^{GUC}*
2	(TGAAA)3	246,455	246,470	*sdh4 - trnV^{GAC}*
3	(CGAAGG)5	446,131	446,161	*matR - trnC^{GCA}*
4	(AAAGA)3	450,464	450,479	*matR - trnC^{GCA}*
5	(CCGGG)3	596,633	596,648	*mttb - nad9*
6	(AGTGA)3	978,493	978,508	*atp9 - rps7*
7	(CTCGG)3	1,037,692	1,037,707	*rps7 - rrn18*
8	(CCTTCG)3	1,047,066	1,047,084	*rps7 - rrn18*
9	(TTCCT)3	1,200,521	1,200,536	*rrn5 - rps1*
10	(TCTTG)3	1,431,830	1,431,845	*nad5 - trn^{GCU}*
11	(ATATAT)4	1,477,325	1,477,349	*nad4 - trnS^{UGA}*
12	(GCCTA)3	1,604,532	1,604,547	*trnD^{GUC} - cox2*
13	(AAGGCG)3	1,645,312	1,645,330	*trnD^{GUC} - cox2*
14	(TTTTC)3	1,649,583	1,649,598	*trnD^{GUC} - cox2*
15	(GCCCA)3	1,671,137	1,671,152	*trnD^{GUC} - cox2*
16	(ACTTTC)3	1,678,787	1,678,805	*trnD^{GUC} - cox2*
17	(CTTTT)3	1,787,375	1,787,390	*nad1 - rps3*
18	(GAGAG)3	1,801,389	1,801,404	*nad1 - rps3*

19	(AAAAG)3	1,805,130	1,805,145	*nad1 - rps3*
20	(CAACT)3	1,847,144	1,847,159	*nad1 - rps3*
21	(CTAAA)3	1,963,196	1,963,211	*trnE^{UUC} - rps4*
22	(TTTCA)3	1,997,401	1,997,416	*rps4*-3'
23	(TAGAA)4	2,007,522	2,007,542	*rps4*-3'

On the other hand, in the mtDNA of *M. huitzilopochtli* were abundant direct and inverted repeated DNA sequences, which were distributed along the whole genome (Fig. 2). In total 4,550 of these repeats were documented, which represent 8.73% (179,215 bp) of the total length of the genome. The most abundant repeats were those shortest: 20-39 bp (2,470 repeats); followed by those of 30-59 bp (1,878), 60-199 bp (183), 100-199 bp (44); and lastly only 17 repeats >200 bp were identified (Fig. 2). We did not observe strong differences in the number of repeats between direct and inverted direction.
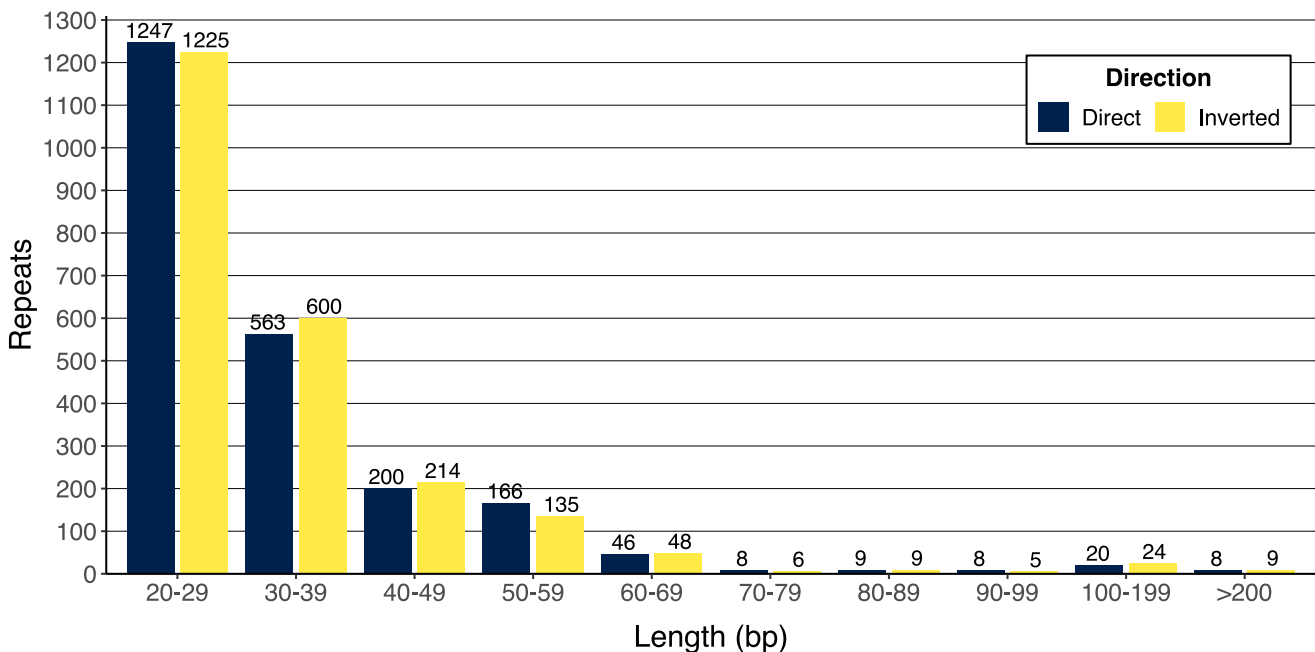
Fig. 2 Length and direction of repeated DNA sequences documented in the mitochondrial genome of *Mammillaria huitzilopochtli.*

In mtDNA of *M. huitzilopochtli*, a total of 34 DNA sequences of plastidic origin that added 10,184 bp (Table 3) were identified. These sequences were complete and fragmented coding genes, as well as, fragments of noncoding sequences located in the IGS of the chloroplast genome. The complete copies of genes were the coding gene *psaC* (including its start-stop codons), and three tRNAs, these were

154 *trnD-GUC* (this had two copies), and a copy of *trnN-GUU* and *trnI-CAU*. The other 31 sequences of

155 plastidic origin were fragments of genes and IGS reported previously in the chloroplast genome (Table

156 3).

157 **Table 3** Genes, intergenic spacers (IGS) and introns of plastid origin recorded in the mitochondrial DNA of *Mammillaria*
158 *huitzilopochtli.* The length, percentage of identity and coordinates obtained by comparison between genomes of mitochon-
159 dria (this study), and chloroplast (MN517612). The percentage of identity, the number of mismatches and of gap opens
160 between these two genomes

	Length (pb)	Identity (%)	Mismatches	Gap opens	Mitochondrial		Chloroplast		Gene/IGS/intron
					start	end	start	end	
1	976	100	0	0	1257330	1258305	92928	93903	*rps7* - Ψ*ndhb*
2	890	99.44	5	0	845158	846047	69624	70513	*ycf2*
3	848	87.38	76	11	1916667	1917495	30751	31586	*psaA*
4	742	90.43	71	8	1921118	1921854	10763	11501	*psbC*
5	731	99.32	5	0	305474	306204	90124	90854	*rrn16*
6	581	100	0	0	1567367	1567947	106298	106878	*atpA* - *atpF*
7	457	100	0	0	2,029455	2029911	3724	4180	*rpoB*
8	441	100	0	0	1326125	1326565	22939	23379	*rbcL*
9	421	99.29	0	0	285799	286219	46334	46754	*psbB*
10	408	91.67	34	3	1105343	1105750	65971	66375	*ycf2*
11	373	100	0	0	3	375	107146	107518	*atpF* (intron)
12	368	100	0	0	1738587	1738954	61284	61651	*matK*
13	301	98.01	6	0	97235	97535	92606	92906	*rps7*
14	264	98.86	2	1	696708	696972	78457	78719	*psaC* *
15	257	100	0	0	1417528	1417784	88364	88620	Ψ*ycf68*
16	226	96.02	8	1	839999	840224	94333	94557	Ψ*ndh* – Ψ*ycf2*
17	168	81.55	22	9	1105085	1105252	65812	65970	*ycf2*
18	164	98.17	3	0	756462	756625	9625	9788	*psbD*
19	160	100	0	0	306775	306934	88800	88959	Ψ*trnI^{GAU}* - *rrn16*
20	153	100	0	0	110668	110820	88023	88175	Ψ*trnA* - Ψ*ycf68*
21	140	95	7	0	1231238	1231377	78948	79087	*ndhD*
22	123	91.06	9	2	1920978	1921099	5950	6071	*trnC^{GCA}* - *petN*
23	122	100	0	0	1929365	1929486	85922	86043	*rrn23*
24	110	84.55	17	1	1388921	1389021	46485	46594	*psbB*
25	100	94	6	0	1078249	1078348	10121	10220	*psbD*
26	83	96.39	2	1	1963453	1963536	8627	8708	*trnD^{GUC}* *
27	82	98.78	1	0	1535734	1535816	8627	8708	*trnD^{GUC}* *
28	81	100	0	0	1211781	1211861	65021	65101	*ycf2*
29	81	93.83	5	0	264614	264694	105467	105547	*atpA*
30	78	94.87	4	0	143568	143645	4218	4295	*rpoB*

| 31 | 77 | 100 | 0 | 0 | 796513 | 796589 | 94862 | 94938 | *trnI*^{CAU}* |
32	77	96.1	3	0	155219	155295	83026	83102	*trnN*^{GUU}*
33	53	100	0	0	1921853	1921905	85887	85939	*rrn23*
34	48	93.75	3	0	490750	490797	65638	65685	*ycf2*

*Complete genes identified in the mitochondrial genome, Ψ indicates a pseudogene that was reported as such in chloroplast genome

Comparison of mitochondrial DNA of *Mammillaria huitzilopochtli* to other land plants.

