Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California

Maestría en Ciencias en Ciencias de la Vida con orientación en Biología Ambiental

---

Filogeografía de *Yucca valida* en la Península de Baja California

Tesis
para cubrir parcialmente los requisitos necesarios para obtener el grado de Maestro en Ciencias

Presenta:

José Alberto López Alemán

Ensenada, Baja California, México
2020

Tesis defendida por

José Alberto López Alemán

y aprobada por el siguiente Comité

Dra. María Clara Arteaga Uribe
Directora de tesis

Dra. Fadia Sara Ceccarelli

Dr. Jaime Gasca Pineda

Dr. Axayácatl Rocha Olivares

Dra. Patricia Juárez Camacho
Coordinadora del Posgrado en Ciencias de la Vida

Dra. Rufina Hernández Martínez
Directora de Estudios de Posgrado

Resumen de la tesis que presenta **José Alberto López Alemán** como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la Vida con orientación en Biología Ambiental.

Filogeografía de *Yucca valida* en la Península de Baja California

Resumen aprobado por:

_____

Dra. María Clara Arteaga Uribe
Directora de tesis

La especie *Yucca valida* (Asparagaceae) es una planta endémica de la península de Baja California que se distribuye en el desierto Central, el desierto del Vizcaíno y las planicies de Magdalena. En esta región se han registrado divergencias genéticas de diversas especies, que se han asociado a eventos históricos geológicos y climáticos ocurridos en la península. Dado que *Y. valida* es un elemento dominante en el paisaje de la región, se espera que sus poblaciones hayan experimentado una fragmentación relacionada con dichos eventos. Con el fin de identificar cuántos linajes genéticos existen a lo largo de la distribución de *Y. valida*, reconstruir su historia demográfica e interpretarla en el contexto de la historia de la península, se evaluó la distribución de la variación genética de la especie por medio de polimorfismos de un solo nucleótido (SNPs) neutrales. Adicionalmente, se estimó el tiempo de origen de la especie a partir de análisis basados en el reloj molecular y para ello se utilizaron secuencias potencialmente provenientes del genoma del cloroplasto. Se genotipificaron 140 individuos provenientes de 20 localidades utilizando 1440 SNPs. Los análisis de estructura genética y las genealogías nucleares fueron congruentes con la identificación de tres linajes monofiléticos divergentes. Las simulaciones de escenarios históricos proporcionaron soporte para la divergencia simultánea de los tres linajes nucleares. Se detectó alta diversidad genética y una asociación positiva y significativa entre las distancias geográficas y la diferenciación genética, indicando un patrón de aislamiento por distancia. Para la estimación del tiempo de divergencia de la especie, se recuperaron 36 secuencias potencialmente provenientes del genoma del cloroplasto de *Y. valida*. En el análisis filogenético, las 36 secuencias de *Y. valida* se agruparon en un único clado, sugiriendo un origen monofilético. La estimación del tiempo de divergencia indica que el origen de la especie ocurrió aproximadamente entre 190,000 y 420,000 años. Considerando la edad estimada de la especie, se propone que el origen de los linajes nucleares detectados está relacionado con una reducción del hábitat adecuado para la especie y una fragmentación de sus poblaciones provocada por cambios climáticos históricos y es mantenido por un flujo genético restringido.

Palabras clave: *Yucca valida*, estructura genética, aislamiento por distancia, tiempo de divergencia, historia demográfica.

Abstract of the thesis presented by **José Alberto López Alemán** as a partial requirement to obtain the Master of Science degree in Life Sciences with orientation in Environmental Biology

***Yucca valida* phylogeography in the Baja California Peninsula**

Abstract approved by:

_____
Dra. María Clara Arteaga Uribe
Thesis Director

Yucca valida (Asparagaceae) is an endemic plant of the Baja California Peninsula that distributes through the Central Desert, the Vizcaino Desert and the Magdalena Plains. Genetic divergences associated with historical geological and climatic events have been recorded for various species in these regions. Since *Y. valida* populations are dominant in this area; it is expected that they have experienced fragmentation related to these events. To identify how many genetic lineages *Y. valida* exist within its geographical distribution, to reconstruct the species' demographic history and interpret it in the context of the history of the Peninsula; the distribution of the genetic variation of the species was evaluated through neutral Single Nucleotide Polymorphisms (SNPs). Additionally, the species' diversification time was estimated from analyzes based on the molecular clock, employing putative chloroplast genome sequences. A total of 140 individuals from 20 localities were genotyped using 1440 SNPs. The genetic structure analyzes and nuclear genealogies were consistent with the identification of three divergent monophyletic lineages. Historical scenario simulations provided support for the simultaneous divergence of the three nuclear lineages. A high genetic diversity was detected. A positive and significant relation between geographic distances to genetic differentiation indicated isolation by distance. To estimate the divergence time of the species, 36 putative chloroplast genome sequences of *Y. valida* were recovered. In the phylogenetic analysis, the 36 sequences of *Y. valida* grouped into a single clade, suggesting a monophyletic origin. The estimated diversification time indicates that the origin of the species occurred between 190,000 and 420,000 years ago. Based on the age estimated for the species, it is proposed that the diversification of the three nuclear lineages is related to the reduction of the suitable habitat and the fragmentation of species' populations caused by historical climatic changes while a restricted genetic flow is maintaining the differentiation signals.

Keywords: *Yucca valida*, genetic structure, isolation by distance, divergence time, demographic history.

Dedicatoria

A Yolanda De La Vega Montealegre,

Andrés Alemán Domínguez,

y Alberto De La Vega Montealegre.

A sus historias.

A todo lo que vive de ellos en mí.

Agradecimientos

Al **Consejo Nacional de Ciencia y Tecnología (CONACyT)**, por el apoyo económico brindado a mi persona para la realización de mis estudios de posgrado y tesis de maestría a través de la beca al CVU 748805, y por el financiamiento del proyecto**:** "Diversidad genética y variación fenotípica de las *Yuccas* (Agavaceae) de la península de Baja California" (CB2014-01-238843) otorgado a la Dra. María Clara Arteaga Uribe.

Al **Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE)** y al **Posgrado en Ciencias de la Vida**. Ha sido un honor alcanzar mi formación como biólogo e investigador en su institución.

Al **Departamento de Biología de la Conservación**, por todas las experiencias. Soy testigo de que cada persona ahí es un acervo invaluable de conocimiento y por si eso fuera poco, con una actitud asombrosa. Académicos y no académicos forman una bella familia en esa área. Me llevo mucho de ustedes conmigo.

A todo el **Comité de tesis**. Ustedes son unos gigantes. Aprecio infinitamente me permitieran ponerme en pie sobre sus hombros para mirar el mundo desde ahí. Les guardo un infinito aprecio y toda mi admiración.

A mi directora de tesis, la **Dra. María Clara Arteaga Uribe**, porque conocerte ha sido todo un parteaguas. Te admiro en muchísimos aspectos y agradezco la oportunidad de haber sido parte en tu laboratorio. Anhelo convertirme en un científico de tu nivel y espero un día tener esa capacidad explicativa tan característica en ti. Gracias por ser claridosa, por cuestionarlo todo. Te considero un gran ejemplo a seguir.

Al **Dr. Jaime Gasca Pineda**, por llenar el camino de pistas, por la constante disposición y confianza en mí. Fuiste la primer persona en recibirme en CICESE y desde entonces charlar contigo es como ver un mapa.

A la **Dra. Fadia Sara Ceccarelli**, por la atención y el interés que mostró al proyecto y sus dudas tan precisas.
Al **Dr. Axayácatl Rocha Olivares**, por los aportes elocuentes y la visión ecléctica que aportó a cada reunión.

Al **Dr. Rafael Bello Bedoy**, por recordarme el valor del trabajo bien hecho y la importancia de la constancia y dedicación. Gracias a ti me fue posible ir dando pasos en el posgrado incluso antes de ser aceptado. Valoro haber estado en tu laboratorio y aún me siento un poco apenado por mi desempeño en tu materia.

A la **Dra. Jimena Carrillo Tripp**, por brindar los mejores consejos en breves pero enriquecedoras charlas.

A los integrantes del **laboratorio de Genética de la Conservación**. Especialmente a la **M. C. Cynthia Rocío Álamo**, la **M. C. Lita Castañeda**, la **M. C. Astrid Luna Ortiz** y el **Dr. Leonardo De La Rosa Conroy** por todos sus aportes al proyecto, por el compañerismo, la actitud, sus reflexiones, experiencias y su calidez humana.

A **Karime Andrade Meneses**, porque tu soporte me brindó mucha estabilidad. Eres una gran guía. Gracias.

A la **PhD Flavia Termignoni**, porque en pocas charlas me diste tonelada y media de inspiración científica.

A la futura **Dra. Julieta Orozco** y al **Dr. Ricardo Mejía**, porque su apoyo incondicional, enseñanzas y confianza fueron clave para que concluyera la licenciatura e ingresara al posgrado. Tienen un amigo en mí.

A **Daniela Félix**, **Daniela Núñez**, **Haran Peiro, Carlos González**, **Alejandro Rodríguez**, **Israel Negrete**, **Julio Fortis** y **Mario González** por su amistad franca. Porque he aprendido mucho en compañía y de cada uno de ustedes. Por todas las anécdotas. Es un orgullo considerarme colega y amigo de biólogos tan biólogos.

A **Myriam** y **Fernando Barillas**, **Luis Enrique** y **Omar Ruiz** y **Miguel Mora**, por aceptarme en su manada. Por la lealtad, el coraje, la devoción y el cariño que han hecho posibles estos años de amistad. Por todas las experiencias vividas y porque siempre estaremos aquí para levantar las manos cuando alguno lo necesite. Porque en la casa de quien fuere, a la mesa todos somos un hijo más. Porque compartir la felicidad con ustedes la hace más real. Son las personas más sensacionales en todo el universo conocido.

A **Esmeralda Bravo Hernández**, por recibirme en tu corazón y en tu hogar. Por recordarme que escuche, que intente ver con otros ojos y por la manera en que alimentamos nuestros sueños. Por los senderos recorridos y los que aguardan. Por el amor mutuo y las observaciones sin juicios. Por mostrarme infinitas posibilidades para el mañana. ¿Qué más vamos a hacer de la vida, sino descubrir lo que queremos de ella?

A **Maribel Alemán De La Vega**, por enseñarme el amor a la letra escrita y al estudio, porque bajo tu tutela empezó mi pasión por la ciencia. Por darle significado a la palabra familia. Porque siempre estás ahí cuando me voy y cuando llego a casa, por ser mi segunda madre. Por resguardar el legado de tantas generaciones.

A **Yolanda De La Vega Montealegre** y **Andrés Alemán Domínguez**, los abuelos que eligieron convertirse en mis padres. Por llevarme de la mano hasta que pude andar, porque llegué hasta aquí gracias a ustedes. Porque a esta alegría le hace falta la luz de sus ojos. Este logro es nuestro. Los llevo conmigo en cada paso.

Tabla de contenido

Lista de figuras

Lista de tablas

Capítulo 1. Introducción

La distribución de la variación genética de las especies está influenciada por procesos estocásticos y determinísticos que pueden ocurrir de manera simultánea a lo largo de distintas escalas temporales y espaciales (Hedrick, 2011). Conocer los efectos de estos procesos sobre la distribución de la variación genética permite relacionar la historia demográfica de las especies con la historia climática y geológica de una región determinada (Avise, 2000; Arbogast y Kenagy, 2001).

Determinar el impacto de la historia evolutiva de las especies sobre la variación genética requiere distinguir las señales que la selección natural, la mutación, la deriva y el flujo imprimen en el genoma. Mientras que la selección produce una fijación diferencial de los alelos dentro una población, la mutación y la deriva tienden a generar fluctuaciones al azar en las frecuencias alélicas entre las generaciones dentro de dicha población. Por su parte, el flujo genético dentro y entre las poblaciones tiene consecuencias en la heterogeneidad o uniformidad de la variación genética, de tal manera que un flujo genético interrumpido o limitado produce diferenciación, estructura, y a grandes escalas temporales la formación de distintos linajes (Riddle y Hafner, 2006).

La interrupción del flujo genético puede ocurrir como consecuencia de barreras geográficas o ambientales (Emerson et al., 2001). La aparición histórica de barreras causadas por diversos eventos geológicos, así como por las oscilaciones climáticas ocurridas durante los últimos periodos glaciales e interglaciales, han provocado discontinuidades espaciales temporales o permanentes (Kumar y Kumar, 2018), dando lugar a la fragmentación de hábitats. Esto repercutió en la distribución geográfica de las especies y provocó divergencia genética en sus poblaciones (Hewitt, 2000; 2004).

Las especies de plantas con una amplia distribución geográfica son buenos modelos para la evaluación de los efectos de las barreras históricas en la diversidad y estructura genética. Dado su hábito sésil y su dispersión de genes dependiente del flujo de polen y semillas, sus poblaciones están sujetas a cambios demográficos ocasionados por eventos geológicos o climáticos que pudieran fragmentar su hábitat y afectar el comportamiento y demografía de sus dispersores a lo largo de su distribución geográfica, actuando como barreras al flujo y moldeando la distribución de la variación genética (Sork et al., 2016; Sweet et al., 2019).

