



Universidad Autónoma de Querétaro  
Facultad de Ciencias Naturales

Exploración de hibridación en agaves mezcaleros  
en la Sierra El Doctor, Querétaro: un enfoque  
morfológico y genómico  
Tesis

Que como parte de los requisitos para obtener  
el Grado de Maestra en Ciencias Biológicas

Presenta  
Patricia Mercedes González Márquez

Dirigido por:  
Dra. Emma Fabiola Magallán Hernández

Co-Director:  
Dra. Erika Aguirre Planter

Querétaro, Qro. a 2024

La presente obra está bajo la licencia:  
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

### Usted es libre de:

**Compartir** — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

### Bajo los siguientes términos:



**Atribución** — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



**NoComercial** — Usted no puede hacer uso del material con [propósitos comerciales](#).



**SinDerivadas** — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

**No hay restricciones adicionales** — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

### Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



Universidad Autónoma de Querétaro  
Facultad de Ciencias Naturales



# Exploración de hibridación en agaves mezcaleros en la Sierra El Doctor, Querétaro: un enfoque morfológico y genómico Tesis

Que como parte de los requisitos para obtener  
el Grado de Maestra en Ciencias Biológicas

Presenta

Patricia Mercedes González Márquez

Dirigido por:

Dra. Emma Fabiola Magallán Hernández

Co-Director:

Dra. Erika Aguirre Planter

Dra. Emma Fabiola Magallán Hernández  
Presidente

Dra. Erika Aguirre Planter  
Secretaria

Dr. Juan Antonio Valencia  
Vocal

Dr. Jaime Gasca Pineda  
Suplente

Dr. Santiago Vergara Pineda  
Suplente

Centro Universitario, Querétaro, Qro.  
Fecha de aprobación por el Consejo Universitario (mes y año)  
México

A México y a todo lo que este grandioso país significa para mí hoy

## AGRADECIMIENTOS

Muchas personas contribuyeron en la realización de esta tesis.

Le agradezco a Fabiola Magallán, por aceptarme como su estudiante y darme esta oportunidad, a pesar del tiempo limitado. Gracias por enseñarme y también por escuchar y respetar mis opiniones.

A Erika Aguirre por su inmensa ayuda en el proceso de extracción de ADN y por llevar la gestión de la secuenciación, también por sus consejos sobre los análisis bioinformáticos y, sobre todo, por su inagotable optimismo.

A Jaime Gasca, por su colaboración en los análisis bioinformáticos, consejos y revisiones.

A Juan Antonio Valencia y Santiago Vergara por su colaboración en el análisis de los datos morfológicos y sus observaciones.

A Irene Martínez, que merece un agradecimiento especial por su invaluable ayuda, tanto tiempo dedicado y su infinita paciencia.

A Luis Eguiarte, gracias por esta maravillosa oportunidad de colaboración y por aceptarme como integrante temporal de su laboratorio

A todo el team Ecología, en especial a Cecile, Lupita y Diego por sus consejos.

Al equipo de muestreo, que pasó una noche de invierno en la sierra: Caro, las Pao, Alviter, Frida, Memo, Marco, Toño, Santiago y Fabby.

A los pobladores de Chavarrías, en especial don Dionisio, Don Fidel, Don Baltazar y Max por permitirnos trabajar en sus tierras y acompañarnos en las salidas de campo.

A Paola Puga que me presentó a Fabiola, sin su intervención quien sabe que tesis random hubiera terminado haciendo.

A Christopher Smith que no ayudó en la fase inicial del planteamiento del problema de investigación.

A Jacinto Treviño Carreón, por compartir su información sobre la sierra de Tamaulipas donde también coexisten estos tres agaves.

A Xóchitl Granados, por sus consejos sobre análisis de hibridación.

Al programa de becas **CONAHCYT** del que fui beneficiaria con número (CVU): 1231175.

Muchas gracias a todos

## ÍNDICE

<b>Resumen .....</b>	<b>9</b>
<b>Abstract .....</b>	<b>10</b>
<b>1. Introducción .....</b>	<b>11</b>
<b>Hipótesis.....</b>	<b>12</b>
<b>Objetivo general .....</b>	<b>12</b>
<b>Objetivos específicos .....</b>	<b>13</b>
<b>2. Marco teórico .....</b>	<b>14</b>
<b>2.1 Hibridación en plantas .....</b>	<b>14</b>
<b>2.2 Genómica .....</b>	<b>17</b>
<b>2.3.1. Genómica de <i>Agave</i> .....</b>	<b>20</b>
<b>2.3 Agaves mezcaleros de la Sierra El Doctor .....</b>	<b>21</b>
<b>3. Materiales y Métodos .....</b>	<b>22</b>
<b>3.1. Área de estudio .....</b>	<b>22</b>
<b>3.2. Análisis morfométrico .....</b>	<b>23</b>
<b>3.3. Extracción de ADN .....</b>	<b>25</b>
<b>3.4. Análisis Bioinformático .....</b>	<b>26</b>
<b>3.5. Diversidad y estructura genómica.....</b>	<b>27</b>
<b>4. Resultados .....</b>	<b>28</b>
<b>4.1 Morfología.....</b>	<b>28</b>
<b>5. Discusión .....</b>	<b>46</b>
<b>5.1 Morfología.....</b>	<b>46</b>
<b>5.2 Genómica .....</b>	<b>48</b>
<b>5.3 Integración morfo-genómica .....</b>	<b>52</b>
<b>6. Conclusiones .....</b>	<b>53</b>
<b>7. Referencias .....</b>	<b>54</b>
<b>8. Anexos.....</b>	<b>64</b>