The phylogenetic tree grouped the 16 Caryophyllids in a single monophyletic ingroup (Fig. 3) support-ed with 100% of bootstrap.
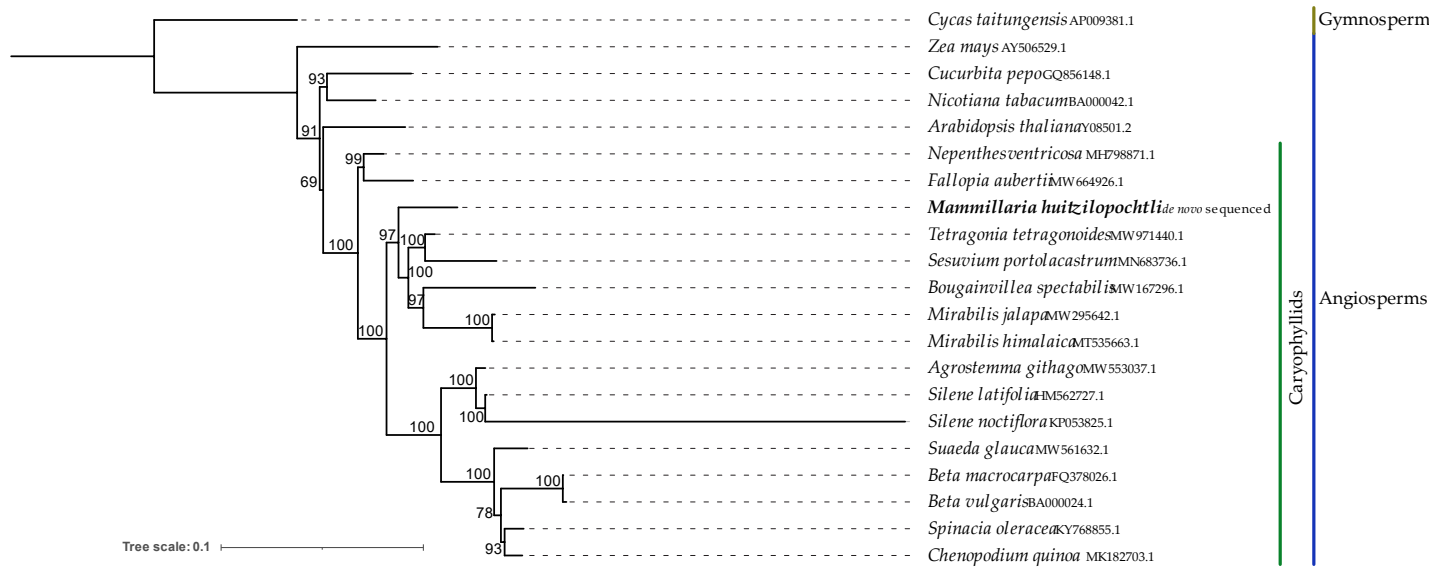
Fig. 3 Maximum Likelihood phylogenetic tree based on 29 orthologous loci. The numbers correspond to the bootstrap per-centages

The mitochondrial genome of *M. huitzilopochtli* had a GC content of 42.97%, which is similar to that of other 15 Caryophyllid species (Fig. 4). The 16 Caryophyllids had on average 43.77± 0.99SD of GC content. In the 21 plant studied species, GC content was negatively correlated to the total length of mitochondrial genome (r=-0.68, p=0.00073), however, this correlation was not significant, when the atypical largest value of S. *noctiflora* was excluded. The lowest GC content was documented in the two species of Caryophyllaceae: *S. latifolia* (42.56%) and *S. noctiflora* (40.82%), which had a genome of 235 kbp and 7.1 Mbp, respectively. In the 21 studied species, the mtDNA of *M. huitzilopochtli*

(2,052,004 bp) was the second largest genome (Fig. 4), behind *Silene noctiflora*. Lastly, the average

number of genes in the 21 species was 59±6.34SD. The total number of genes contained in the mtDNA

of these analyzed species was not correlated to the total length of the mtDNA (N=21, r=-0.14, p= 0.56).

In fact, the lowest number of genes was documented in the Caryophyllid *S. latifolia* (41 genes); and the

highest number in the gymnosperm *C. taitungensis* (70).
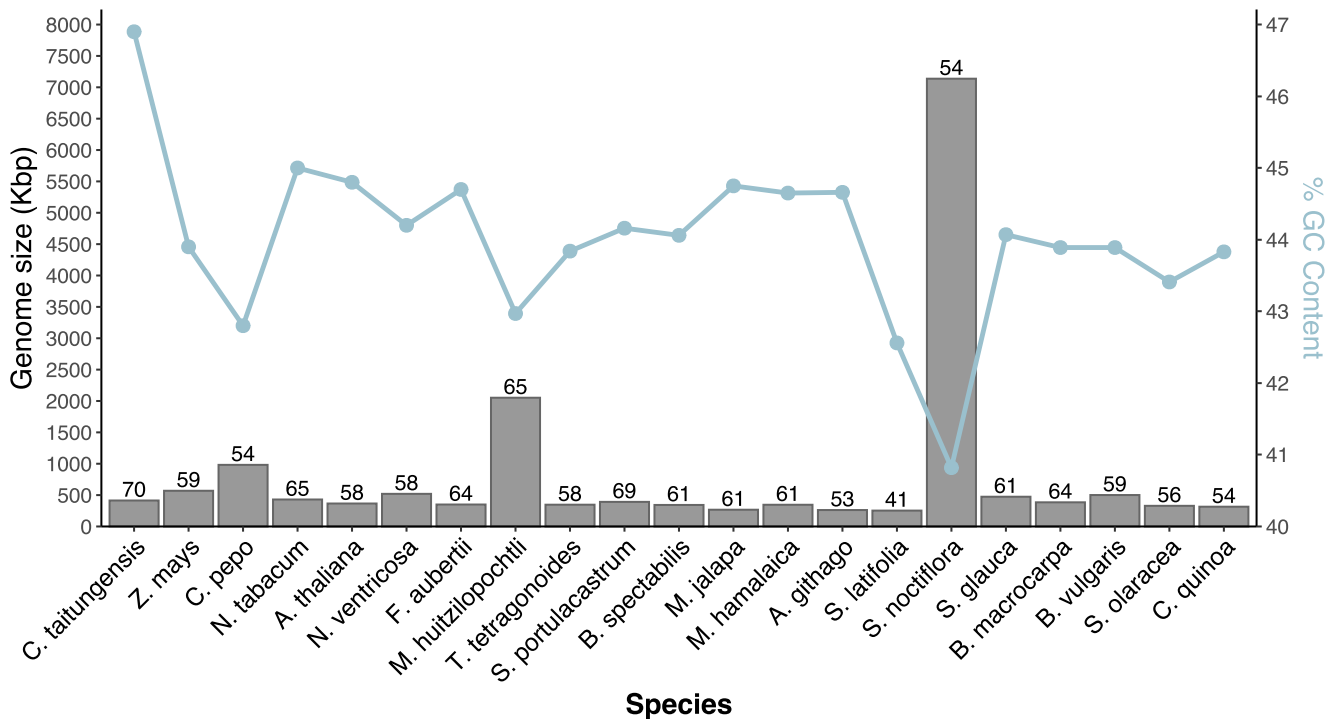
Fig. 4 Comparison of the genome size (bars) and GC content (line) of *Mammillaria huitzilopochtli* to other 20 land plants.
The number above the bar indicates the total number of genes of each genome

      With respect to the identity of the genes that composed the mitochondrial genomes, we document-

ed that the 21 species had the three typical ribosomal units (*rrn5, rrn18* and *rrn26*) reported to land

plants. In contrast, the number and identity of PCGs showed variation among the 21 species. The gym-

nosperm contained the largest number of PCGs (41 genes), however, most of the 20 angiosperms had a

complete set of 24 PCGs, which are those recognized as core genes for angiosperms; however, some

PCGs lacked (white squares, Fig. 5); or these were incomplete sequences (pseudogenes; grey squares