La península de Baja California es una región con una historia geológica y climática dinámica (Leaché et al., 2007; Garrick et al., 2009; Dolby et al., 2015). La historia geológica de la península se remonta

aproximadamente a los últimos seis millones de años, tiempo en el que se estima ocurrió la separación del continente y la formación del Golfo de California. Posteriormente parece haber sufrido de tres fenómenos geológicos relevantes en la historia natural de sus poblaciones (Brusca, 2015). Dos de ellos habrían ocurrido durante el Plioceno, hace tres millones de años: el primero fue la inundación del Istmo de la Paz, aislando a la región del Cabo del resto de la península y el segundo fue una intrusión marítima al norte del Golfo, separando por completo la península del resto del continente. El tercero tendría lugar hace 1.6 millones de años, cuando la hipotética formación de un canal marítimo en la región del Vizcaíno (ubicada entre los 26° y los 28°), pudo atravesar del Océano Pacífico al Golfo de California, conectando las lagunas de San Ignacio y Ojo de Liebre, y aislando históricamente a las poblaciones distribuidas en éstas áreas (Hafner y Riddle, 2011). Por su parte, la evidencia de la historia climática ha demostrado que las oscilaciones ocurridas desde el Pleistoceno tardío (hace ∼140,000 años) moldearon la distribución de los hábitats peninsulares provocando contracciones y expansiones en las poblaciones naturales (Grismer, 2002; Holmgren et al., 2011).

Específicamente en la región del Vizcaíno, la existencia de una intrusión marítima como un evento vicariante se considera la explicación más parsimoniosa para explicar la presencia de linajes de diversas especies de reptiles, mamíferos, aves y artrópodos que presentan una distribución alopátrica, y cuya divergencia es consistente con el tiempo hipotetizado para la existencia de dicha barrera (Upton y Murphy 1997; Zink et al., 1997; Riddle et al., 2000; Rodríguez-Robles y De Jesús-Escobar, 2000; Murphy y Aguirre-León, 2002). Sin embargo, la ausencia de sedimentos marinos a lo largo de la región ha provocado que ésta hipótesis se mantenga cuestionada (Grismer, 2002; Hafner y Riddle, 2011). De forma paralela, diversas especies de plantas (Clark-Tapia y Molina-Freaner, 2003; Fehlberg y Ranker, 2009; Rebernig et al., 2010; Lira-Noriega et al., 2015; Gutiérrez-Flores et al., 2016; Klimova et al., 2017; Martínez-Noguez et al., 2020), reptiles (González-Rubio et al., 2016; Harrington et al., 2018), artrópodos (González-Trujillo et al., 2016) y mamíferos (Latch et al., 2009) muestran evidencia de haber experimentado divergencia y formación de linajes intraespecíficos como consecuencia de los cambios demográficos provocados por la fragmentación de hábitats y la formación de refugios glaciales e interglaciales, así como zonas de expansión y contacto secundarias (Hewitt 2000, 2001). En conjunto, los patrones de distribución de la variación genética en la península sugieren que la historia demográfica, estructura y diversidad genética de las especies presentes en esta región podrían estar mediadas por ambos factores históricos, el geológico y el climático (Dolby et al., 2015).

Un género representativo de Norteamérica y de la península de Baja California es *Yucca* (Asparagaceae), un taxón de plantas arbóreas monocotiledóneas, longevas y semiperennes con tallos leñosos, hojas verdes

suculentas lanceoladas e inflorescencias blancas en forma de roseta. Se caracteriza por un ciclo de vida prolongado y tasas de crecimiento y reclutamiento bajas (Wiggins., 1980; Turner et al., 1995). Su dispersión por polen es a cortas distancias y depende de polinizadores obligados de polillas pertenecientes a los géneros *Tegeticula* y *Parategeticula* (Pellmyr et al., 2008), mientras que la dispersión por semillas parece haber estado relacionada con la megafauna norteamericana ahora extinta, y actualmente pudiera estar mediada por roedores, aves y rumiantes (Pellmyr y Leebens-Mack, 1999; Pellmyr, 2003; Vander et al., 2006).

Existen tres especies registradas del género *Yucca* a lo largo de la península: *Y. schidigera, Y*. *capensis* y *Y. valida*, siendo estas últimas dos endémicas y consideradas como especies hermanas (Pellmyr et al., 2007), además de ser polinizadas por una única especie de polilla, *Tegeticula baja*. Mientras que *Y. schidigera* se distribuye al norte, de los 36°N a los 29.5°N y *Y. capensis* al sur, de los 24°N a los 23°N; las poblaciones de *Y. valida* se extienden desde los 29.8°N hasta los 25.6°N, atravesando el Desierto Central, el Desierto del Vizcaíno y las Planicies de Magdalena (Turner et al., 1995; Thiede, 2020).

La presencia de *Y. valida* es dominante y continua a lo largo de su distribución norteña, mientras que parece estar fragmentada entre los 26°N y los 27°N (Figura 1). Sus poblaciones atraviesan las regiones geográficas en las que se han observado divergencias genéticas en diversas especies, que han sido asociadas a 1) la hipotética intrusión marina de la región del Vizcaíno y 2) los ciclos climáticos ocurridos durante los últimos periodos glaciales (Upton y Murphy, 1997; Grismer, 2002; Riddle et al., 2006). Esto sugiere que la especie también pudo haber experimentado fragmentación con una subsecuente reducción en las tasas de flujo genético, resultando en una divergencia histórica. Conocer la edad de origen de *Y. valida* permitiría asociar su historia demográfica con la historia geológica y climática de la península, permitiendo establecer el intervalo de tiempo en el que se espera que sus poblaciones pudieran haber sufrido divergencia histórica. Mientras que el origen del género *Yucca* se ha estimado en aproximadamente seis millones de años (Smith et al., 2008), se desconoce cuándo ocurrió la colonización de la península por parte de las especies endémicas de *Yucca* que actualmente se distribuyen en ella y si este proceso ocurrió por la vía de dispersión o vicarianza. Con el fin de identificar cuantos linajes genéticos hay a lo largo de la distribución de *Y. valida*, reconstruir su historia demográfica e interpretarla en el contexto de la historia geológica y climática de la península, se evaluó la distribución de la variación genética de la especie por medio de marcadores genómicos a lo largo de toda su distribución geográfica.

1.1 Hipótesis

La incursión marina que tuvo lugar en el centro de la península de Baja California durante el Pleistoceno y las fluctuaciones climáticas ocurridas durante los últimos periodos glaciales actuaron como barreras que moldearon la distribución de la diversidad genética de *Y. valida*.

Predicción: Los límites de la distribución genética de los linajes de *Y. valida* coincidirán con las barreras históricas y los tiempos de divergencia de los linajes de *Y. valida* serán congruentes con los periodos en los que ocurrieron.

1.2 Objetivos

1.2.1 Objetivo general

- Evaluar la distribución de la diversidad genética neutral de *Y. valida*.
- Reconstruir la historia demográfica de *Y. valida*.

1.2.2. Objetivos específicos

- Determinar cuántos linajes genéticos hay a lo largo de la distribución de *Y. valida*.
- Determinar las relaciones genealógicas de los linajes *Y. valida*.
- Estimar el tiempo de divergencia de *Y. valida* como especie.

Capítulo 2. Metodología

2.1 Colecta de muestras y extracción de ADN

Para el presente estudio se colectaron muestras de tejido de hoja de 160 individuos en 20 localidades a lo largo de la distribución de la especie (Figura 1, Tabla 1), abarcando una extensión de alrededor de 450 kilómetros desde los 29.8°N hasta los 25.6°N. El número de muestras promedio por localidad fue de 7, con un mínimo de 2 y un máximo de 14. Las muestras fueron preservadas en gel de silicato desecante.

Tabla 1. Coordenadas y tamaño de muestra colectada a lo largo de la distribución de la especie *Y. valida*.

Localidad	N	Latitud	Longitud
100	3	29.36467	-114.36412
16	6	29.25202	-114.16822
15	6	29.11758	-114.14969
26	5	29.05869	-114.00852
105	10	28.83702	-114.12524
104	4	28.65867	-114.03167
103	10	28.40771	-113.99003
14	12	28.20819	-114.00066
1	12	27.73669	-113.38708
13	7	27.31694	-112.79994
7	9	27.31363	-113.12936
8	7	27.30480	-113.07413
9	5	27.2925	-112.98147
4	12	27.29052	-113.15716
6	8	27.24591	-113.21036
10	2	27.23780	-112.86741
102	8	27.15399	-112.88222
101	12	27.06119	-112.96024
90	7	26.000366	-112.17968
91	15	25.936442	-112.09022

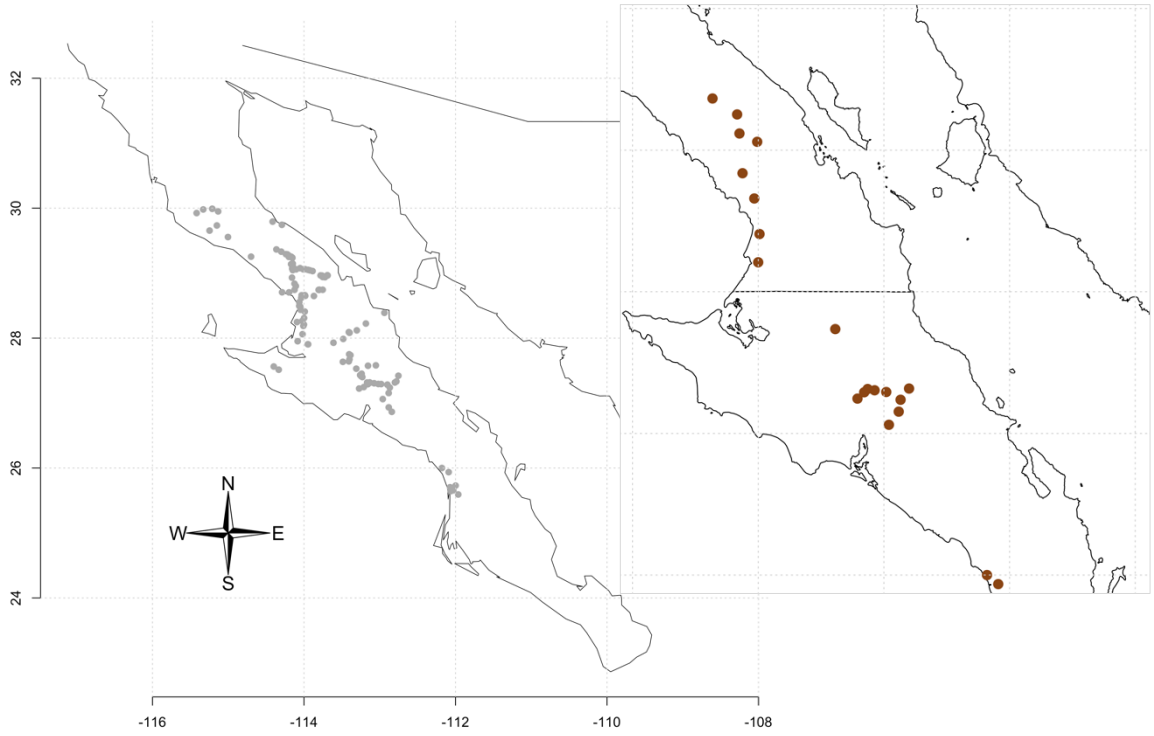

Figura 1. Izquierda: Registros georreferenciados de la distribución actual de *Y. valida* tomados del Sistema Nacional de Información sobre Biodiversidad (SNIB) Derecha: Ubicación de las 20 localidades de muestreo de *Y. valida*.

Para la extracción de ADN se utilizaron 100 mg de cada muestra desecada siguiendo el protocolo del equipo comercial Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Alemania). La evaluación de la concentración y calidad de las extracciones se llevó a cabo por medio de geles de agarosa GelRed al 1.5% y un espectrofotómetro NanoDrop (Thermo Fisher Scientific, Waltham, MA, Estados Unidos).

2.2 Preparación de bibliotecas genómicas

El ADN genómico se compiló en bibliotecas de genotipado por secuenciación de nextRAD (SNPsaurus, LLC) (Russello et al., 2015). Posteriormente, se fragmentó con reactivo Nextera (Illumina, Inc.) ligando secuencias adaptadoras cortas a los extremos de los fragmentos. La reacción de Nextera se escaló para fragmentar 20 ng de ADN genómico y se utilizaron 50 ng de ADN genómico para compensar la cantidad de ADN degradado en las muestras y para aumentar el tamaño de los fragmentos. El ADN fragmentado se amplificó durante 27 ciclos a 74°C. Las bibliotecas nextRAD se secuenciaron en un HiSeq4000 con una línea de lecturas (single-end) de 150 pb (Universidad de Oregon).