## ÍNDICE DE CUADROS

### Cuadro 1

Número de individuos muestreados por especie y por sitio,  
después del filtrado por calidad y datos faltantes.....24

### Cuadro 2

Comparación de medias de variables morfológicas entre las  
especies de agaves mezcaleros de la Sierra El Doctor.....27

### Cuadro 3

Número de SNPs obtenidos a partir del filtrado con vcftools.....34

### Cuadro 4

Diversidad genética estimada usando SNPs.  $H_o$ : heterocigosidad observada,  
 $H_s$ : heterocigosidad esperada,  $F_{IS}$ : coeficiente de endogamia,  
 $F_{st}$ : coeficiente de diferenciación genética,  $N$ : tamaño de muestra.....36

### Cuadro 5

Coeficiente de diferenciación genética pareado ( $F_{ST}$ ).....36

### Cuadro 6

Comparación entre los intervalos de las principales variables morfológicas  
medidas en el presente estudio y los registrados por García et al., en 2022.....44

## ÍNDICE DE FIGURAS

Figura 1. Zonas de muestreo dentro de la Sierra El Doctor, Cadereyta, Querétaro.....	21
Figura 2. MDS con variables morfológicas cuantitativas por especies .....	28
Figura 3: Prueba Kruskal- Wallis de variables morfológicas por especie. A. Altura de la roseta, B. Número de hojas, C. Ancho de la hoja, D. Longitud de la espina terminal, E. Longitud de la hoja, F. Ancho de la roseta.....	28
Figura 4. MDS con variables morfológicas cuantitativas por sitio. A. <i>Agave montana</i> . B. <i>Agave gentryi</i> . C. <i>Agave salmiana</i> . D. <i>Agave intermedio</i> .....	30
Figura 5. Análisis de componentes principales. A. se muestra la contribución de cada variable. B. contribución de cada variable. C. se muestra el agrupamiento de cada individuo asignado a una especie <i>a priori</i> .....	31
Figura 6. FAMD con datos morfológicos donde se muestra el agrupamiento de cada individuo asignado a una especie <i>a priori</i> .....	32
Figura 7. A. Análisis de clustering K-Medoids con datos morfológicos, k=3. B. Matriz de confusión.....	33
Figura 8. El dendrograma muestra la relación de similitud entre los individuos, donde las ramas más largas representan una mayor distancia entre los individuos. La distancia entre los individuos se calculó utilizando la métrica de distancia euclidiana y se agruparon utilizando el método de agrupamiento jerárquico con el método de Ward. (Am: <i>Agave montana</i> , Ag: <i>Agave gentryi</i> , As: <i>Agave salmiana</i> , Ai. <i>Agave intermedio</i> ).....	35
Figura 9. Análisis de componentes principales de los 92 individuos. Los primeros dos componentes explican 11.92 % de la variabilidad de los datos.....	35
Figura 10. Análisis Stucture (K=2, K=3, K=4), para el conjunto de 92 individuos. El menor valor de k obtenido fue K=2.....	37
Figura 11. Matriz de coancestría FineRadStructure que incluye a los 92 individuos. Cuanto más oscuro es el cuadrado, mayor es la similitud genética entre un par de individuos. Las etiquetas	



marcadas con color granada agrupan a los individuos de *A. montana* y los intermedios afines, mientras que las etiquetas mostaza corresponden a *A. gentryi*, *A. salmiana* y sus intermedios afines.....38

Figura 12. A. Análisis de componentes principales de *A. montana* por sitio de muestreo. B. Dendrograma de los 27 individuos de *A. montana*.....39

Figura 13. Análisis Structure de los 27 individuos de *A. montana* muestreados. El menor valor de k obtenido fue K=1.....40

Figura 14. A. Dendrograma de los 21 individuos de *A. gentryi* B. Análisis de componentes principales de *A. gentryi* por sitio de muestreo. C. Dendrograma de los 21 individuos de *A. salmiana*. D. Análisis de componentes principales de *A. salmiana* por sitio de muestreo.....40