Fig. 5). In the angiosperms *M. jalapa* lacked the genes *cob* and *cox1*; and in *S. glauca*, were not identi-

fied the genes *nad4* and *nda6*. In contrast, more variation was documented in the 17 genes named as variable PCGs (those non-core genes). Particularly, we documented complete lacks and pseudogenization in the subunits of ribosomal proteins (*rps*) and of succinate dehydrogenase (*sdh*). The cactus *M. huitzilopochtli* lacks eight of these genes; and a total of 24 pseudogenes were documented among 12 different species. With respect to tRNAs the most frequent lacks were documented in *trnL-UAA* (20 species), *trnR-UCU* (20), trnV-*UAC* (20), *trnI-GAU* (19) and *trnL-CAA* (17) (Fig. 5). Pseudogenization in tRNAs was documented in four tRNAs but only in two species (*A. thaliana* and *S. noctiflora*). The species *S. noctiflora* and *S. latifolia*, had the higher number of pseudogenes, 6 and 5, respectively; and in the cactus *M. huitzilopochtli* had only one pseudogene (Ψ*rps14*; Fig. 5).
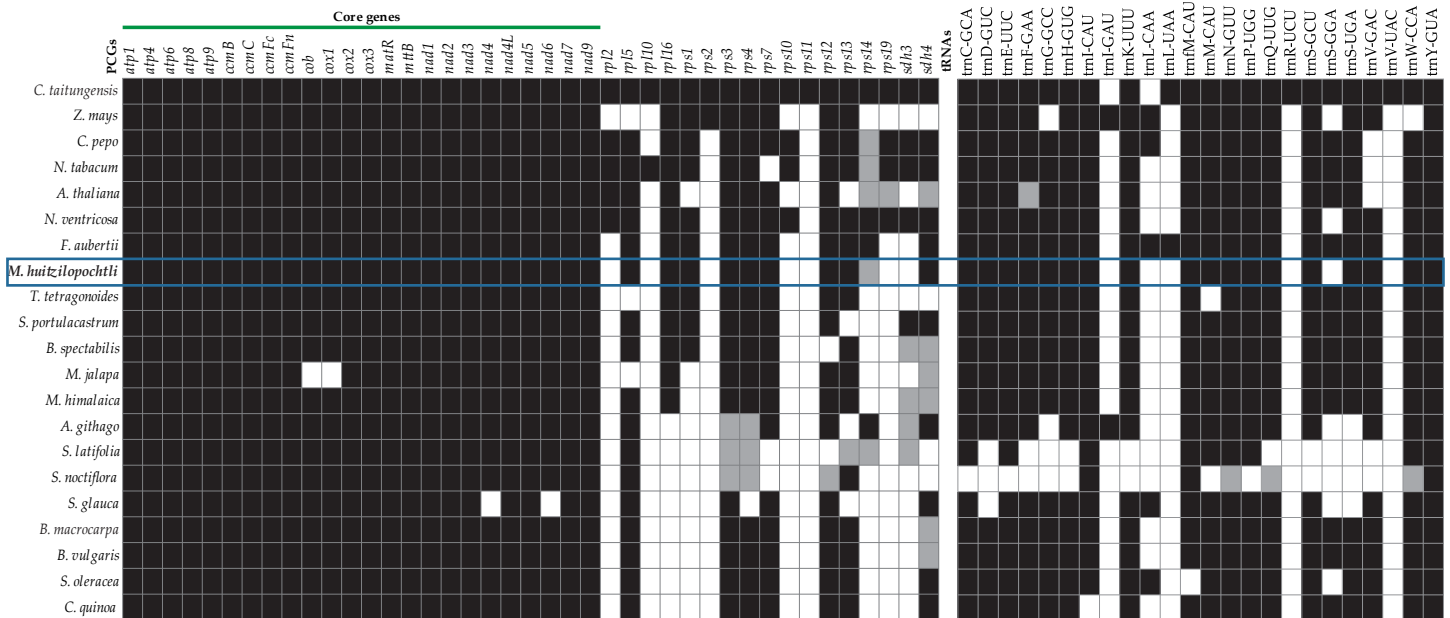
Fig. 5 Comparison of gene content of protein coding genes and tRNAs of mitochondrial DNA of *Mammillaria huitzilopochtli* to other 20 land plant species. The color of the squares indicates if the gene was recorded (dark), absent (white), and grey (pseudogene)

The comparison of substitution rates in 25 genes between *M. huitzilopochtli* to other six angiosperm species (Fig. 6), showed that 23 genes had values that indicated negative selection (Ka/Ks<1, below red horizontal line, Fig. 6). Positive selection (Ka/Ks>1) was estimated only in the comparison

of the gene *atp6* of *C. quinoa* and in *ccmB* of *A. thaliana* and *N. tabacum* (Fig. 6). Neutral selection was not identified.
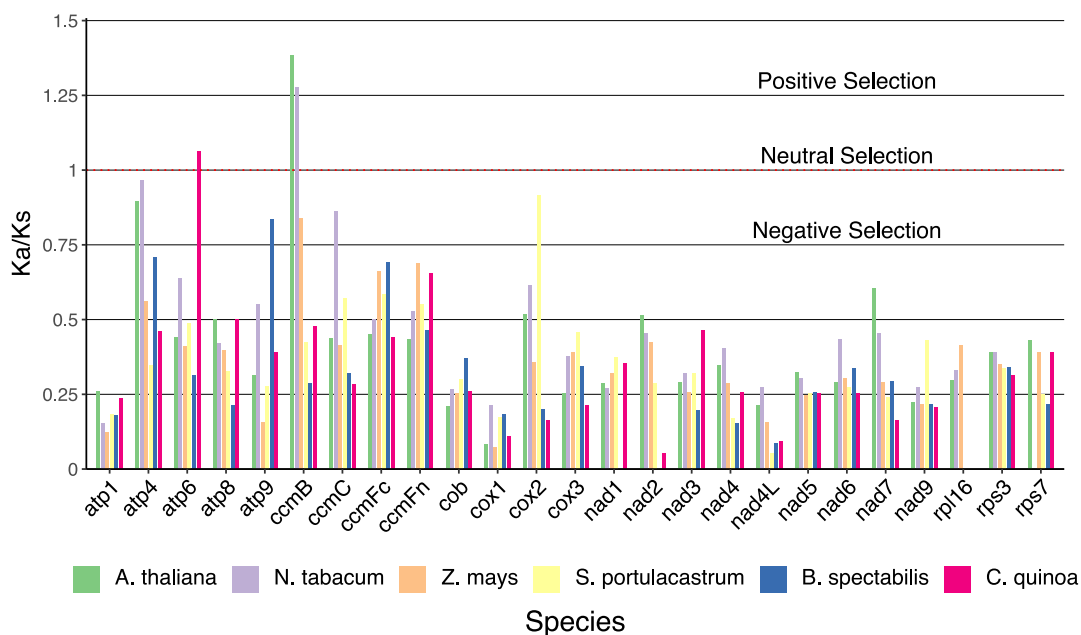
Fig. 6 The values of Ka/Ks of 25 protein coding genes compared between *Mammillaria huitzilopochtli* to six angiosperm species

Discussion

This study pioneered the analysis of complete mitochondrial genome on cacti species; and we consider these results will open new perspectives for phylogenetic analysis of these plants. However, the lack of data limited our comparisons to other land plants not phylogenetically related, however, the comparisons focused in Caryophyllids (Amaranthaceae, Aizoaceae, Caryophyllaceae, Nepenthaceae, Nyctaginaceae and Polygonaceae) showed similar gene content, although the strong differences in size and structural arrangement. Our findings showed that *M. huitzilopochtli* possess the third largest mitochondrial genome (2.05 Mbp) followed to the Caryophyllids *S. conica* (11.3 Mbp) [10] and *S. noctiflora* (7.1 Mbp) [35]. Our comparisons among 21 species suggested that the total size of the genome does not determine: 1) the structural complexity (i.e., arrangement in more than one chromosome), 2) GC content; and 3) total number of genes.

We identified that the variation in the total size of mtDNA of the 21 studied species, was caused by the expansion and contraction of noncoding sequences, primarily by the length of IGS, and secondly of introns. Consequently, in these land plants their total size of mtDNA expands/contracts by length of noncoding sequences, not by gain coding sequences. In addition, we identified that the enlargement of IGS was associated to the abundant repeated sequences of different nature, such as, microsatellites, direct and inverted repeats. The abundant repeats in IGS of mtDNA of land plants is a common feature [19, 36, 37], and some studies [16, 37] have suggested that IGS may receive DNA sequences from foreign genomes. Presently, the functional role of these repeats sequences in the mtDNA has not been clearly elucidated, however, it has been postulated that these repeats may participate in the replication of complete mtDNA [23], and in the repeat-mediated recombination [38, 39]. This last process has been proposed to have an important role in the structural rearrangements of mtDNA [7, 39, 40].