2.3 Procesamiento de lecturas crudas, obtención y filtrado de sitios variantes

La evaluación de la calidad de las lecturas demultiplexadas se llevó a cabo por medio de FastQC (Andrews, 2010) y MultiQC. La limpieza por calidad y la remoción de adaptadores Nextera se realizó a través de la herramienta bbduk (BBTools) (Bushnell et al., 2017) conservando solamente fragmentos de 100 pb con un nivel Phred igual o mayor a 30. Considerando que se obtuvo un promedio de dos millones de lecturas por muestra y con la finalidad de optimizar el ensamblado de lecturas, se removieron las muestras que poseían menos de un millón de lecturas después de la limpieza por calidad.

El ensamblado de lecturas se realizó por medio de un mapa *de novo* siguiendo el flujo de trabajo de Stacks 2.41 (Catchen et al., 2013). El mínimo de lecturas idénticas para que una secuencia se tomara en cuenta fue de 8 (*-m 8*). Se estableció un máximo de tres bases de diferencia (*missmatches*) entre una lectura y otra para ser considerada la misma secuencia (conocida como *stack*) dentro de una misma muestra (*-M 3*), y un máximo de cuatro bases de diferencia entre el *stack* de una muestra y el de otra para ser considerado el mismo locus putativo (*-n 4*). Se mantuvieron sólo aquellas muestras con una profundidad promedio ≤15x. La calibración de estos parámetros se llevó a cabo siguiendo las recomendaciones de Paris y colaboradores (2017).

El filtrado del listado de secuencias se realizó en PLINK 1.07 (Purcell et al., 2007) considerando solamente aquellas con una frecuencia del alelo menor (MAF) por encima de 0.05 y en equilibrio de Hardy-Weinberg con una $p < 0.001$. Se conservaron los loci presentes en al menos el 80% de la población y las muestras con un máximo de datos faltantes (*missing data*) del 20% (Tabla 2, Figura 2). La obtención del listado final de secuencias se realizó a través de Populations (Stacks 2.41) seleccionando solamente un polimorfismo de un solo nucleótido (conocido como SNP) por locus de manera aleatoria a través de la función *write_random_snp*. Se generaron los archivos tipo Genepop y VCF y la conversión de formatos requerida para los posteriores análisis se realizó por medio de PGDSpider 2.1.1.5 (Lischer y Excoffier 2012).

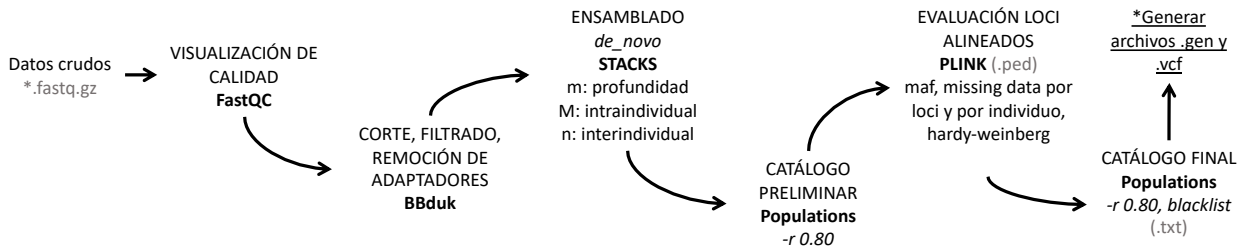

Figura 2. Flujo de trabajo para la obtención de SNPs de *Y. valida*.

Tabla 2. Parámetros del flujo de trabajo para la obtención de SNPs de *Y. valida*.

Parámetro	Valores
Calidad Phred	≤30
Tipo de adaptadores	Nextera
Tamaño inicial de fragmentos	150 pb
Tamaño final de fragmentos	100 pb
Profundidad mínima por secuencia	8x
Missmatches	3 pb
Distancia entre stacks	4 pb
Profundidad promedio por muestra	≤15x
MAF	<0.05
Valor p de equilibrio H-W	<0.001
Máx. missing data por loci/muestra	20%

2.4 Identificación de SNPs potencialmente bajo selección

Puesto que el interés del presente estudio era determinar la estructura genética neutral, la cual es resultado de procesos de la deriva genética y el flujo, se requirió la identificación y exclusión de loci con frecuencias fuera de lo esperado al azar, conocidos como atípicos u outliers. Mientras que en los sitios neutrales la distribución de las frecuencias alélicas son una consecuencia de las mutaciones, el flujo y la deriva genética, los sitios atípicos presentan una variación que puede ser derivada de algún tipo de selección.

Tomando en cuenta lo anterior, se realizaron pruebas de identificación de loci outliers. Los análisis se realizaron usando tres métodos de detección. Primero se llevó a cabo una búsqueda de alelos con índices de fijación distintos a lo esperado según la neutralidad, por medio del análisis frecuentista de la función Fdist2 vía Arlequin 3.5 (Excoffier y Lischer 2010) que evalúa la relación entre el índice F_{ST} global y la heterocigosidad para cada locus. La segunda prueba fue un análisis bayesiano a través de Bayescan 2.1 (Foll y Gaggiotti, 2008), considerando las frecuencias alélicas de cada sitio muestreado basados en un modelo jerárquico de islas siguiendo los parámetros preestablecidos por el programa, es decir, realizando 20 corridas de 5,000 iteraciones con un *burn-in* de 50,000 pasos y un *thinning factor* (factor de adelgazamiento) igual a 10. El resultado del análisis fue evaluado bajo un factor FDR de 0.25. La tercera prueba, basada en la estructura de los componentes principales de la totalidad de las muestras (número de clústeres o agrupaciones genéticas) se realizó por medio de PCAdapt 4.0.1 (Luu, Bazin, y Blum, 2017).

Se clasificaron como atípicos aquellos loci identificados en al menos dos de los métodos utilizados. Los resultados de la identificación de sitios outliers llevaron a la formación de dos catálogos: sitios neutrales y sitios potencialmente bajo selección. Con la finalidad de determinar las funciones de los sitios potencialmente bajo selección, se realizó una búsqueda mediante la anotación del catálogo de sitios potencialmente bajo selección por medio de BLAST+ (Camacho et al., 2009) a través del servidor remoto del NCBI basándonos en un criterio de similitud del 80% con toda la base de datos de nucleótidos existente.

2.5 Análisis de estructura y diversidad genética

La evaluación de la estructura y diversidad genética de *Y. valida* se realizó a partir del catálogo de 1,440 sitios neutrales identificados en 140 muestras tras los métodos previamente descritos. Para evaluar la estructura genética se ejecutó un análisis de los componentes principales (PCA) por medio del paquete Adegenet en R (R Core Team, 2016) y un análisis de asignación por medio de Structure 2.3.4 (Pritchard et al., 2000) y ADMIXTURE 1.3.0 (Alexander y Lange, 2011). El análisis en Structure se implementó utilizando el script StrAuto v1.0 con un *burn-in* de 300,000 pasos y 300,000 iteraciones post *burn-in*. Se probó un número de K entre 1-20 y se ejecutaron 10 iteraciones por cada valor de K. El número óptimo de linajes genéticos (K) se determinó utilizando el método de Evanno y colaboradores (2005) en STRUCTURE HARVESTER 0.6.94. El análisis en Admixture se realizó con 10 iteraciones para cada K=1-20, identificando el número óptimo de linajes utilizando el enfoque de la prueba de validación cruzada para cada *K* evaluada (Alexander y Lange, 2011).

Adicionalmente, se convirtieron los datos de los 1,440 SNPs neutrales en secuencias lineales para cada muestra en formato Phylip según los códigos de ácidos nucleicos de la IUPAC y se generó una red por medio del algoritmo NEIGHBORNET dentro de SPLITSTREE 5 (Huson y Bryant, 2006). Todos los análisis de estructura fueron concordantes en sus resultados y puesto que los análisis de asignación mostraron mezcla de componentes en ciertas muestras, la pertenencia a cada linaje observado (tres) se baso en la diferenciación observada por medio del análisis de componentes principales. Posteriormente, para cada linaje identificado (ver resultados: norte, centro y sur) así como para el total de la muestra, se calcularon los parámetros descriptores de la diversidad genética (H_E, H_O y F_{IS}) por medio del paquete Adegenet en R (Jombart, 2008). Se realizaron pruebas de varianza molecular (AMOVA) y una estimación de F_{ST} pareada entre los distintos linajes utilizando el programa GENODIVE 3.04 (Meirmans, 2020).

Finalmente, para evaluar la correlación entre la distancia geográfica y genética entre las localidades muestreadas, se efectuó una prueba de Mantel utilizando la distancia genética estandarizada de Rousset ($F_{ST}/1-F_{ST}$) y la distancia geográfica lineal (km) entre cada localidad. La significancia de la prueba de Mantel se evaluó por 99,999 permutaciones. Con el fin de visualizar la existencia de un patrón de aislamiento por distancia se trazaron las distancias genéticas y geográficas por localidad en un gráfico de densidad de kernel bidimensional utilizando la función kde2d en el paquete MASS 7.3 en R.

2.6 Análisis filogenéticos

Para conocer las relaciones genealógicas entre los individuos muestreados y los linajes identificados, se construyó un árbol de máxima verosimilitud utilizando PhyML 3.0 (Guindon et al., 2010). El grupo externo utilizado fue *Y. capensis*, especie hermana, de la que se agregaron dos secuencias. La obtención de SNPs para este análisis se llevó a cabo siguiendo los métodos del flujo de trabajo del procesamiento de lecturas crudas, obtención y filtrado de sitios variantes y las pruebas de remoción de sitios atípicos u outliers antes descritos. El análisis filogenético se realizó a partir de una base de datos de 1,230 SNPs neutrales en los 140 individuos de *Y. valida* y los dos individuos de *Y. capensis*. La selección del modelo de sustitución se realizó por medio del criterio de Akaike, según las recomendaciones de Lefort y colaboradores (2017). El modelo de sustitución utilizado fue el reversible de tiempo general asumiendo una distribución gamma con cuatro categorías de tasas de sustitución y estimando la proporción de sitios invariantes (GTR + G + I).

2.7 Historia demográfica

Para comprender la sucesión de eventos que pudieron originar los linajes identificados, se realizó una reconstrucción de la historia demográfica de *Y. valida* a través de un enfoque computacional bayesiano aproximado (ABC) por medio de DIYABC v2.1.0 (Cornuet et al., 2014). Se realizaron 100,000 simulaciones exploratorias para cada uno de tres escenarios propuestos (Figura 3, A-C) con una distribución previa uniforme de los tiempos de divergencia y de los tamaños efectivos de población para cada linaje. Con base en los valores de distribución posteriores, se optimizaron los valores previos para realizar 1,000,000 de simulaciones bajo un modelo de distribución normal para el tiempo de divergencia y los tamaños efectivos poblacionales en escenario planteado (Csilléry et al., 2010; Wilkinson, 2013). La selección del modelo se realizó generando conjuntos de datos pseudo-observados mediante el muestreo y reemplazo de cada modelo específico y los valores de los parámetros asociados a partir de las 1,000 simulaciones más cercanas a los datos observados. Para cada conjunto de datos pseudo-observados se calculó la probabilidad posterior evaluando la proporción de veces que el modelo tuvo la probabilidad más alta.

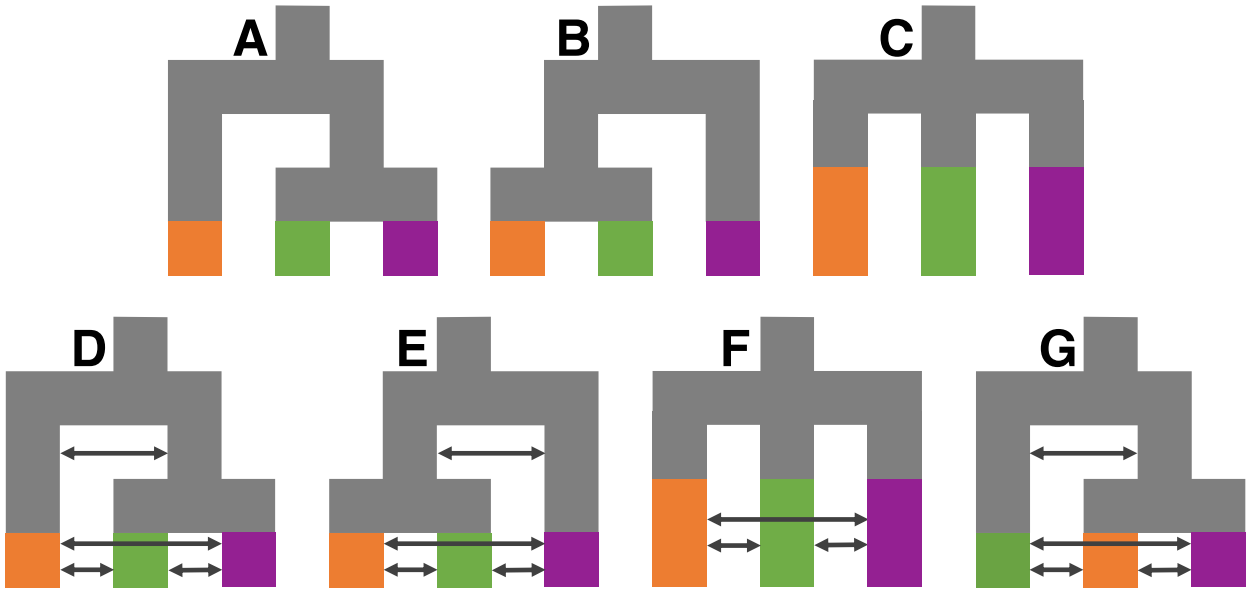

Figura 3. Escenarios demográficos simulados para determinar la secuencia de origen de los linajes de *Y. valida*. Cada color representa los linajes identificados por los análisis de estructura (norte en naranja, centro en verde y sur en morado). Las flechas representan la simulación y estimación de las tasas de migración.