Figura 15. Análisis Structure A. individuos de *A. gentryi* y todos los intermedios. B. individuos de *A. montana* y todos los intermedios. En ambos casos el menor valor de k obtenido fue K=2.....42

Figura 16. Análisis Structure de *A. montana* e intermedios afines. El menor valor de k obtenido fue K=1.....43

Figura 17. Análisis Structure de *A. gentryi*, *A. salmiana* e intermedios afines. El menor valor de k obtenido fue K=1.....43

## Resumen

En los últimos años el consumo y la comercialización del mezcal han aumentado de forma significativa. En la Sierra El Doctor, Cadereyta, Querétaro, encontramos principalmente tres especies de agaves mezcaleros: *A. montana*, *A. gentryi* y *A. salmiana*. Además, hay individuos con características morfológicas intermedias, que no pueden ser claramente asignados a una de estas especies. El objetivo de este trabajo fue determinar, a través de información morfológica y genómica, si estos individuos con rasgos intermedios son híbridos. Para ello se caracterizaron morfológicamente 114 individuos de las tres especies y los intermedios. Se secuenciaron 96 individuos con la técnica de Double Digest Restriction-site Associated DNA sequencing (ddRADseq). Para el ensamblaje se utilizó como referencia el transcriptoma de *A. tequilana*. Los diferentes análisis morfológicos fueron congruentes en sus resultados y permitieron agrupar, pero no separar completamente cada una de las especies. En todos los casos *A. montana* se separa de las otras dos especies, mientras que *A. gentryi* y *A. salmiana* se superponen y no pueden ser claramente separados. En el caso de los análisis genómicos, *A. montana* también se separa notablemente del resto y no muestra señales de mezcla. Mientras que *A. salmiana*, *A. gentryi* y los intermedios afines se presentaron como variante genética única, con cierta mezcla de la variante de *A. montana* en algunos casos. Se detectó un déficit de heterocigotos y alto coeficiente de endogamia en *A. montana*, en contraste con *A. gentryi* y *A. salmiana*, que mostraron exceso de heterocigotos y bajo coeficiente de endogamia. A pesar de que la variabilidad morfológica y la plasticidad fenotípica dificultan la detección de mezcla, la evidencia genómica sugiere un flujo genético entre las especies, especialmente de *A. montana* hacia *A. gentryi* y *A. salmiana*, apoyando la existencia de procesos de hibridación en curso en lugar de variabilidad debida únicamente a la plasticidad fenotípica.

**Palabras clave:** SNPs, *Agave montana*, *Agave gentryi*, *Agave salmiana*, híbrido.

## Abstract

In recent years, the consumption and commercialization of mezcal have significantly increased. In Sierra El Doctor, Cadereyta, Querétaro, we primarily find three mezcal agaves: *A. montana*, *A. gentryi*, and *A. salmiana*. Additionally, there are individuals with intermediate morphological traits that cannot be clearly assigned to any of these species. The objective of this study was to determine, through morphological and genomic information, whether these individuals with intermediate traits are hybrids. For this purpose, 114 individuals of the three species and the intermediates were morphologically characterized. Ninety-six individuals were sequenced using the Double Digest Restriction-site Associated DNA sequencing (ddRADseq) technique. The transcriptome of *A. tequilana* was used as a reference for the assembly. The different morphological analyses were congruent in their results and allowed grouping but not completely separating each of the species. In all cases, *A. montana* was distinct from the other two species, while *A. gentryi* and *A. salmiana* overlapped and could not be clearly separated. In the genomic analyses, *A. montana* also stood out from the rest and showed no signs of admixture. Meanwhile, *A. salmiana*, *A. gentryi*, and the related intermediates appeared as a single genetic variant, with some instances of admixture with the *A. montana* variant. A deficit of heterozygotes and a high inbreeding coefficient were detected in *A. montana*, in contrast to *A. gentryi* and *A. salmiana*, which showed an excess of heterozygotes and a low inbreeding coefficient. Despite the fact that morphological variability and phenotypic plasticity complicate the detection of admixture, genomic evidence suggests genetic flow between the species, especially from *A. montana* to *A. gentryi* and *A. salmiana*, supporting the existence of ongoing hybridization processes rather than variability solely due to phenotypic plasticity.