Moreover, our results indicated that mitochondrial genome of land plants tends to maintain a gene composition and gene number, irrespective of the total size, or the structural complexity in which a certain genome is arranged. Particularly the four ribosomal units, and the set of 24 PCGs seem to have a key role in the 21 studied species. It is probable that the identity and the number of genes contained in a certain mtDNA are influenced by phylogenetic factors, and not by its structural features. Since, we documented that the gymnosperm *C. taitungensis* had the highest number of different genes, which is similar to the results published for other two conifers *Larix sibirica* (77 genes, [41]), and *Picea sitchensis* (71, [42]). Our results showed that the 20 studied angiosperms had a lower total number of genes than the gymnosperm *C. taitungensis*; and we identified that it was caused by the loss of various types of PCGs and tRNAs. It seems that in angiosperms the evolutionary trend is a reduction in the total number of genes with respect to gymnosperms, however, this should be verified when more data of complete mitochondrial genomes are available for flowering plants. On the other hand, we concluded that in *M. huitzilopochtli* the natural selection restricts mutations in the coding genes, since most of

these PCGs showed Ka/Ks<1 (negative selection). Consequently, these sequences are highly conserved as has been recognized for most angiosperm species (e.g., [2, 12]).

We documented migration of DNA sequences of plastidic origin to the mtDNA of the cactus *M. huitzilopochtli*. This migration consisted in complete coding genes, and fragmented sequences of genes and IGS. In fact, the migration of DNA sequences from chloroplast to mitochondria is a common process that has been documented for other angiosperms in previous studies [19, 37, 43]. However, the migration of complete coding genes from chloroplast to mtDNA is an unusual process reported in angiosperms [44], as well as, in gymnosperms [45]. In this study, we documented for *M. huitzilopochtli* the presence of the complete coding gene *psaC* in the mtDNA, which was previously reported in the plastid genome of this species [33]. Presently, it has not been identified if these copies of plastidic origin have a functional role in the mtDNA [17, 44]. In addition, the migration of tRNAs from chloroplasts to mitochondria is common in land plants [43]. In this study, four tRNAs were documented in the *M. huitzilopochtli*, which were previously reported in the chloroplast genome of this cactus [33]. The functional role of these tRNAs in mtDNA was assigned to the synthesis of proteins in the mtDNA [43]. On the other hand, in plants the migration from nuclear genome to mtDNA has not been studied in detail, however, it may occur. Since, in the whole sequenced genome of the cantaloupe, *Cucumis melo* (Cucurbitaceae) was reported that nearly 46.47% of its mtDNA was of nuclear origin [16]. In our study we did not evaluate sequences of nuclear origin because currently it has not been published a complete nuclear genome for *M. huitzilopochtli*.

We have to mention, that although the primary purpose of this study was not to obtain the phylogenetic relationships of *M. huitzilopochtli* to other Caryophyllids, due to few available data of complete mtDNA; the phylogenetic tree obtained recovered a concordant topology to that obtained in previous studies based on plastidic loci [46]. Accordingly, the seven families of Caryophyllales here studied were organized according to the previous published phylogeny of 40 families of this order based on 83 plastidic loci [47]. Moreover, the 16 Caryophyllid species here studied were clustered in a monophylet-

22

ic ingroup. Based on these phylogenetic results, we concluded that the 29 mitochondrial loci here used had enough phylogenetic resolution to separate families of the order Caryophyllales. We expect that in the future, as more complete mitochondrial genomes are being published, the value of mtDNA for phylogenetic analysis will be reevaluated. For example, recently, the study of Rydin et al. [26] analyzed 53 species of Rubiaceae (Gentianales); based on mitochondrial and chloroplast genomes. The phylogenetic trees showed phylogenetic discordances, which suggested that the phylogenetic studies should include mitochondrial, nuclear and plastid loci in order to study in detail the evolution of plants.

Conclusions

This newly assembled and annotated complete mitochondrial genome of the cactus *M. huitzilopochtli* has findings that will allow for further comparisons of mtDNA of members of Cactaceae. We expect that our study contributes to elucidate biological, taxonomic, and systematic issues still extant in Cactaceae. In the context of angiosperms, we consider that presently, we are far from understanding the processes that drive the structural organization of mtDNA of plants. It seems that the low mutation rates documented in coding regions of mtDNA in this cactus are driven by natural selection, which allows synonymous substitutions in DNA sequences without effects on the amino acids chains. Thus, we encourage the sequencing of complete mitochondrial genomes in order to assemble the evolutionary puzzle of plants.

Methods

Genomic DNA extraction and massive sequencing

Tissue samples of *Mammillaria huitzilopochtli* D.R. Hunt were collected in 2016 from a wild population close to the municipality of San Juan Bautista Cuicatlán, Oaxaca. These tissue samples were im-

299 mediately stored in liquid nitrogen until its experimental processing in the laboratory, where tissue

300 samples are maintained at -80°C to long term for genetic research.

301 Frozen tissue samples of 70-100 mg from single individual of *Mammillaria huitzilopochtli* were inde-

302 pendently processed following the indications of manufacturer of DNAeasy Plant Mini Kit (Qiagen,

303 Germany), in order to obtain one microgram of gDNA of high molecular weight and 260/280 ≥1.7.

304 This whole gDNA was supplied to sequencer provider, who prepared PE libraries with a mean size of

305 ~600 bp of insert and sequenced in 2 x 150 cycles in TruSeq Nano DNA 350 (Illumina, USA).

306 **Mitochondrial genome assembly and annotation**

307 The quality of the reads of raw data was checked with FastQC v0.11.9 [48]. Since 91.66% the reads

308 had Qphred ≥30, and attached adapters were not identified, these reads did not require filtering. Since

309 in this set of reads the three genomes were represented, we proceeded to extract only those reads of

310 mitochondrial origin. For this, the reads of plastid origin were mapped with BWA-0.7.17 [49], taking

311 as reference the cpDNA published for *M. huitzilopochtli* [33]. These plastid reads were discarded with

312 SAMtools 1.15 [50]. The remaining reads were *de novo* assembled with NovoPlasty 4.3 [51]. This as-

313 sembly obtained various and large supercontigs (~10 - 290 kbp), which did not compose a single con-

314 tinuous sequence. The plant mitochondrial origin of the reads included in these supercontigs was veri-

315 fied with BLASTN [52]. All of these mitochondrial reads verified were directly extracted from raw

316 data; and were newly assembled with the pipeline Unicycler v.0.4.9 [53] that uses SPAdes 3.15 [54] as

317 assembler. This assembler formed various and independent large supercontigs of ~300 kbp, which were

318 input in the program Bandage v0.8.1 [55] in order to identify their adjacency. The original reads of

319 these flanking sequences were searched in the raw data with BBDuk [56], in this way we achieved to

320 merge all supercontigs. Finally, since some few and short gaps remained in some of the supercontigs,

321 we searched in the original reads of raw data those sequences in order to fill by hand these gaps. Once

322 the genome was completely assembled, it was fully annotated with Mitofy [17]; and all the genes iden-

tified were manually curated with BLASTN [52] The complete mitochondrial genome assembled, annotated, and manually curated of *M. huitzilopochtli* was plotted with OGDRAW [57]. The mtDNAMhui was featured on the total size, number of chromosomes, gene composition based on three types of genes: protein coding genes (PCG) that were classified in families according to its functional role; tRNAs and rRNAs. For each PCG, its length, start and stop codons, and the length of the amino acid chain transcribed were described. In addition, we characterized the abundant and diverse types of repeated sequences with MISA-web [58]. Those repeats of microsatellite type (i.e. DNA sequences repeated in tandem) were assessed. We characterized the microsatellite as simple (i.e. an unique type of motif) and composed (i.e. two or more types of motifs) microsatellites, with a number of repeats of three to eight. In addition, the direct and inverted repeats of at least 20 bp also were assessed with REPuter [59]. Lastly, we searched DNA sequences of plastid origin by comparing mtDNAMhui to the cpDNA accessed at NCBI (MN517612) previously reported [33]. This comparison was carried out with BLASTN [52] based on parameters of matching rate ≥ 70%, E-value ≤ 1e − 10, and length ≥ 40.