Además, se evaluaron modelos demográficos alternativos que incluyeron la estimación de la migración a partir del espectro de frecuencias alélicas (SFS) de los SNPs neutrales por medio del método de probabilidad compuesta implementado en FASTSIMCOAL2.6 (Excoffier et al., 2013). Se compararon cuatro modelos demográficos distintos con tres linajes correspondientes al norte, centro y sur. Estos incluyeron

todos los escenarios de divergencia posibles para los tres linajes así como la divergencia simultánea. Cada escenario fue simulado considerando el flujo histórico de genes, representado como la proporción de una población compuesta por migrantes (*m*) de otra población (Figura 3, D-G). Dada la necesidad de una tasa de mutación, se utilizó el valor de $1.4x10^{-7}$ mutaciones/sitio/generación estimado para SNPs de *Y. brevifolia* (Smith et al., datos no publicados). Se realizaron 50 réplicas para cada modelo con 100,000 simulaciones para cada réplica. La selección del escenario más probable se realizó con base en criterio de Akaike.

2.8 Obtención de fragmentos de cloroplasto

Dado que la extracción de ADN total contiene fragmentos nucleares y de plástidos, se implementó un flujo de trabajo para explorar el potencial del método nextRAD para recuperar secuencias de cloroplastos. Se utilizaron las 160 muestras genotipificadas de *Y. valida*, así como 40 muestras de *Y. capensis*. Las muestras se filtraron por calidad y limpiaron de adaptadores Nextera por medio de FastQC, MultiQC y la herramienta bbduk (BBTools), conservando solamente fragmentos de 100 pb con un nivel Phred igual o mayor a 30.

Se realizó una normalización de los fragmentos con la herramienta bbnorm (BBTools) para descartar las lecturas con una profundidad menor a 10x. Debido a que no existe un genoma de referencia de *Y. valida*, las lecturas se mapearon y alinearon teniendo como referencia el genoma de la especie más emparentada, *Yucca schidigera* (Pellmyr et al., 2007), utilizando Bowtie 2.2.6 (Langmead y Salzberg, 2012) y Samtools.

Los sitios sin coincidencias durante el alineamiento fueron sustituidos con bases enmascaradas (“N”) y tratados como datos faltantes (*missing data*). Posteriormente se seleccionaron las muestras con la menor cantidad de datos faltantes por especie (de *Y. valida* y *Y. capensis* de manera independiente), considerando la muestra con mayor número de bases recuperadas como un 0% de datos faltantes. Se conservaron solamente las muestras con un porcentaje de datos faltantes igual o menor al 20% para cada especie.

Se recuperaron 36 secuencias de *Y. valida* con una extensión máxima de 140,823 y mínima de 112,548 bases nitrogenadas por muestra. Los fragmentos recuperados de *Y. capensis* presentaron una alta cantidad de datos faltantes, por lo que solamente se conservó una secuencia con 41,801 bases nitrogenadas repartidas a lo largo del genoma mapeado. Las diferencias entre la información recuperada para cada especie parece estar asociada al proceso de secuenciación de nextRAD (SNPsaurus, LLC) *per se*, pues las muestras de *Y. valida* fueron generadas en dos celdas de flujo distintas mientras que las muestras de *Y.*

capensis compartieron una celda de flujo con muestras de *Y. valida*. La información proveniente de esta última celda mostró de manera general, un menor porcentaje de recuperación de fragmentos asociados al cloroplasto, de tal forma que las muestras de *Y. valida* de esta celda tuvieron un porcentaje de datos faltantes mayor al 20% con respecto a la muestra con mayor cantidad de bases recuperadas en la celda individual, por lo que no se recuperó ninguna secuencia de *Y. valida* proveniente de la celda compartida. Las 36 secuencias recuperadas de *Y. valida* provenían de individuos de 15 localidades, de las cuales, 17 individuos fueron asignados al linaje nuclear del norte, 14 al del centro y 5 al del Sur.

2.9 Relaciones filogenéticas y tiempo de divergencia de *Y. valida*

Las secuencias potencialmente provenientes de cloroplasto de ambas especies se alinearon por medio de ClustalW (Sievers et al., 2011) con los genomas de referencia de los cloroplastos de 15 especies de la subfamilia *Agavoideae* (*Agave attenuata, Beschorneria septentrionalis, Camassia scilloides, Chlorogalum pomeridianum, Hesperaloe campanulata, Hesperaloe parviflora, Hesperocallis undulata, Hesperoyucca whipplei, Hosta ventricosa, Manfreda virginica, Schoenolirion croceum, Yucca brevifolia, Yucca filamentosa, Yucca queretaroensis* y *Yucca schidigera*) presentes en la base de datos de NCBI. El alineamiento de las 36 muestras de *Y. valida* y la muestra de *Y. capensis* con los genomas de referencia de los cloroplastos de la subfamilia *Agavoideae* dio como resultado una matriz de 161,262 bases.

Para los pasos subsiguientes, los sitios con bases enmascaradas (“N”) y huecos (conocidos como gaps, representados con el símbolo “-”) resultado del proceso de alineamiento, fueron considerados como datos faltantes (*missing data*). Se ha demostrado que el peso relativo de los datos faltantes en la resolución de filogenias puede ser contrarrestado por los métodos basados en coalescencia que son implementados en los análisis de máxima verosimilitud y bayesianos, permitiendo obtener inferencias robustas de conjuntos de secuencias incompletas (Burleigh et al., 2009; Zhenxiang et al., 2016). A partir de estas premisas, se reconstruyeron las relaciones filogenéticas dentro de la subfamilia *Agavoideae* con el fin de realizar estimaciones del tiempo de diversificación de *Y. valida* y *Y. capensis* dentro del género *Yucca* basados en un modelo de reloj molecular (Zuckerkandl y Pauling, 1965).

Previo a los análisis filogenéticos, el alineamiento de las 15 especies de *Agavoideae*, las 36 muestras de *Y. valida* y la muestra de *Y. capensis* fue evaluado por medio jModelTest2 (Darriba et al., 2012) para identificar el modelo de sustitución más adecuado. El modelo elegido fue reversible de tiempo general

asumiendo una distribución gamma con cuatro categorías de tasas de sustitución y estimando la proporción de sitios invariantes (GTR + G + I). Se utilizó una tasa de mutación fija bajo un modelo de reloj estricto como parámetro de calibración, basada en la estimación de Smith y colaboradores (no publicado) para los genomas de los cloroplastos de *Agavoideae* ($\mu=3x10^{-4}$, $\sigma=3x10^{-5}$ sustituciones por sitio por cada millón de años).

El método de máxima verosimilitud se implementó a través de RelTime-CC por medio de MEGA X (Tamura et al., 2012; Kumar et al., 2018) con 100 iteraciones a partir de un árbol inicial de máxima parsimonia. Con el árbol resultante se identificaron las medias y los intervalos de confianza (95%) de las edades estimadas para cada nodo. El método bayesiano fue implementado a través de BEAST v1.10.4 con la ayuda de BEAUti v1.10.4 (Drummond y Rambaut, 2007; Drummond et al., 2012). Asumiendo un tamaño poblacional constante, se establecieron distribuciones gamma previas para cada tipo de sustitución y previas uniformes (0 a 1) para las frecuencias de las bases y la proporción de los sitios invariables. El grupo externo fue *Hosta ventricosa*, forzando la monofilia de las demás especies, siguiendo las recomendaciones de Drummond y Bouckaert (2015). Se realizaron dos corridas independientes, cada una de $1x10^7$ generaciones en el servidor de CIPRES Science Gateway (Miller et al., 2010), con un muestreo cada 10,000 generaciones. Se utilizó Tracer v1.7.1 para evaluar la convergencia de las MCMC y se construyó un árbol de máxima credibilidad utilizando TreeAnnotator v1.10.4 considerando un burn-in del 10% con las medianas y los intervalos de confianza (95%) de las edades estimadas para cada nodo.

Capítulo 3. Resultados

3.1 Procesamiento de lecturas crudas, obtención y filtrado de sitios variantes

Se procesaron 400 millones de secuencias de 150 pb provenientes de 160 muestras (2.4 millones de secuencias promedio por muestra). Tras el filtrado de calidad y la limpieza de adaptadores, se conservaron 317 millones de secuencias (2 millones de secuencias promedio por muestra) de 100 pb distribuidas en 154 muestras (Figura 4). El ensamble *de novo* en Stacks recuperó 3,439 loci. Después de los filtros de profundidad por muestra y de MAF y HW por locus se obtuvieron 3,159 loci, de los cuales 1,460 fueron polimórficos. El catálogo final de secuencias consistió de 1,460 sitios bialélicos en 140 muestras.

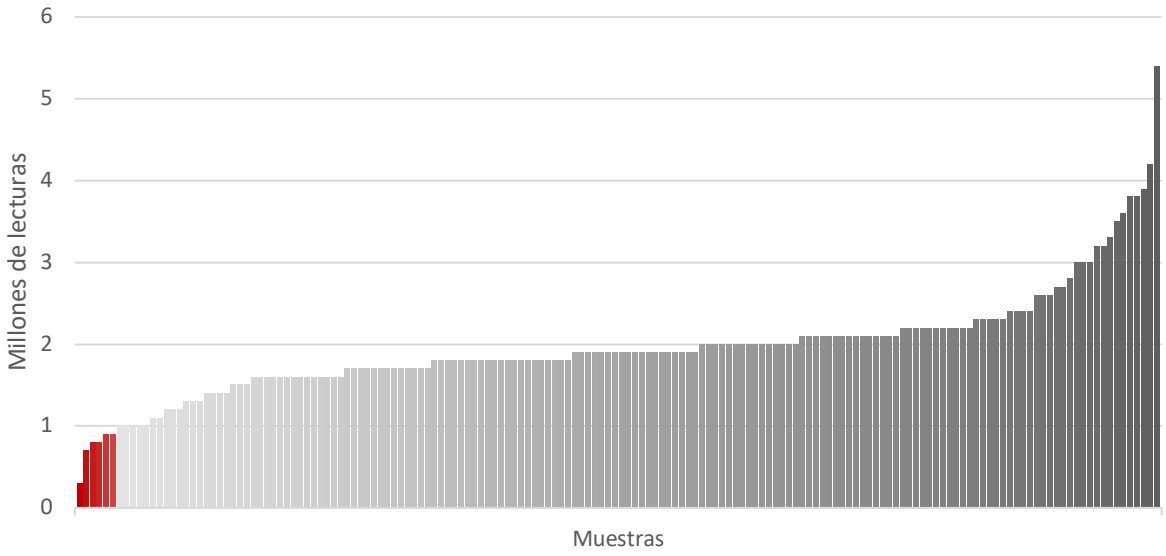

Figura 4. Distribución de lecturas por muestra de las 160 muestras de *Y. valida* tras la limpieza por calidad y la remoción de adaptadores. Las barras en rojo indican las muestras que no superaron el número mínimo de lecturas (un millón) para ser utilizadas en los subsiguientes análisis.

3.2 Identificación de SNPs potencialmente bajo selección

Se identificaron 20 loci atípicos con al menos dos de los tres métodos de análisis. Se encontraron 49 sitios atípicos por medio de Arlequin 3.5, 13 por medio de Bayescan 2.1 y ocho por medio de PCAdapt 4.0.1. La identificación de SNPs outliers a través función Fdist2 vía Arlequin 3.5 fue la única que identificó 29 sitios que no concordaron con los otros dos métodos de exploración. A partir de estos resultados, se construyó

un catálogo de 1,440 sitios neutrales y un catálogo 20 sitios potencialmente bajo selección. La anotación de secuencias en BLAST+ no identificó genes coincidentes para los loci potencialmente bajo selección.