Comparison of mitochondrial genome of *Mammillaria huitzilopochtli* to other land plant species

We compared gene composition and structural characteristics between mtDNAMhui to the mtDNA previously reported for other 15 Caryophyllids, and other four angiosperms *Arabidopsis thaliana*, *Cucurbita pepo*, *Nicotiana tabacum* and *Zea mays* with detailed mtDNA published. The gymnosperm *Cycas taitungensis* was used as external group in phylogenetic analysis (species evaluated listed in Online Resource 1). The phylogenetic tree was obtained for these 21 studied species, and it was based on 29 orthologous loci (26,849 bp) identified with OrthoFinder 2.5.4 [60]. The DNA sequences of these loci were both coding sequences as noncoding sequences IGS. The DNA sequences of these loci were concatenated and aligned with MAFFT 7.471 [61]. The best substitution model that the program ModelFinder [62] identified was IVM; and Maximum Likelihood ran with 1000 bootstraps in IQ-TREE 1.6.12 [63] were used to obtain this tree. We used this phylogenetic tree to organize the taxa

order in the comparisons carried out. We compared among the 21 species the percentage of GC content, the total size, and number and identity of genes. Since in mtDNA of land plants there is a group of genes recognized as core genes, which includes PCGs (e.g. [2, 13] and tRNAs, we described in detail the number and type of these genes for each of the 21 studied species, as well as, if these were pseudogenes or were completely lack in the respective mtDNA. We tested the statistical relationship between GC content to the total length of the 21 genomes analyzed with Pearson correlation following the procedure described in Sokal and Rohlf [64]. In order to evaluate the natural selection on the DNA sequences of PCGs we estimated the rate of synonymous (Ks) and nonsynonymous (Ka) substitutions of 25 PCGs between mtDNAMhui and other six angiosperm species (*A. thaliana*, *Bougainvillea spectabilis*, *Chenopodium quinoa*, *N. tabacum* and *Z. mays*) to *M. huitzilopochtli*. These 25 PCGs were extracted from the complete mtDNA of each of these seven species; and then aligned with MAFFT 7.471 [61]. The rate Ka/Ks was estimated with codeml [65], which was ran online in the PAL2NAL site [66]. Accordingly, the effect of natural selection was assigned as negative selection when Ka/Ks < 1, positive selection if Ka/Ks > 1 and neutral selection if Ka/Ks = 1 [67].

Abreviations

M. huitzilopochtli: Mammillaria huitzilopochtli,

mtDNA: mitochondrial DNA,

cpDNA: chloroplast DNA,

PCG: protein coding gene

Supplementary information

Table S1: Taxonomic classification of the 21 studies species evaluated in this study.

Declarations

Ethics approval and consent to participate

The cactus species analyzed in this study is included in the Mexican Red List Species (NOM-059-SEMARNAT-2010), the sampling was authorized to S.S. with the collecting permission number SGPA/DGVS/06880/16, in accordance with the national regulations established for protected species sampled for research purposes. Dr. Salvador Arias, specialist in taxonomy of cacti, confirmed the taxonomic identity of the specimen.

Consent for publication

Not applicable.

Availability of data and materials

List of the species investigated, and their accession IDs is included in Table S1. The genome sequenced here reported is free available at GenBank site (ID: OP081771).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

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

Ciencias Biológicas, UNAM. The company Macrogen Inc., Seoul, South Korea provided sequencing

service of whole genome of *M. huitzilopochtli*.

References

1. Roger AJ, Muñoz-Gómez SA, Kamikawa R. The Origin and Diversification of Mitochondria. Curr Biol. 2017;27:R1177–92.

2. Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, et al. The alternative reality of plant mitochondrial DNA: One ring does not rule them all. PLoS Genet. 2019;15:e1008373.

3. Jacoby RP, Li L, Huang S, Lee CP, Millar AH, Taylor NL. Mitochondrial Composition, Function and Stress Response in PlantsF. J Integr Plant Biol. 2012;54:887–906.

4. Liberatore KL, Dukowic-Schulze S, Miller ME, Chen C, Kianian SF. The role of mitochondria in plant development and stress tolerance. Free Radic Biol Med. 2016;100:238–56.

5. Van Aken O, Van Breusegem F. Licensed to Kill: Mitochondria, Chloroplasts, and Cell Death. Trends Plant Sci. 2015;20:754–66.

6. Kim Y-J, Zhang D. Molecular Control of Male Fertility for Crop Hybrid Breeding. Trends Plant Sci. 2018;23:53–65.

7. Mahapatra K, Banerjee S, De S, Mitra M, Roy P, Roy S. An Insight Into the Mechanism of Plant Organelle Genome Maintenance and Implications of Organelle Genome in Crop Improvement: An Update. Front Cell Dev Biol. 2021;9:671698.

8. Chevigny N, Schatz-Daas D, Lotfi F, Gualberto JM. DNA Repair and the Stability of the Plant Mitochondrial Genome. Int J Mol Sci. 2020;21:328.

9. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. Proc Natl Acad Sci USA. 2015;112:E3515–24.

10. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, et al. Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. PLoS Biol. 2012;10.

11. Li R, Wei X, Wang Y, Zhang Y. The complete mitochondrial genome of a mangrove associated plant: Sesuvium portulacastrum and its phylogenetic implications. Mitochondrial DNA B Resour. 5:3112–3.

12. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 2003;29:380–95.

13. Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci USA. 2002;99:9905–12.

14. Rao RSP, Salvato F, Thal B, Eubel H, Thelen JJ, Møller IM. The proteome of higher plant mitochondria. Mitochondrion. 2017;33:22–37.

15. Zhao N, Wang Y, Hua J. The Roles of Mitochondrion in Intergenomic Gene Transfer in Plants: A Source and a Pool. Int J Mol Sci. 2018;19:547.

16. Rodríguez-Moreno L, González VM, Benjak A, Martí MC, Puigdomènech P, Aranda MA, et al. Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. BMC Genomics. 2011;12:424.

17. Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the Evolution of Mitochondrial Genome Size from Complete Sequences of Citrullus lanatus and Cucurbita pepo (Cucurbitaceae). Mol Biol Evol. 2010;27:1436–48.

18. Goremykin VV, Salamini F, Velasco R, Viola R. Mitochondrial DNA of *Vitis vinifera* and the Issue of Rampant Horizontal Gene Transfer. Mol Biol Evol. 2008;26:99–110.

19. Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. Origins and Recombination of the Bacterial-Sized Multichromosomal Mitochondrial Genome of Cucumber. Plant Cell. 2011;23:2499–513.

20. Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, et al. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. Science. 2013;342:1468–73.

21. Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm Amborella. Proc Natl Acad Sci USA. 2004;101:17747–52.

22. Sanchez-Puerta MV, Edera A, Gandini CL, Williams AV, Howell KA, Nevill PG, et al. Genome-scale transfer of mitochondrial DNA from legume hosts to the holoparasite *Lophophytum mirabile* (Balanophoraceae). Mol Phylogenet Evol. 2019;132:243–50.

23. Cupp JD, Nielsen BL. Minireview: DNA replication in plant mitochondria. Mitochondrion. 2014;19:231–7.

24. Drouin G, Daoud H, Xia J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol. 2008;49:827–31.

25. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA. 1987;84:9054–8.

26. Rydin C, Wikström N, Bremer B. Conflicting results from mitochondrial genomic data challenge current views of Rubiaceae phylogeny. Am J Bot. 2017;104:1522–32.

27. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci USA. 2010;107:4623–8.

28. Gitzendanner MA, Soltis PS, Wong GK-S, Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. Am J Bot. 2018;105:291–301.

29. Hunt DR. The new cactus lexicon, text and atlas. Milborne Port, UK: DH Books; 2006.

30. Hernández-Hernández T, Hernández HM, De-Nova JA, Puente R, Eguiarte LE, Magallón S. Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). Am J Bot. 2011;98:44–61.

31. Butterworth CA, Wallace RS. Phylogenetic Studies of Mammillaria (Cactaceae): Insights from Chloroplast Sequence Variation and Hypothesis Testing Using the Parametric Bootstrap. Am J Bot. 2004;91:1086–98.

32. Harpke D, Peterson A, Hoffmann M, Röser M. Phylogenetic evaluation of chloroplast trnL–trnF DNA sequence variation in the genus Mammillaria (Cactaceae). Schlechtendalia. 2006;14:7–16.

33. Solórzano S, Chincoya DA, Sanchez-Flores A, Estrada K, Díaz-Velásquez CE, González-Rodríguez A, et al. De Novo Assembly Discovered Novel Structures in Genome of Plastids and Revealed Divergent Inverted Repeats in Mammillaria (Cactaceae, Caryophyllales). Plants. 2019;8:392.

34. Chincoya DA, Sanchez-Flores A, Estrada K, Díaz-Velásquez CE, González-Rodríguez A, Vaca-Paniagua F, et al. Identification of High Molecular Variation Loci in Complete Chloroplast Genomes of Mammillaria (Cactaceae, Caryophyllales). Genes. 2020;11:830.

35. Wu Z, Cuthbert JM, Taylor DR, Sloan DB. The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. Proc Natl Acad Sci USA. 2015;112:10185–91.

36. Cheng Y, He X, Priyadarshani SVGN, Wang Y, Ye L, Shi C, et al. Assembly and comparative analysis of the complete mitochondrial genome of Suaeda glauca. BMC Genomics. 2021;22:167.