3.3 Análisis de estructura y diversidad genética

Los análisis de estructura se realizaron para el catálogo de 1,440 SNPs neutrales en las 140 muestras de *Y. valida*. Los análisis en Structure y Admixture coincidieron en que tres linajes (K=3) proporcionan el mejor ajuste a los datos. Ambos análisis mostraron un único componente en la mayoría de los individuos y mezcla de componentes en las localidades 102 y 101 (Figura 5). La asignación de los individuos a cada linaje se basó en el mayor porcentaje de un componente, y para aquellos individuos con alto nivel de mezcla, se basó en su posición en el gráfico del análisis de componentes principales (Figura 6). La red de agrupaciones a nivel individual realizada por medio de SPLITSTREE (Figura 7) mostró congruencia con los resultados del PCA y de los análisis de asignación. Los linajes identificados se nombraron según su distribución geográfica como norte, centro y sur (Figura 8).

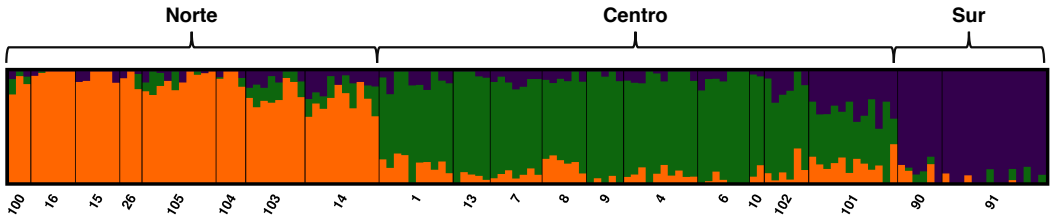

Figura 5. Resultados de ADMIXTURE para el valor óptimo (*K=3*) de la estructura genética de *Y. valida*. Cada barra vertical representa la composición genética por individuo. Cada color representa un componente genético. Los números representan la localidad geográfica a la que pertenece cada individuo, de mayor a menor latitud.

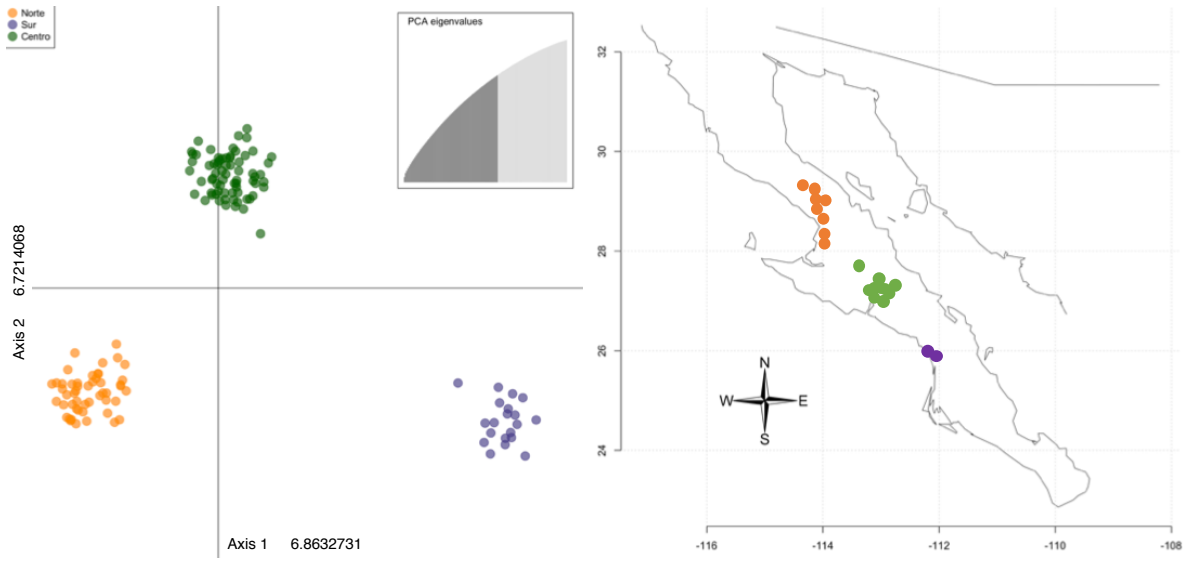

Figura 6. Izquierda: Análisis de componentes principales de la estructura genética de *Y. valida*. El panel superior derecho representa la varianza acumulativa de los componentes principales utilizados (80) para este análisis. Derecha: Mapa de distribución de los linajes genéticos inferidos como resultado de los análisis de estructura genética neutral de *Y. valida*. Cada color representa un linaje asignado *a posteriori*.

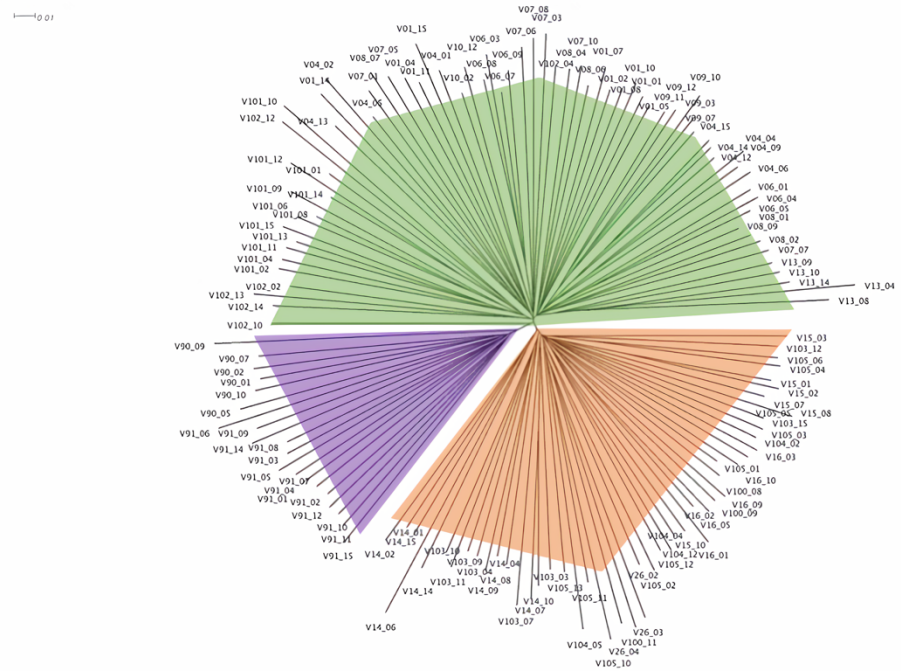

Figura 7. Red (NEIGHBORNET) de las agrupaciones a nivel individual de las muestras de *Y. valida.* Cada color representa un linaje asignado *a posteriori*.

Los tres linajes genéticos encontrados se diferenciaron significativamente en los análisis de F_{ST} y AMOVA que se realizaron con el conjunto de 1,440 SNPs neutrales. Se observó una mayor estructuración entre los linajes norte y sur que entre los linajes contiguos (Tabla 3) . Además se encontró que el 3.4% de la varianza es explicada por la diferenciación entre los tres linajes de manera significativa (Tabla 4).

Tabla 3. Valores de F_{ST} pareada entre los linajes inferidas a partir de los análisis de estructura genética de 1,440 SNPs neutrales de *Y. valida.*

FST	NORTE	CENTRO	SUR
NORTE	-	-	-
CENTRO	0.02799*	-	-
SUR	0.05510*	0.04769*	-

*valores estadísticamente significativos (*p=0.001*)

Tabla 4. Análisis de varianza molecular AMOVA entre los linajes inferidos a partir de los análisis de estructura genética de 1,440 SNPs neutrales de *Y. valida.*

FUENTE DE VARIACIÓN	GL	SC	COMPONENTE DE VARIACIÓN	PORCENTAJE DE VARIACIÓN
ENTRE LINAJES	2	730.28	2.77	3.4
ENTRE LOCALIDADES DENTRO DE LOS LINAJES	17	1973.03	3.03	3.72
DENTRO DE LOS LINAJES	260	19692.98	75.74	92.88
TOTAL	279	22396.3	81.54	100.00

La prueba de Mantel detectó una asociación positiva y significativa entre las distancias geográficas y la diferenciación genética entre las localidades muestreadas (p=0.0001, r=0.44, y=0.0001x+0.055), sugiriendo un patrón de aislamiento por distancia (Figura 8).

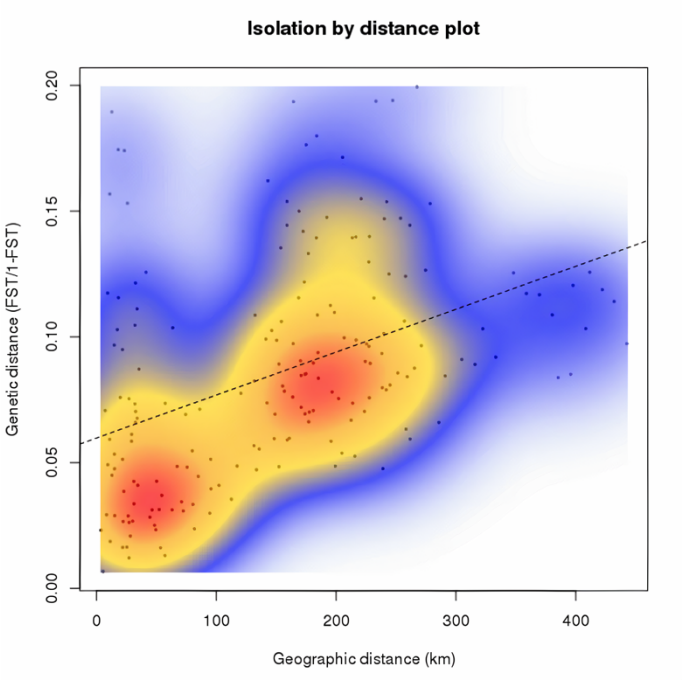

Figura 8. Asociación entre las distancias geográficas (km) y genéticas de Rousset (F_{ST}/1-F_{ST}) entre los individuos de *Y. valida.*

Los análisis de diversidad genética encontraron una heterocigosidad observada global (H_O) de 0.2991, una heterocigosidad esperada (H_E) de 0.3073 y un coeficiente de endogamia (F_{IS}) de 0.0600. La heterocigosidad observada (H_O) osciló en los linajes entre 0.2991 y 0.3051, y fue menor a la heterocigosidad esperada (0.3043 a 0.3248), sin embargo, esta diferenciación no resultó significativa (Tabla 5).

Tabla 5. Diversidad genética global de *Y. valida* y por linaje a partir de 1,440 SNPs neutrales.

LINAJE	N	H_O	H_E	F_{IS}
TOTAL	140	0.2991	0.3073	0.0589
NORTE	50	0.2929	0.3043	0.0541
CENTRO	70	0.3029	0.3061	0.0546
SUR	20	0.3051	0.3248	0.0680

3.4 Análisis filogenéticos

El árbol genealógico construido fue congruente con los análisis de estructura, distinguiendo tres linajes monofiléticos divergentes (Figura 9).

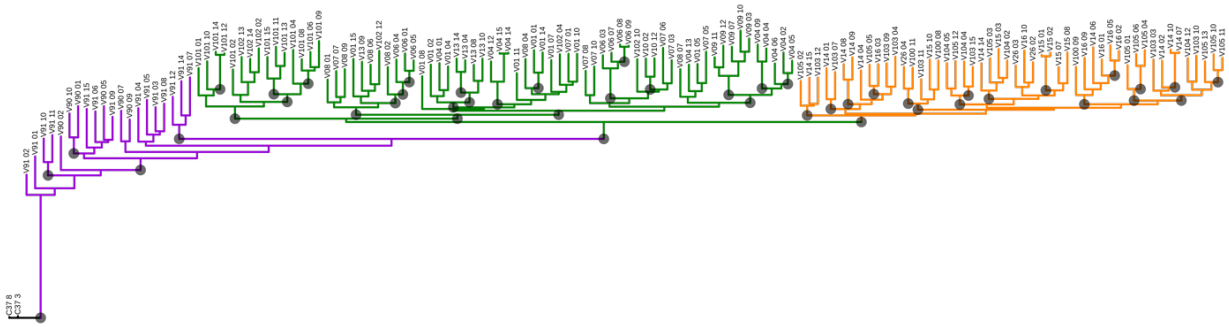

Figura 9. Árbol genealógico de 140 individuos de *Y. valida*. Se utilizaron 2 individuos de *Y. capensis* como grupo externo. El color de las ramas representa los linajes identificados por los análisis de estructura (norte en naranja, centro en verde y sur en morado). Los puntos representan los valores de soporte (*bootstrap*) mayores al 90%.

3.5 Historia demográfica

La simulación bayesiana de escenarios históricos por medio de DIY-ABC y el método de probabilidad compuesta implementado en FASTSIMCOAL2.6 proporcionaron soporte para la divergencia simultánea de los tres linajes (Figura 3, Escenarios C y F). Adicionalmente, FASTSIMCOAL estimó que los linajes están compuestos por migrantes con una proporción de 3.94×10^{-07} entre linajes contiguos y de 6.50×10^{-08} entre los linajes limítrofes. El tamaño poblacional de cada linaje es una variable no conocida, por lo que se debe considerar a estos valores como un índice de relación.