37. Cui H, Ding Z, Zhu Q, Wu Y, Qiu B, Gao P. Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer. Sci Rep. 2021;11:1595.

38. Stern DB, Palmer JD. Recombination sequences in plant mitochondrial genomes: diversity and homologies to known mitochondrial genes. Nucl Acids Res. 1984;12:6141–57.

39. Cole LW, Guo W, Mower JP, Palmer JD. High and Variable Rates of Repeat-Mediated Mitochondrial Genome Rearrangement in a Genus of Plants. Mol Biol Evol. 2018. https://doi.org/10.1093/molbev/msy176.

40. Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. The plant mitochondrial genome: Dynamics and maintenance. Biochimie. 2014;100:107–20.

41. Putintseva YA, Bondar EI, Simonov EP, Sharov VV, Oreshkova NV, Kuzmin DA, et al. Siberian larch (Larix sibirica Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. BMC Genomics. 2020;21:654.

42. Jackman SD, Coombe L, Warren RL, Kirk H, Trinh E, MacLeod T, et al. Complete Mitochondrial Genome of a Gymnosperm, Sitka Spruce (Picea sitchensis), Indicates a Complex Physical Structure. Genome Biol Evol. 2020;12:1174–9.

43. Warren JM, Salinas-Giegé T, Triant DA, Taylor DR, Drouard L, Sloan DB. Rapid Shifts in Mitochondrial tRNA Import in a Plant Lineage with Extensive Mitochondrial tRNA Gene Loss. Mol Biol Evol. 2021;38:5735–51.

44. Wang D, Wu Y-W, Shih AC-C, Wu C-S, Wang Y-N, Chaw S-M. Transfer of Chloroplast Genomic DNA to Mitochondrial Genome Occurred At Least 300 MYA. Mol Biol Evol. 2007;24:2040–8.

45. Chaw S-M, Chun-Chieh Shih A, Wang D, Wu Y-W, Liu S-M, Chou T-Y. The Mitochondrial Genome of the Gymnosperm Cycas taitungensis Contains a Novel Family of Short Interspersed Elements, Bpu Sequences, and Abundant RNA Editing Sites. Mol Biol Evol. 2008;25:603–15.

46. Brockington SF, Alexandre R, Ramdial J, Moore MJ, Crawley S, Dhingra A, et al. Phylogeny of the Caryophyllales Sensu Lato: Revisiting Hypotheses on Pollination Biology and Perianth Differentiation in the Core Caryophyllales. Int J Plant Sci. 2009;170:627–43.

47. Yao G, Jin J-J, Li H-T, Yang J-B, Mandala VS, Croley M, et al. Plastid phylogenomic insights into the evolution of Caryophyllales. Mol Phylogenet Evol. 2019;134:74–86.

48. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 4 Feb 2020.

49. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.

50. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

51. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45:e18–e18.

52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

53. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13:e1005595.

54. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Curr Protoc Bioinformatics. 2020;70.

55. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies: Fig. 1. Bioinformatics. 2015;31:3350–2.

56. Bushnell B, Rood J, Singer E. BBMerge – Accurate paired shotgun read merging via overlap. PLoS ONE. 2017;12:e0185056.

57. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:W59–64.

58. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33:2583–5.

59. Kurtz S. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29:4633–42.

60. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.

61. 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;30:3059–66.

62. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.

63. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015;32:268–74.

64. Sokal RR, Rohlf FJ. Introduction to biostatistics. 2nd ed., Dover ed. Mineola, N.Y: Dover Publications; 2009.

65. Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol. 1994. https://doi.org/10.1093/oxfordjournals.molbev.a040153.

541 66. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corre-
542 sponding codon alignments. Nucleic Acids Res. 2006;34 Web Server:W609–12.

543 67. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18:486–7.

544

545

546

547

548

549

550

551 **Table 1** Gene composition of the mitochondrial genome of *Mammillaria huitzilopochtli* grouped by protein coding genes,
552 ribosomal and transfer RNAs. Protein coding genes were of ten different gene families; for each of these genes is showed its
553 length, its start and stop codons; and the number of amino acids transcribed

Type of genes	Gene name	Length (bp)	Start codon	Stop codon	Amino acid
I. Protein Coding Genes					
1. NADH dehydrogenase	*nad1*[a]	978	ACG	TAA	325
	nad2[a]	1467	ATG	TAA	488
	nad3	357	ATG	TAA	118
	nad4[a]	1488	ATG	TGA	495
	nad4L	273	ATG	TAA	90
	nad5[a]	1992	ATG	TAA	663
	nad6	690	ATG	TAA	229
	nad7[a]	1092	ATG	TAG	363
	nad9	579	ATG	TAA	192
2. ATP synthase	*atp1*	1530	ATG	TGA	509
	atp4	552	ATG	TAA	183
	atp6	726	ATG	TAG	241
	atp8	489	ATG	TAA	162
	atp9	225	ATG	CGA	74
3. Cytochrome c biogenesis	*ccmB*	621	ATG	TGA	206
	ccmC	720	ATG	TGA	239
	ccmFC[a]	1341	ATG	TAG	446
	ccmFN	1740	ATG	TGA	579
4. Cytochrome c oxidase	*cox1*	1575	ATG	TAA	524
	cox2[a]	834	ATG	TAG	277
	cox3	798	ATG	TGA	265
5. Maturase	*matR*	1992	ATG	TAG	663
6. Ubiquinol cytochrome c reductase	*cob*	1182	ATG	TGA	393
7. Ribosomal proteins (LSU)	*rpl5*	561	ATG	TAA	186
	rpl16	477	GTG	TAA	158
8. Ribosomal proteins (SSU)	*rps1*	606	ATG	TAA	201
	rps3[a]	1686	ATG	TGA	561
	rps4	1098	TTG	TAA	365
	rps7	447	ATG	TAA	148
	rps12	378	ATG	TGA	125
	rps13	351	ATG	TGA	116
9. Methyltransferase	*mttb*	783	ATA	TAA	260

10. Cytochrome b	*sdh4*	423	ATG	TGA	140
II. Ribosomal RNAs					
11. rrn	*rrn5*	119			
	rrn18	1818			
	rrn26	2859			
III. Transfer RNAs					
12. trn	*trnC-GCA*	73			
	trnD-GUC (1)[b]	75,74			
	trnE-UUC	72			
	trnF-GAA	74			
	trnfM-CAU	60			
	trnG-GCC	72			
	trnH-AUG	70			
	trnH-GUG	74			
	trnI-CAU	74			
	trnK-UUU (1)	73,73			
	trnM-CAU (4)	72,74,73,73, 72			
	trnN-GUU	72			
	trnP-UGG (1)	75,74			
	trnQ-UUG (1)	72,72			
	trnS-GCU	88			
	trnS-UGA	87			
	trnY-GUA	83			
	trnV-GAC[b]	72			
	trnW-CCA (1)	74,74			

554   Note: In parenthesis is presented the number of additional copies annotated

555   a. genes with introns; b. genes of plastid origin

DISCUSIÓN GENERAL

El ensamble del genoma mitocondrial de *Mammillaria huitzilopochtli* resultó en una sola secuencia con una longitud de 2,050,004 pb. Esta secuencia de ADNmt representa la primera reportada para la familia Cactaceae. Las comparaciones realizadas en término de tamaño, contenido de GC, contenido total de genes y composición genética, permitieron observar los altos niveles de variación estructural que presenta el ADNmt de plantas.

En particular, la especie *M. huitzilopochtli* tiene el tercer genoma más grande reportado a la fecha para el orden Caryophyllales, por debajo de *Silene noctiflora* (7.1 Mpb) (Wu et al., 2015) y *S. conica* (11.3 Mpb) (Sloan et al., 2012). Sin embargo, en el universo de las plantas terrestres, el alerce siberiano (*Larix sibirica*) tiene, hasta el momento, el ADNmt más grande con un total de 11.7 Mpb (Putintseva *et al.*, 2020). Los procesos que determinan esta variación en el tamaño aún no están del todo claros. Además, las expansiones o contracciones parecen estar limitadas a regiones intergénicas, dada la poca variación que hubo en la composición genética de las especies comparadas. Una evaluación y caracterización de las grandes regiones intergénicas podría develar el rol que cumplen en el ADNmt de plantas. Algunos estudios han encontrado grupos de marcos abierto de lectura (ORFs) de función desconocida qué son expresados (Negruk, 2013; Omelchenko *et al.*, 2020) y que podrían tener varios roles en la función mitocondrial. La comparación de ORFs, ha permitido proponer genes candidatos con implicaciones en la esterilidad citoplásmica de los machos en algodón (*Gossypium,* Li *et al.*, 2018) y arroz (*Oryza,* Omukai *et al.*, 2021). Resultados como estos permiten proponer al ADNmt de plantas como una fuente de genes huérfanos (genes exclusivos de un linaje) que pueden ser de gran relevancia para el entendimiento de la evolución de las plantas (O'Conner y Li, 2020).

Las secuencias repetidas localizadas en la mitocondria suelen ser de tipo microsatélite, es decir, pueden ser directos o invertidos. Estas secuencias aportan a la expansión de los genomas y constituyen una alta proporción del genoma total. Por ejemplo, en el pepino (*Cucumis sativus*), se identificó que hasta el 36% del genoma está conformado de regiones repetitivas, siendo aquellos menores a 50 pb los más abundantes (Alverson et al., 2011). En varios estudios se ha propuesto que estos repetidos participan en la replicación del ADNmt (Cupp y Nielsen, 2014) o en la recombinación mediada por repetidos (Cole et al., 2018), y por lo tanto, podrían cumplir un rol importante en los rearreglos estructurales del ADNmt (Mahapatra et al., 2021). Otro tipo de repetidos comunes en el ADNmt son los microsatélites, esto podría implicar que el genoma mitocondrial es una importante fuente de variación que

puede tener gran utilidad en estudios poblacionales, y su abundancia, polimorfismo y facilidad de detección por PCR facilita su uso (Powell et al., 1996).

La composición de genes del ADNmt en plantas parece tener implicaciones filogenéticas, sin importar los factores estructurales del ADNmt. Al comparar la composición genéticas de las 21 especies, se observó poca variación entre las especies cercanamente relacionadas, sin embargo, entre las plantas terrestres, las gimnospermas parecen tener una mayor cantidad de genes únicos que las angiospermas. Por ejemplo, *Cycas taitungensis* tiene la mayor cantidad de genes únicos con excepción de *trnI-GAU* y *trnL-CAA*, lo cual representa una composición de genes muy similar a otras especies de gimnospermas (Jackman et al., 2020; Putintseva et al., 2020). Sin embargo, es importante mencionar, que la ausencia de genes en el ADNmt no implica la pérdida de estos, sino su migración al núcleo (Cui *et al.*, 2021).

Al comparar el genoma mitocondrial con el de cloroplasto de *M. huitzilopochtli*, se pudieron identificar aquellas regiones que han migrado de la mitocondria hacia el cloroplasto. Un total de cinco genes de cloroplasto con una copia adicional en el ADNmt (*psaC*, dos copias de *trnD-GUC*, *trnN-GUU* y *trnI-CAU*), fueron identificados. En contraste, la presencia de ARNs de transferencia de origen del cloroplasto en el ADNmt es común en plantas. De hecho, hay estudios previos que indican que estos ARNs de transferencia mantienen su función en la mitocondria y participan en la síntesis de proteínas (Joyce y Gray, 1989; Warren et al., 2021). Conocer la función que cumple el gen *psaC* en la mitocondria no es posible por métodos bioinformáticos, por lo tanto, no podemos, por el momento, elucidar las razones y las consecuencias de la migración de este gen codificante.

El análisis de las tasas de sustituciones no sinónimas sobre sustituciones sinónimas (Ka/Ks) y su comparación con otras especies de angiospermas revela algunas de las fuerzas evolutivas a las que está sujeto el ADNmt, los cuales son consistentes con otros estudios realizados en plantas (Cheng et al. 2021). La mayoría de los genes codificantes de proteínas están sujetos a selección negativa, lo que nos indica que estos son funcionalmente importantes y son conservados en la mayoría de las plantas terrestres (Cheng et al. 2021).

El árbol filogenético obtenido en este trabajo muestra una topología concordante con las filogenias previamente obtenidas a partir de loci de cloroplasto. Las siete familias se organizan en un grupo monofilético y se organizan dentro de los mismos clados recuperados en una filogenia reconstruida por Yao y colaboradores (2019), a partir del uso de 83 marcadores de cloroplasto. En este análisis filogenético, se recuperó el clado de inclusión globular (Cactaceae + Aizoaceae + Nyctaginaceae), Centrospermae (Amaranthaceae +

Chenopodiaceae + Caryophyllaceae) y Non-core Caryophyllales (Nepenthaceae + Polygonaceae). Estos resultados nos indican que los marcadores mitocondriales usados para la reconstrucción filogenética tienen la suficiente resolución para separar familias del orden Caryophyllales. A partir de estos resultados, se considera que la facilidad de secuenciación y técnicas como *genome skimming* permitirán obtener una mayor cantidad de datos mitocondriales (Dodsworth, 2015; Nevill *et al.*, 2020) y favorecerán futuros estudios filogenómicos del ADNmt que resalten la importancia de este genoma como fuente de variación. De hecho, recientemente, algunos estudios se han propuesto evaluar esto. Así, están disponibles las filogenias reconstruidas en angiospermas para la familia Rubiaceae (Gentianales) a partir de loci de cloroplasto y mitocondria, los cuales mostraron discordancias. Los autores atribuyen estas discordancias a eventos de hibridación de las especies y enfatizaron la importancia del genoma mitocondrial como fuente de información filogenética (Rydin *et al.*, 2017). Resultados similares se han reportado en la familia Poaceae (Poales), en la cual las discordancias no sólo se limitan a las filogenias reconstruidas con genomas de orgánulos, sino también a aquellas inferidas a partir de loci nucleares. En este caso, los autores sugieren que la intricada historia evolutiva causada por la radiación adaptativa, la poliploidía y la hibridación de muchos grupos de plantas, podría ser la causante de estas conflictivas señales evolutivas (Wu *et al.*, 2022).

CONCLUSIONES GENERALES

En este estudio se ensambló y anotó el genoma mitocondrial de *M. huitzilopochtli.* El genoma es lineal con una longitud de 2,050,004 pb. Un total de 65 genes fueron anotados. De ellos 34 son genes codificantes de proteínas, 28 corresponden a ARNs de transferencia y tres a ARNs ribosomales. Los resultados obtenidos mostraron diferencias de tamaño entre las especies comparadas, los valores de GC y la composición de genes conservados. La estimación de tasas Ka/Ks reveló que los genes codificantes de proteínas son conservados y que están bajo una selección negativa. Las relaciones filogenéticas de las familias de Caryophyllales fueron concordantes con lo previamente reportado con para los loci de cloroplasto.

Este trabajo permitió reevaluar la importancia del genoma mitocondrial como una fuente importante de variación molecular y estructural en plantas terrestres. Esta variación debe ser tomada en cuenta para responder preguntas evolutivas de plantas y para construir las relaciones filogenéticas de las especies. Sin embargo, la complejidad de este genoma aún representa un reto bioinformático, ya que es necesario el uso de tecnologías de secuenciación de última generación que generen lecturas largas para determinar los rearreglos estructurales de este genoma. Además, se requiere profundizar en el estudio del genoma mitocondrial para conocer en detalle el origen, causas y consecuencias de su alta variación molecular y estructural, así como para caracterizar en detalle las grandes regiones intergénicas de la mitocondria en plantas.

REFERENCIAS BIBLIOGRÁFICAS

Adams K L, Palmer J D. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution* **29**:380-395.

Adams K L, Qiu Y-L, Stoutemyer M, Palmer J D. 2002. Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proceedings of the National Academy of Sciences* **99**:9905-9912.

Alverson A J, Rice D W, Dickinson S, Barry K, Palmer J D. 2011. Origins and Recombination of the Bacterial-Sized Multichromosomal Mitochondrial Genome of Cucumber. *The Plant Cell* **23**:2499-2513.

del Ángel-Piña C. 2005. Análisis cromosómico y determinación del tamaño del genoma por citometría de flujo en especies de Mammillaria de la serie Supertextae (Cactaceae). Tesis de Maestría, UNAM, México, D.F.

Arakaki M, Christin P-A, Nyffeler R, Lendel A, Eggli U, Ogburn R M, Spriggs E, Moore M J, 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.

Araújo W L, Nunes-Nesi A, Fernie A R. 2014. On the role of plant mitochondrial metabolism and its impact on photosynthesis in both optimal and sub-optimal growth conditions. *Photosynthesis Research* **119**:141-156.

Bárcenas R T, Yesson C, Hawkins J A. 2011. Molecular systematics of the Cactaceae. *Cladistics* **27**:470-489.

Bensasson D, Zhang D-X, Hartl D L, Hewitt G M. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution* **16**:314-321.

Bergthorsson U, Richardson A O, Young G J, Goertzen L R, Palmer J D. 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm Amborella. *Proceedings of the National Academy of Sciences* **101**:17747-17752.

Blanchard J L, Schmidt G W. 1995. Pervasive migration of organellar DNA to the nucleus in plants. *Journal of Molecular Evolution* **41**:397-406.

Brockington S F, Alexandre R, Ramdial J, Moore M J, Crawley S, Dhingra A, Hilu K, Soltis D E, Soltis P S. 2009. Phylogeny of the Caryophyllales Sensu Lato: Revisiting Hypotheses on Pollination Biology and Perianth Differentiation in the Core Caryophyllales. *International Journal of Plant Sciences* **170**:627-643.