3.6 Relaciones filogenéticas y tiempo de divergencia de *Y. valida*

Las relaciones filogenéticas y edades estimadas para cada especie dentro del género *Yucca* fueron consistentes usando los métodos de inferencia de máxima verosimilitud y bayesiano, con todos los nodos interespecíficos representados con un soporte alto (>90%). Las 36 secuencias de *Y. valida* se agruparon en un único clado, lo que indica un origen monofilético (Figura 8). No se observaron divergencias a nivel intraespecífico para *Y. valida*, a diferencia de lo observado en el árbol de datos nucleares (Figura 9). Dentro del clado *Y. valida* todos los nodos tuvieron un soporte bajo, sugiriendo la ausencia de diferenciación del genoma del cloroplasto dentro de la especie.

El análisis de máxima verosimilitud estimó una edad de 240,000 años (100,000-550,000) para el ancestro común de *Y. valida* y *Y. capensis* y un tiempo de divergencia aproximado de 27,000 años (3,200-26,000) para *Y. valida*, mientras que el análisis bayesiano estimó una edad de 660,000 años (380,000-980,000) para el ancestro común de *Y. valida* y *Y. capensis* y un tiempo de divergencia aproximado de 300,000 años (190,000-420,000) para *Y. valida* (Figura 10).

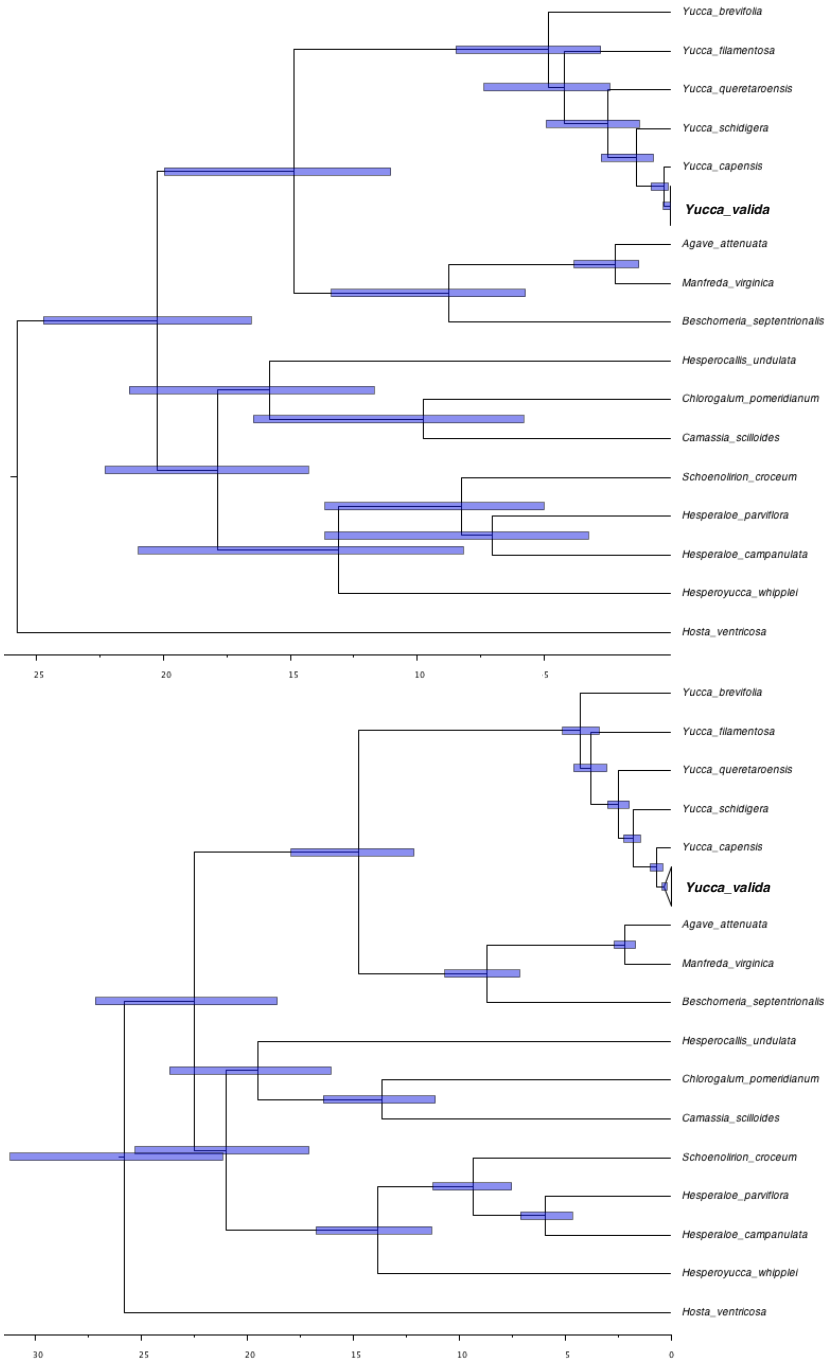

Figura 10. Relaciones filogenéticas e intervalos de confianza del 95% (barras azules) de los tiempos de divergencia estimados por los análisis de máxima verosimilitud (parte superior) y bayesiano (parte inferior) entre las secuencias del genoma del cloroplasto de la subfamilia *Agavoideae* a partir de un modelo de reloj molecular estricto. Las escalas inferiores representan millones de años a partir de la actualidad.

Capítulo 4. Discusión

Los patrones filogeográficos reportados en la península de Baja California sugieren que la distribución de la diversidad genética de las especies en esta región ha sido influenciada por los procesos geológicos que dieron origen a ella, así como por las hipotéticas incursiones marinas durante el Plio-pleistoceno y las oscilaciones climáticas ocurridas durante los pasados ciclos glacial-interglaciales (Dolby et al., 2015; Riddle et al., 2000). En el presente trabajo se evaluó la distribución de la variación genética de *Y. valida*, y a través del análisis de 1,440 SNPs nucleares neutrales, se identificaron tres linajes alopátricos monofiléticos cuya diferenciación tuvo lugar en la región media peninsular de manera simultánea. La estimación del tiempo de divergencia de la especie a partir de secuencias potencialmente provenientes de cloroplasto sostiene que el ancestro común más reciente de *Y. capensis y Y. valida* tiene una edad menor a un millón de años, que los genomas del cloroplasto de *Y. valida* pertenecen a un único linaje y que el origen de la especie habría ocurrido hace aproximadamente 190,000 a 420,000 años. Considerando la edad de la especie estimada, el papel de la hipotética incursión marina hace 1.6 millones de años quedaría descartado de los procesos que moldearon la estructura genética nuclear de la especie.

El origen de los tres linajes alopátricos detectados a lo largo de la distribución de *Y. valida* puede ser consecuencia de dos procesos no excluyentes: la pérdida de hábitat durante los cambios climáticos históricos y un patrón de aislamiento por distancia resultado de un flujo genético reducido. De acuerdo con la distribución potencial de *Y. valida*, durante el último interglacial (∼120,000 años) no existía hábitat disponible para esta especie (Arteaga et al., 2020), hecho que pudo provocar una reducción significativa de sus poblaciones. Diversos estudios de plantas señalan la importancia del papel de las oscilaciones climáticas durante los ciclos glaciales e interglaciales del Cuaternario en la distribución de la diversidad genética en la península de Baja California (Nason et al., 2002; Clark-Tapia y Molina-Fraener, 2003; Fehlberg y Ranker, 2009; Garrick et al., 2009; Klimova et al., 2017). Sin embargo, esta evidencia por sí misma podría no explicar la distribución de los tres linajes nucleares observados. Basados en que a partir del último máximo glacial (∼20,000 años) el hábitat disponible para *Y. valida* se distribuía de manera continua con una amplia extensión en el centro de la península y que por tanto, no existían barreras al flujo, se debe considerar el peso de la distancia geográfica en la diferenciación genética de la especie. Una limitada capacidad de dispersión asociada a la distancia puede llevar a que la diferenciación de las poblaciones vegetales ocurra intrínsecamente (Irwin, 2002).

Los análisis de estructura genética y las genealogías nucleares fueron congruentes con la asignación de los individuos a tres linajes monofiléticos. Dentro de cada linaje no se observaron patrones de divergencia,

por lo que cada una las tres agrupaciones observadas (norte, centro y sur) fungen como poblaciones. Álamo-Herrera (2019) detectó bajos niveles de diferenciación entre localidades a lo largo de la distribución de la especie. En similitud con estos resultados, se observó la existencia de homogeneidad genética a nivel intralinaje. Las estimaciones de las distancias de dispersión indican que el movimiento de polen en *Y. valida* es de aproximadamente 50 m (Álamo-Herrera, 2019), y que la dispersión mediada por semillas también ocurre a cortas distancias (Waitman et al., 2012). En este sentido, el modelo de flujo genético escalonado (stepping-stone) permite mantener la conectividad dentro los linajes. La estructura genética observada podría ser el resultado de una fragmentación histórica reforzada por la acumulación de una dispersión restringida a largas distancias. Aunado a lo anterior, la región geográfica en donde se observó la diferenciación genética entre los linajes coincide con la transición entre el Desierto del Vizcaíno y las Planicies de Magdalena (Minnich et al., 2014; González-Abraham et al., 2010) por lo que la distancia ambiental también podría tener un peso relativo en la distribución de la variación genética.

Se estimó un valor bajo de diferenciación genética a lo largo de la distribución de la especie (F_{ST}=0.037), sin embargo significativo. Este valor es similar a lo identificado anteriormente, siendo menor al reportado en microsatélites para *Y. valida* (F_{ST}=0.059; Álamo-Herrera, 2019), *Y. schidigera* (F_{ST}=0.067; De la Rosa et al., 2020) y *Y. brevifolia* (F_{ST}=0.061; Starr et al.,2013), y mayor a la estimación de Luna (2018) para la especie hermana, *Y. capensis* (F_{ST}=0.022). La baja diferenciación observada puede ser explicada por tamaños efectivos poblacionales históricos grandes que, bajo el modelo de flujo genético de stepping-stone descrito, llevarían a que las limitantes a la dispersión no produzcan mayores efectos en la estructura. Este escenario se sustenta por la amplia distribución potencial presente a partir del último máximo glacial (Arteaga et al., 2020). Por otra parte, los valores estimados en la prueba de Mantel (r=0.44, p=0.0001) fueron semejantes a reportes previos (r =0.45, p=0.02; Álamo-Herrera, 2019), confirmando el efecto de la distancia geográfica sobre la diferenciación genética (Wright, 1943).

El valor más alto de F_{ST} pareada se observó entre los linajes limítrofes (norte-sur=0.055), reflejando la existencia de un mayor flujo histórico entre los linajes vecinos. Este flujo entre linajes vecinos puede explicar la mezcla de componentes genéticos observada en el linaje central, principalmente para los individuos localizados en los límites con los linajes al norte y el sur. Además, sustenta la baja diferenciación observada entre linajes (3.4% de variación). De manera teórica, un migrante por generación sería suficiente para erosionar por completo la diferenciación genética (Wright, 1965), por lo que la baja proporción de migrantes por generación estimada por las simulaciones demográficas ($6.50x10^{-08}$ a $3.94x10^{-07}$) es consistente con las señales de diferenciación genética observadas entre linajes.

Los niveles de diversidad genética observados (H_O=0.2991, H_E=0.3073) son mayores a los descritos anteriormente para esta especie (H_O=0.1119, H_E=0.1736; Arteaga et al., 2020) usando el mismo tipo de marcadores moleculares. Esta disparidad entre los valores de diversidad podría ser causada por las diferencias en el ensamble de lecturas *de novo* en Stacks y el número de individuos de *Y. valida* usados en ambos estudios. El flujo de trabajo de Stacks para la construcción del catálogo de loci busca maximizar el número de SNPs recuperados, compartidos por una cantidad representativa de individuos (Rochette y Catchen, 2017). Arteaga y colaboradores (2020) realizaron un ensamble de loci a partir de muestras de distintas especies (*Y. valida* y *Y. capensis*) así como de poblaciones híbridas en conjunto. Formar un catálogo de loci con muestras de una sola especie puede modificar la matriz de SNPs recuperados y por tanto, las estimaciones de los parámetros descriptores de la diversidad genética (Catchen et al., 2013). Además, el tamaño de muestra en el estudio de Arteaga y colaboradores en relación al número de individuos y localidades (35 individuos en siete localidades de *Y. valida*) es menor al de este estudio (140 individuos en 20 localidades), lo que puede generar diferencias en los estimadores de diversidad. Nuestros resultados (H_O=0.2991, H_E=0.3073) son semejantes a las estimaciones previas realizadas con datos de microsatélites usando un tamaño de muestra similar al nuestro (*Ho*=0.55 *He*=0.69; Álamo-Herrera, 2019), considerando que los SNPs pertenecen a los marcadores bialélicos, de tal forma que el máximo valor teórico de la heterocigosidad es de 0.5 (Singh, et al., 2013). La diversidad genética en esta especie parece estar igualmente distribuida entre los linajes, ya que encontramos niveles de heterocigosis similares y altos (∼0.3). El coeficiente de endogamia presentó niveles bajos (F_{IS}=0.0589), en comparación al registrado previamente (Arteaga et al., 2020, F_{IS}=0.3522). De igual manera, la variación de resultados puede deberse a las diferencias asociadas al ensamble *de novo* y el tamaño de muestra utilizados en ambos trabajos. Los altos niveles de diversidad genética y el bajo nivel de endogamia detectados en este trabajo pueden ser explicados por la posible existencia de tamaños poblacionales históricos grandes y el sistema de reproducción de *Yucca*, que se caracteriza por favorecer el entrecruzamiento y limitar la autopolinización (Pellmyr et al., 1997; Huth y Pellmyr, 2000).

Los resultados de los modelos de demografía histórica basados en simulaciones señalan que la divergencia entre los linajes nucleares está asociada a un solo evento temporal. Aunque no fue posible realizar una estimación de cuándo habría ocurrido este evento, la edad establecida para *Y. valida* (300,000 años, IC95%=190,000-420,000) sugiere que los cambios climáticos del Cuaternario, los cuales moldearon la distribución de la variación genética de diversas especies de plantas en la península (Nason et al., 2002; Clark-Tapia y Molina-Fraener, 2003; Fehlberg y Ranker, 2009; Garrick et al., 2009; Klimova et al., 2017), pudieron provocar una reducción significativa del tamaño poblacional durante el último interglacial, en el que los modelos de distribución potencial señalan la ausencia de hábitat disponible para la especie hace

aproximadamente 120,000 años (Arteaga et al., 2020). Esta fragmentación, aunada a la limitada dispersión de la especie y las diferencias ecológicas entre los desiertos Central, Del Vizcaíno y las Planicies de Magdalena pudieron tener consecuencias en la diferenciación genética al provocar una pérdida de conectividad en las poblaciones. De manera complementaria, no debería descartarse la contribución de la selección natural en la historia demográfica de la especie (Rivas et al., 2015).

En el análisis filogenético a partir de las secuencias potencialmente provenientes del cloroplasto se observó la ausencia de diferenciación intraespecífica en *Y. valida*. Este resultado es consistente con una divergencia nuclear reciente y un genoma del cloroplasto altamente conservado (Palmer et al., 1986). La reconstrucción del tiempo de divergencia de *Y. valida* obtuvo un buen soporte con los análisis de máxima verosimilitud y bayesiano. Las relaciones filogenéticas y las estimaciones de los tiempos de divergencia basadas en un reloj molecular estricto indican que el ancestro común de *Y. valida* y *Y. capensis* tiene una edad aproximada de entre 240,000 (IC 95%=100,000-550,000) y 670,000 (IC 95%=380,000-970,000) años. Las estimaciones de máxima verosimilitud y bayesiana tuvieron una superposición en la edad de divergencia de *Y. valida* y *Y. capensis*, sin embargo, indicaron valores muy diferentes en la edad calculada para *Y. valida*, con 27,000 años (3,200-26,000) y 300,000 años (190,000-420,000) respectivamente. El algoritmo de máxima verosimilitud implementado en RelTime-CC ha demostrado confiabilidad y coincidencia con respecto a las estimaciones bayesianas de BEAST con genomas provenientes de cloroplasto (Haga et al., 2019), además de permitir aproximaciones computacionalmente rápidas (Tamura et al., 2012). Sin embargo, considerando la evidencia de RelTime por subestimar la divergencia intraespecífica (Miura et al., 2020) y que los métodos bayesianos han demostrado mayor exactitud en el manejo de datos faltantes (Burleigh et al., 2009; Zheng y Wiens., 2015; Zhenxiang et al., 2016), se considera que la estimación más robusta se encuentra alrededor de los 300,000 años, indicando que el origen de la especie debió ocurrir en el periodo previo al último interglacial.

Capítulo 5. Conclusiones

Se identificó la existencia de tres linajes nucleares monofiléticos y un único linaje del genoma del cloroplasto en *Y. valida*. Nuestros resultados sugieren que los patrones de distribución de la diversidad genética de *Y. valida* tienen un origen reciente y que la estructura genética nuclear es resultado de la fragmentación histórica del hábitat y de un patrón de aislamiento por distancia.

Una evaluación detallada de los loci potencialmente bajo selección permitiría incluir el papel de esta fuerza en la historia evolutiva de la especie. Adicionalmente, una secuenciación específicamente destinada al genoma completo del cloroplasto de *Y. valida* confirmaría la solidez de los análisis filogenéticos y del reloj molecular. Así mismo, la realización de análisis de estimación de los tiempos de divergencia basados en SNPs, en conjunto con los resultados provistos del tiempo de diferenciación de *Y. capensis* y *Y. valida*, podrían dar una aproximación de la edad de los linajes nucleares. Finalmente, modelaciones de la distribución potencial en otros periodos de tiempo y simulaciones del aislamiento por distancia podrían ampliar la comprensión de los efectos de los procesos históricos en la distribución de la variación genética.

Literatura citada

Alamo Herrera, C.R. 2019. Evaluación del flujo genético de *Yucca valida* (Asparagaceae) en diferentes escalas espaciales. Tesis de Maestría en Ciencias. Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California. 45 pp.

Alexander, D y Lange, K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12, 246. BMC bioinformatics. 12. 246. 10.1186/1471-2105-12-246.

Andrews, S. 2010. FASTQC. A quality control tool for high throughput sequence data [En línea]. Disponible en http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arbogast, B.S. y Kenagy, G.J. 2001, Comparative phylogeography as an integrative approach to historical biogeography. Journal of Biogeography, 28:819-825. 10.1046/j.1365-2699.2001.00594.x.

Arteaga, M., Bello-Bedoy, R. y Gasca-Pineda, J. 2020. Hybridization Between Yuccas from Baja California: Genomic and Environmental Patterns. Frontiers in Plant Science. 11. 10.3389/fpls.2020.00685.

Avise, J.C. 2000. Phylogeography: The history and formation of species. Harvard.

Brusca, R.C. 2015. A Brief Geological History of Northwestern Mexico. Disponible en www.rickbrusca.com

Burleigh, J., Hilu, K. y Soltis, D. 2009. Inferring phylogenies with incomplete data sets: A 5-gene, 567-taxon analysis of angiosperms. BMC evolutionary biology. 9. 61. 10.1186/1471-2148-9-61.

Bushnell, B., Rood, J., Singer, E. 2017. BBMerge - Accurate paired shotgun read merging via overlap. PLoS One. 10.1371/journal.pone.0185056.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T. 2009. BLAST+:architecture and applications. BMC Bioinformatics 10:421. BMC bioinformatics. 10. 421.

Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., Cresko, W. 2013. Stacks: An analysis tool set for population genomics. Molecular ecology. 22. 10.1111/mec.12354.

Clark-Tapia R., Molina-Freaner F. 2003. The genetic structure of a columnar cactus with a disjunct distribution: *Stenocereus gummosus* in the Sonoran Desert. Heredity. 90(6):443-450.

Cornuet, J.M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.M., Estoup, A. 2014. DIYABCv2.0:a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data. Bioinformatics (Oxford, England). 30. 10.1093/bioinformatics/btt763.

Csillery, K., Blum, M.G.B., Gaggiotti, O.E., Francois, O. 2010. Approximate Bayesian Computation (ABC) in practice. Trends in Ecology and Evolution, 25(7), 410-418. 10.1016/j.tree.2010.04.001.

Huson, D.H. y Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution. 10.1093/molbev/msj030.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 10.1038/nmeth.2109.

De la Rosa-Conroy, L., Gasca-Pineda, J., Bello-Bedoy, R., Eguiarte, L.E., Arteaga, M.C. 2020. Genetic patterns and changes in availability of suitable habitat support a colonization history of a North American perennial plant. Plant Biology J. 22:233-242.

Dolby, G.A., Bennett, S.E.K., Lira-Noriega, A., Wilder, B.T., Munguía-Vega, A. 2015. Assessing the Geological and Climatic Forcing of Biodiversity and Evolution Surrounding the Gulf of California. Journal of the Southwest 57(2), 391-455. 10.1353/jsw.2015.0005.

Drummond, A.J. y Bouckaert, R.R. 2015. Bayesian evolutionary analysis with BEAST. Cambridge University Press.

Drummond, A.J. y Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology, 7, 214.

Drummond, A., Suchard, M.A., Xie, D., Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution. 22. 1185-1192.

Emerson, B.C., Paradis, C. Thébaud. 2001. Revealing the demographic histories of species using DNA sequences. Trends in Ecology and Evolution, 16 (12):707-716.

Evanno, G., Regnaut, S., Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology, 14:2611-2620. 10.1111/j.1365-294X.2005.02553.x.

Excoffier, L. y Lischer, H.E.L. 2010. Arlequin suite version 3.5:A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, 10:564-567.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., Foll, M. 2013. Robust demographic inference from genomic and SNP data. PLOS Genetics, 9(10):e1003905. 10.1371/journal.pgen.1003905

Fehlberg, S. y Ranker, T. 2009. Evolutionary history and phylogeography of *Encelia farinosa* (Asteraceae) from the Sonoran, Mojave, and Peninsular Deserts. Molecular Phylogenetics and Evolution. 50. 326-335. 10.1016/j.ympev.2008.11.011.

Foll, M. y Gaggiotti, O. 2008. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics. 180. 977-93. 10.1534/genetics.108.092221.

Garrick, R.C., Nason, J.D., Meadows, C.A., Dyer, R.J. 2009. Not just vicariance: Phylogeography of a Sonoran Desert euphorb indicates a major role of range expansion along the Baja peninsula. Molecular Ecology, 18:1916-1931. 10.1111/j.1365-294X.2009.04148.x.

Gonzalez-Abraham, C., Garcillán, P., Ezcurra, E. 2010. Ecorregiones de la península de Baja California: Una síntesis. Boletín de la Sociedad Botánica de México. 87. 10.17129/botsci.302.

González-Rubio C., García-De León F.J., Rodríguez-Estrella R. 2016. Phylogeography of endemic Xantus' hummingbird (*Hylocharis xantusii*) shows a different history of vicariance in the Baja California Peninsula. Molecular Phylogenetics and Evolution. 10.1016/j.ympev.2016.05.039.

Gonzalez Trujillo, R., Correa-Ramirez, M., Ruiz-Sanchez, E., Salinas, E., Jiménez, M.-L., Garcia De Leon, F. 2016. Pleistocene refugia and their effects on the phylogeography and genetic structure of the Wolf spider *Pardosa sierra* (Araneae: Lycosidae) on the Baja California Peninsula. Journal of Arachnology. 44. 367-379. 10.1636/R15-84.1.

Grismer, L.L. 2002. A re-evaluation of the evidence for a mid-Pleistocene mid-peninsular seaway in Baja California: A reply to riddle et al. Herpetological Review. 33. 15-16.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology. 59(3):307-21.

Gutierrez-Flores, C., Garcia De Leon, F., León-de la Luz, J., Cota-Sánchez, J.H. 2016. Microsatellite genetic diversity and mating systems in the columnar cactus *Pachycereus pringlei* (Cactaceae). Perspectives in Plant Ecology, Evolution and Systematics. 22. 10.1016/j.ppees.2016.06.003.

Hafner, D. y Riddle, B. 2011. Boundaries and barriers of North American warm deserts: An evolutionary perspective. En: Upchurch, P., McGowan, A., C. Slater (eds.). Palaeogeography and palaeobiogeography: Biodiversity in space and time, The Systematics Association Special Volume Series, CRC Press, Boca Raton. p. 75-113.

Haga, N., Kobayashi, M., Michiki, N., Takano, T., Baba, F., Kobayashi, K., Ohyanagi, H., Ohgane, J., Yano, K., y Yamane, K. 2019. Complete chloroplast genome sequence and phylogenetic analysis of wasabi (*Eutrema japonicum*) and its relatives. Scientific Reports. 9. 1-10. 10.1038/s41598-019-49667-z.

Harrington, S.M., Hollingsworth, B.D., Higham, T.E., Reeder, T.W. 2018. Pleistocene climatic fluctuations drive isolation and secondary contact in the red diamond rattlesnake (*Crotalus ruber*) in Baja California. Journal of Biogeography. 10.1111/jbi.13114.

Hedrick, P.W. 2011. Genetics of populations. Jones and Bartlett Publishers. Sudbury.

Hewitt, G.M. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907-913. Nature. 405. 907-13. 10.1038/35016000.

Hewitt, G.M. 2004. The structure of biodiversity – insights from molecular phylogeography. Frontiers in Zoology 1, 4. 10.1186/1742-9994-1-4

Holmgren, C., Betancourt, J., Rylander, K. 2011. Vegetation history along the eastern, desert escarpment of the Sierra San Pedro Mártir, Baja California, Mexico. Quaternary Research. 75. 647-657. 10.1016/j.yqres.2011.01.008.

Huth, C.J. y Pellmyr, O. 2000. Pollen-mediated selective abortion in Yuccas and its consequences for the plant-pollinator mutualism. Ecology, 81:1100-1107. 10.1890/0012-9658(2000)081[1100:PMSAIY]2.0.CO;2.