Burger G, Gray M W, Franz Lang B. 2003. Mitochondrial genomes: anything goes. *Trends in Genetics* **19**:709-716.

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**:1086-1098.

Cota-Sánchez J H, Bomfim-Patrício M C. 2010. Seed morphology, polyploidy and the evolutionary history of the epiphytic cactus Rhipsalis baccifera (Cactaceae). *Polibotánica*:107-129.

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, E.U.A.

Cui H, Ding Z, Zhu Q, Wu Y, Qiu B, Gao P. 2021. Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer. *Scientific Reports* **11**:1595.

Dodsworth S. 2015. Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science* **20**:525-527.

Dong S, Chen L, Liu Y, Wang Y, Zhang S, Yang L, Lang X, Zhang S. 2020. The draft mitochondrial genome of Magnolia biondii and mitochondrial phylogenomics of angiosperms. *PLOS ONE* **15**:e0231020.

Goremykin V V, Salamini F, Velasco R, Viola R. 2008. Mitochondrial DNA of *Vitis vinifera* and the Issue of Rampant Horizontal Gene Transfer. *Molecular Biology and Evolution* **26**:99-110.

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**:1382-1397.

Hernández-Hernández T, Hernández H M, De-Nova J A, Puente R, Eguiarte L E, Magallón S. 2011. Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *American Journal of Botany* **98**:44-61.

Hu J, Huang W, Huang Q, Qin X, Yu C, Wang L, Li S, Zhu R, Zhu Y. 2014. Mitochondria and cytoplasmic male sterility in plants. *Mitochondrion* **19**:282-288.

Hunt D R. 2006. The new cactus lexicon, text and atlas. DH Books, Milborne Port, UK.

Joppa L N, Roberts D L, Pimm S L. 2011. How many species of flowering plants are there? *Proceedings of the Royal Society B: Biological Sciences* **278**:554-559.

Kim C-K, Kim Y-K. 2018. The multipartite mitochondrial genome of Fallopia multiflora (Caryophyllales: Polygonaceae). *Mitochondrial DNA Part B* **3**:155-156.

Knoop V, Volkmar U, Hecht J, Grewe F. 2011. Mitochondrial Genome Evolution in the Plant Lineage. *Plant Mitochondria*:3-29.

Kozik A, Rowan B A, Lavelle D, Berke L, Schranz M E, Michelmore R W, Christensen A C. 2019. The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLOS Genetics* **15**:e1008373.

Lang B F, Gray M W, Burger G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annual Review of Genetics* **33**:351-397.

Leushkin E V, Sutormin R A, Nabieva E R, Penin A A, Kondrashov A S, Logacheva M D. 2013. The miniature genome of a carnivorous plant Genlisea aurea contains a low number of genes and short non-coding sequences. *BMC Genomics* **14**:476.

Li S, Chen Z, Zhao N, Wang Y, Nie H, Hua J. 2018. The comparison of four mitochondrial genomes reveals cytoplasmic male sterility candidate genes in cotton. *BMC Genomics* **19**:775.

López-Ortiz N. 2017. Filogeografía y diversificación de Mammillaria. Tesis de Maestría, UNAM, Ciudad de México.

Mering S von et al. 2019. The Global Caryophyllales Initiative: Towards an updated taxonomic backbone and a dynamic monograph of a major plant group. *Biodiversity Information Science and Standards* **3**.

Mower J P, Sloan D B, Alverson A J. 2012. Plant Mitochondrial Genome Diversity: The Genomics Revolution. *Plant Genome Diversity Volume 1*:123-144.

Negruk V. 2013. Mitochondrial Genome Sequence of the Legume Vicia faba. *Frontiers in Plant Science* **4**.

Nevill P G, Zhong X, Tonti-Filippini J, Byrne M, Hislop M, Thiele K, van Leeuwen S, Boykin L M, Small I. 2020. Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics. *Plant Methods* **16**:1.

O'Conner S, Li L. 2020. Mitochondrial Fostering: The Mitochondrial Genome May Play a Role in Plant Orphan Gene Evolution. *Frontiers in Plant Science* **11**:600117.

Omelchenko D O, Makarenko M S, Kasianov A S, Schelkunov M I, Logacheva M D, Penin A A. 2020. Assembly and Analysis of the Complete Mitochondrial Genome of Capsella bursa-pastoris. *Plants* **9**:469.

Omukai S, Arimura S, Toriyama K, Kazama T. 2021. Disruption of mitochondrial *open reading frame 352* partially restores pollen development in cytoplasmic male sterile rice. *Plant Physiology* **187**:236-246.

Palmer J D, Adams K L, Cho Y, Parkinson C L, Qiu Y-L, Song K. 2000. Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences* **97**:6960-6966.

Pellicer J, Fay M F, Leitch I J. 2010. The largest eukaryotic genome of them all? *Botanical Journal of the Linnean Society* **164**:10-15.

Putintseva Y A et al. 2020. Siberian larch (Larix sibirica Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. *BMC Genomics* **21**:654.

Rice D W et al. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. *Science (New York, N.Y.)* **342**:1468-1473.

Richardson A O, Palmer J D. 2007. Horizontal gene transfer in plants. *Journal of Experimental Botany* **58**:1-9.

Rodríguez-Moreno L, González V M, Benjak A, Martí M C, Puigdomènech P, Aranda M A, Garcia-Mas J. 2011. Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC Genomics* **12**:424.

Roger A J, Muñoz-Gómez S A, Kamikawa R. 2017. The Origin and Diversification of Mitochondria. *Current Biology* **27**:R1177-R1192.

Röschenbleck J, Wicke S, Weinl S, Kudla J, Müller K F. 2016. Genus-Wide Screening Reveals Four Distinct Types of Structural Plastid Genome Organization in Pelargonium (Geraniaceae). *Genome Biology and Evolution* **9**:64-76.

Rydin C, Wikström N, Bremer B. 2017. Conflicting results from mitochondrial genomic data challenge current views of Rubiaceae phylogeny. *American Journal of Botany* **104**:1522-1532.

Saxena R K, Edwards D, Varshney R K. 2014. Structural variations in plant genomes. *Briefings in Functional Genomics* **13**:296-307.

Skippington E, Barkman T J, Rice D W, Palmer J D. 2015. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. *Proceedings of the National Academy of Sciences* **112**:E3515-E3524.

Sloan D B, Alverson A J, Chuckalovcak J P, Wu M, McCauley D E, Palmer J D, Taylor D R. 2012. Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. *PLoS Biology* **10**.

Sloan D B, Alverson A J, Štorchová H, Palmer J D, Taylor D R. 2010. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm Silene latifolia. *BMC Evolutionary Biology* **10**:274.

Solórzano S, Chincoya D A, Sanchez-Flores A, Estrada K, Díaz-Velásquez C E, González-Rodríguez A, Vaca-Paniagua F, Dávila P, Arias S. 2019. De Novo Assembly Discovered Novel Structures in Genome of Plastids and Revealed Divergent Inverted Repeats in Mammillaria (Cactaceae, Caryophyllales). *Plants* **8**:392.

Van Aken O, Van Breusegem F. 2015. Licensed to Kill: Mitochondria, Chloroplasts, and Cell Death. *Trends in Plant Science* **20**:754-766.

Ward B L, Anderson R S, Bendich A J. 1981. The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). *Cell* **25**:793-803.

Warren J M, Salinas-Giegé T, Triant D A, Taylor D R, Drouard L, Sloan D B. 2021. Rapid Shifts in Mitochondrial tRNA Import in a Plant Lineage with Extensive Mitochondrial tRNA Gene Loss. *Molecular Biology and Evolution* **38**:5735-5751.

Won H, Renner S S. 2003. Horizontal gene transfer from flowering plants to Gnetum. *Proceedings of the National Academy of Sciences* **100**:10824-10829.

Wu H, Yang J-B, Liu J-X, Li D-Z, Ma P-F. 2022. Organelle Phylogenomics and Extensive Conflicting Phylogenetic Signals in the Monocot Order Poales. *Frontiers in Plant Science* **12**:824672.

Wu Z, Cuthbert J M, Taylor D R, Sloan D B. 2015. The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **112**:10185-10191.

Xi Z, Wang Y, Bradley R K, Sugumaran M, Marx C J, Rest J S, Davis C C. 2013. Massive Mitochondrial Gene Transfer in a Parasitic Flowering Plant Clade. *PLoS Genetics* **9**.

Xiong A-S et al. 2008. Gene duplication and transfer events in plant mitochondria genome. *Biochemical and Biophysical Research Communications* **376**:1-4.

Yao G et al. 2019. Plastid phylogenomic insights into the evolution of Caryophyllales. *Molecular Phylogenetics and Evolution* **134**:74-86.