Irwin, D. 2002. Phylogeographic breaks without barriers to gene flow. Evolution. 56. 2383 - 2394. 10.1111/j.0014-3820.2002.tb00164.x.

Jombart, T. 2008. Adegenet: An R package for the multivariate analysis of genetic markers. Bioinformatics. 24. 1403-1405. 10.1093/bioinformatics/btn129/.

Kimura, M. 1968. Evolutionary Rate at the Molecular Level. Nature 217, 624–626. 10.1038/217624a0.

Klimova, A., Hoffman, J.I., Gutierrez-Rivera, J.N., Leon de la Luz, J., Ortega-Rubio, A. 2017. Molecular genetic analysis of two native desert palm genera, *Washingtonia* and *Brahea*, from the Baja California Peninsula and Guadalupe Island. Ecology and evolution, 7(13), 4919–4935. 10.1002/ece3.3036.

Kumar, R. y Kumar, V. 2018. A review of phylogeography: Biotic and abiotic factors, Geology, Ecology, and Landscapes, 2:4, 268-274, 10.1080/24749508.2018.1452486

Kumar S., Stecher G., Li M., Knyaz C., Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution. 35(6):1547-1549. 10.1093/molbev/msy096.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359. Nature methods. 9. 357-9. 10.1038/nmeth.1923.

Latch, E.K., Heffelfinger, J.R., Fike, J.A., Rhodes, O.E. Jr. 2009. Species-wide phylogeography of North American mule deer (*Odocoileus hemionus*): Cryptic glacial refugia and postglacial recolonization. Molecular Ecology. 18(8):1730-1745. 10.1111/j.1365-294X.2009.04153.x.

Leaché, A.D., Crews, S.C., Hickerson, M.J. 2007. Two waves of diversification in mammals and reptiles of Baja California revealed by hierarchical Bayesian analysis. Biology letters, 3(6), 646–650. 10.1098/rsbl.2007.0368.

Lefort, V., Longueville, J.E., Gascuel, O. 2017. SMS: Smart Model Selection in PhyML. Molecular Biology and Evolution, Volume 34, Issue 9, 2422-2424.

Lira-Noriega A., Toro-Núñez O., Oaks J.R., Mort, M.E. 2015. The roles of history and ecology in chloroplast phylogeographic patterns of the bird-dispersed plant parasite *Phoradendron californicum* Nutt. (Viscaceae) in the Sonoran Desert. American Journal of Botany 102:149–164. 10.3732/ajb.1400277.

Lischer H.E., Excoffier L. 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 28(2):298-299. 10.1093/bioinformatics/btr642.

Luna, P.A. 2018. Diversidad genética de *Yucca capensis* (Asparagaceae), planta endémica de la Sierra de la Laguna. Tesis de Maestría en Ciencias. Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California. 45 pp.

Luu, K., Bazin, E., Blum, M.G. 2017. Pcadapt: An R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources. 17(1):67-77. 10.1111/1755-0998.12592.

Martínez-Noguez, J. J., León de la Luz, J. L., Delgadillo Rodríguez, J., León, G. D., y Francisco, J. 2020. Phylogeography and genetic structure of an iconic tree of the Sonoran Desert, the Cirio (*Fouquieria columnaris*), based on chloroplast DNA, Biological Journal of the Linnean Society, Volume 130, Issue 3, July 2020, Pages 433–446. 10.1093/biolinnean/blaa065.

Meirmans, Patrick. 2020. GenoDive v. 3.0:Easy-to-use software for the analysis of genetic data of diploids and polyploids. Molecular Ecology Resources. 20. 10.1111/1755-0998.13145.

Miller, M. A., Pfeiffer, W. y Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov.2010, New Orleans, LA pp 1-8.

Minnich, R.A., Franco-Vizcaíno, E.F., Goforth, B.R. 2014. Distribution of chaparral and pine-oak ''skyislands'' in central and southern Baja California and implications of packrat midden records on climate change since the Last Glacial Maximum. En: Whencke E., Lara-Lara R., Álvarez-Borrego S., Ezcurra, E. (eds) Conservation science in NW Mexico. Instituto Nacional de Ecología, CICESE, UC MEXUS, México.

Miura, S., Tamura, K., Tao, Q., Huuki, L.A., Kosakovsky-Pond, S.L., Priest, J., Deng, J., Kumar, S. 2020. A new method for inferring timetrees from temporally sampled molecular sequences. Plos Computational Biology. 16(1):e1007046. 10.1371/journal.pcbi.1007046.

Murphy, R. W. y Aguirre-León, G. 2002. The nonavian reptiles: Origins and evolution. A new island biogeography of the Sea of Cortés. pp.181-220. Oxford University Press, New York.

Nason, J.D., Hamrick, J.L., Fleming, T.H. 2002. Historical Vicariance and Postglacial Colonization Effects on the Evolution of Genetic Structure in *Lophocereus*, a Sonoran Desert Columnar Cactus. Evolution: International Journal of Organic Evolution, 56, 2214-2226. 10.1111/j.0014-3820.2002.tb00146.x.

Palmer, J. y Stein, D. 1986. Conservation of chloroplast genome structure among vascular plants. Current Genetics. 10, 823–833. 10.1007/BF00418529

Paris, J.R., Stevens, J.R., Catchen, J.M. 2017. Lost in parameter space: A road map for stacks. Methods in Ecology and Evolution, 8:1360-1373. 10.1111/2041-210X.12775.

Pellmyr, O. y Leebens-Mack, J. 1999. Forty million years of mutualism: Evidence for eocene origin of the yucca-yucca moth association. Proceedings of the National Academy of Sciences of the United States of America, 96(16), 9178–9183. 10.1073/pnas.96.16.9178.

Pellmyr, O. 2003. Yuccas, Yucca Moths, and Coevolution: A Review. Annals of the Missouri Botanical Garden, 90(1), 35-55. 10.2307/3298524.

Pellmyr, O., Balcazar-Lara, M., Segraves, K.A., Althoff, D.M., y Littlefield, R.J. 2008. Phylogeny of the pollinating yucca moths, with revision of Mexican species (*Tegeticula* and *Parategeticula*; Lepidoptera, Prodoxidae). Zoological Journal of the Linnean Society, 152:297-314. 10.1111/j.1096-3642.2007.00361.x.

Pellmyr, O., Massey, L.K., Hamrick, J.L., y Feist, M.A. 1997. Genetic consequences of specialization: *Yucca* moth behavior and self-pollination in yuccas. Oecologia. 109(2):273-278. 10.1007/s004420050083.

Pellmyr, O., Segraves, K.A., Althoff, D.M., Balcázar-Lara, M., Leebens-Mack, J. 2007. The phylogeny of yuccas. Molecular Phylogenetics and Evolution. 43(2):493-501. 10.1016/j.ympev.2006.12.015.

Pritchard, J. K., Stephens, M., y Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155(2):945-959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics. 81(3):559-575.

R Core Team. 2016. R:A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponible en http://www.R-project.org/

Rebernig, C.A., Schneeweiss, G.M., Bardy, K.E., Schönswetter, P., Villaseñor, J.L., Obermayer, R., Stuessy, T.F., Weiss-Schneeweiss, H. 2010. Multiple Pleistocene refugia and Holocene range expansion of an abundant southwestern American desert plant species (*Melampodium leucanthum*, Asteraceae). Molecular Ecology, 19:3421-3443. 10.1111/j.1365-294X.2010.04754.x.

Riddle, B.R. y Hafner, D.J. 2006. Phylogeography in historical biogeography: Investigating the biogeographic histories of populations, species, and young biotas. En: Ebach, M.C. y Tangney, R.S. Biogeography in a changing world. CRC Press, Boca Raton, FL. pp 161-176.

Riddle, B.R., Hafner, D.J., Alexander, L.F., Jaeger, J.R. 2000. Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. Proceedings of the National Academy of Sciences 97 (26) 14438-14443. 10.1073/pnas.250413397.

Rivas, M.J., Domínguez-García, S., Carvajal-Rodríguez, A. 2015. Detecting the Genomic Signature of Divergent Selection in Presence of Gene Flow. Current genomics. 16(3), 203–212. 10.2174/1389202916666150313230943

Rochette, N. y Catchen, J. 2017. Deriving genotypes from RAD-seq short-read data using Stacks. Nature Protocols. 12, 2640–2659. 10.1038/nprot.2017.123.

Rodríguez-Robles, J.A. y De Jesús-Escobar, J.M. 2000 Molecular systematics of New World gopher, bull, and pinesnakes (*Pituophis*: Colubridae), a transcontinental species complex. Molecular Phylogenetics and Evolution. 14, 35-50. 10.1006/mpev.1999.0698.

Russello, M.A., Waterhouse, M.D., Etter, P.D., Johnson, E.A. 2015. From promise to practice: Pairing non-invasive sampling with genomics in conservation. PeerJ 3:e1106. 10.7717/peerj.1106.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology, 7, 539. 10.1038/msb.2011.75.

Singh, N., Choudhury, D.R., Singh, A.K., Kumar, S., Srinivasan, K., Tyagi, R.K., Singh, R. 2013. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. PLoS One, 8(12):e84136. 10.1371/journal.pone.0084136.

Smith, C.I., Pellmyr, O., Althoff, D.M., Balcázar-Lara, M., Leebens-Mack, J., Segraves, K.A. 2008. Pattern and timing of diversification in *Yucca* (Agavaceae):specialized pollination does not escalate rates of diversification. Proceedings of the Royal Society B: Biological Sciences, 275(1632), 249–258.

Sork, V.L., Gugger, P.F., Chen, J., Werth, S. 2016. Evolutionary lessons from California plant phylogeography. Proceedings of the National Academy of Sciences of the United States of America. 113 29, 8064-71. 10.1073/pnas.1602675113.

Starr, T.N., Gadek, K.E., Yoder, J.B., Flatz, R., Smith, C.I. 2013. Asymmetric hybridization and gene flow between Joshua trees (Agavaceae: *Yucca*) reflect differences in pollinator host specificity. Molecular Ecology, 22, 437-449.

Sweet, L.C., Green, T., Heintz, J.G.C., Frakes, N., Graver, N., Rangitsch, J.S., Rodgers, J.E., Heacox, S., Barrows, C.W. 2019. Congruence between future distribution models and empirical data for an iconic species at Joshua Tree National Park. Ecosphere 10(6):e02763. 10.1002/ecs2.2763.

Tamura, K., Battistuzzi, F., Billing-Ross, P., Murillo, O., Filipski, A. y Kumar, S. 2012. Estimating divergence times in large molecular phylogenies. Proceedings of the National Academy of Sciences of the United States of America. 109 (47) 19333-19338. 10.1073/pnas.1213199109.

Thiede, J. 2020. *Yucca*. Agavaceae. En: Eggli, U. Nyffeler, R. (Eds.) Monocotyledons. Springer. pp. 363-421.

Turner, R.M., Bowers, J.E., Burgess, T.L. 1995. Sonoran Desert plants: an ecological atlas. University of Arizona Press.

Upton, D. E., Murphy, R. W. (1997). Phylogeny of the side-blotched lizards (Phrynosomatidae: *Uta*) based on mtDNA sequences :support for midpeninsular seaway in Baja California. Molecular Phylogenetics and Evolution, 8(1):104-113. 10.1006/mpev.1996.0392.

Vander Wall, S. B., Esque, T., Haines, D., Garnett, M., y Waitman, B. A. 2006. Seed-caching rodents disperse joshua tree (*Yucca brevifolia*) seeds. Ecoscience, 13(4):539-543.

Waitman B.A., Vander Wall S.B., Esque T.C. 2012. Seed dispersal and seed fate in Joshua tree (*Yucca brevifolia*). Journal of Arid Environments, 81:1-8 10.1016/j.jaridenv.2011.12.012.

Wiggins, I.L. 1980. Flora of Baja California. Stanford, CA: Stanford University Press.

Wilkinson, R. 2013. Approximate Bayesian computation (ABC) gives exact results under the assumption of model error, Statistical Applications in Genetics and Molecular Biology, 12(2), 129-141.

Wright, S. 1943. Isolation by distance. Genetics, 28(2):114-138.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution. 19:395-420. 10.1111/j.1558-5646.1965.tb01731.x.

Zheng, Y. y Wiens, J.J. 2015. Do missing data influence the accuracy of divergence-time estimation with BEAST? Molecular Phylogenetics and Evolution, 85:41-49. 10.1016/j.ympev.2015.02.002.

Zhenxiang, X., Liang, L., Charles, C.D. 2016. The Impact of Missing Data on Species Tree Estimation. Molecular Biology and Evolution, 33(3):838–860. 10.1093/molbev/msv266

Zink, R. M., Blackwell, R. y Rojas-Soto, O. 1997. Species limits on the Le Conte's Thrasher. Condor. 99:132-138.

Zuckerkandl, E., y Pauling, L. 1965. Molecules as documents of evolutionary history. Journal of Theoretical Biology, 8(2):357-366.