**Keywords:** SNPs, *Agave montana*, *Agave gentryi*, *Agave salmiana*, hybrid.

## 1. Introducción

El género *Agave* L. pertenece a la familia Asparagaceae, subfamilia Agavoideae (APG IV, 2016). Comprende más de 200 especies (García-Mendoza et al., 2019), que se encuentran en una amplia variedad de climas, desde tropicales hasta templados (Figueredo-Urbina y Torres-García, 2017). Estas plantas suculentas son originarias del continente americano, y es en México donde se encuentra el centro de mayor diversidad, con aproximadamente 75 % de las especies del género (Eguiarte et al., 2021). Este grupo vegetal ha tenido gran importancia para las culturas indígenas americanas debido a los diversos usos que se le han dado desde hace 10,000 años (Colunga-GarcíaMarín et al., 2017). Estos usos incluyen la obtención de fibras, la construcción de viviendas, la creación de artesanías, conformación de cercos vivos y refuerzo de bordes de terrazas en terrenos con pendiente alta. Además, han sido consumidos como alimento y con fines medicinales (Pérez-Zavala, 2020). Otro de sus usos más relevantes es la preparación de bebidas alcohólicas como tequila, pulque y mezcal (Vega-Ramos et al., 2022).

El mezcal es un licor destilado a partir del tallo del agave. Durante la última década, el consumo y la comercialización de mezcal han experimentado un notable auge tanto en el mercado nacional como internacional (Valiente-Banuet, 2023). A diferencia del tequila, que generalmente se produce con agaves cultivados en grandes plantaciones, en la elaboración del mezcal se valora más la producción artesanal, los métodos tradicionales y el vínculo con la cultura y las comunidades mezcaleras. En consecuencia, la extracción de agaves mezcaleros del medio silvestre ha aumentado.

Alrededor de 50 especies de *Agave* son consideradas aptas para la producción de mezcal (Eguiarte et al., 2021). Entre ellas, las más utilizadas son *Agave vivipara* (antes *A. angustifolia*), *A. rhodacantha*, *A. potatorum*, *A. nussaviorum*, *A. seemanniana*, *A. hookeri*, *A. desmettiana*, *A. marmorata*, *A. karwinskii*, *A. cupreata* y *A. americana* (Torres et al., 2015). Mientras que existen otras especies que por el tamaño reducido de sus

poblaciones o lo restringido de su distribución no son tan frecuentemente utilizadas, pero sí producen mezcal de alta calidad (Jacques-Hernández *et al.*, 2007). Una de estas especies es *Agave montana* Villarreal (coms. pers Dr. Jacinto Treviño, 2023).

*Agave montana* fue descrito en 1996 en la Sierra Madre Oriental y aunque se cree que está relacionada con *A. gentryi* y *A. parrasana* (Villarreal, 1996), aún no se han realizado estudios filogenéticos para confirmarlo. Su distribución abarca Nuevo León, Tamaulipas y Coahuila (Villarreal, 1996; Treviño-Carreón *et al.*, 2011). Recientemente (Magallán *et al.*, en preparación), se registró una población en el estado de Querétaro, específicamente en la Sierra El Doctor, Cadereyta, donde no se utiliza para elaborar mezcal, pero sus flores se usan extensivamente, como alimento regional.

En la Sierra El Doctor podemos encontrar otras especies de agaves mezcaleros. Una de ellas es silvestre (*A. gentryi*) y otra, que puede ser encontrada tanto en zonas de cultivo como en áreas silvestres (*A. salmiana*). En la zona también se pueden encontrar individuos que presentan características morfológicas intermedias y no pueden ser claramente asignados a una especie en particular, aunque presentan rasgos morfológicos que los relacionan con las especies mencionadas anteriormente. Esta variabilidad podría ser el resultado de la alta diversidad intrínseca dentro de las especies (Mora-López *et al.*, 2011; Parker *et al.*, 2014). Otra posible explicación, es que estos individuos sean el resultado de hibridación, ya que se sabe que el género *Agave* tiende a hibridar con frecuencia (Gentry, 1982; Roll *et al.*, 2021).

## **Hipótesis**

Los individuos de agaves mezcaleros con características morfológicas intermedias en la Sierra El Doctor son el resultado de un proceso de hibridación.

## **Objetivo general**

Determinar, a través de información morfológica y genómica, si los individuos de agaves mezcaleros con rasgos intermedios en la Sierra El Doctor son híbridos.

## **Objetivos específicos**

1-Caracterizar la variabilidad morfológica dentro de las poblaciones de agaves mezcaleros en la Sierra El Doctor, incluyendo a aquellos individuos que presentan rasgos intermedios.

2-Caracterizar la diversidad genética y la estructura de poblaciones de los agaves mezcaleros de la Sierra El Doctor, con un enfoque particular en los individuos con características intermedias.

3-Determinar, apoyado en datos morfológicos y genómicos, si existen evidencias de hibridación entre las especies de agaves mezcaleros en la Sierra El Doctor.

## **Justificación**

La creciente demanda de agaves para la producción de mezcal y otros usos está ejerciendo una presión adicional sobre las poblaciones naturales de *A. montana*, *A. gentryi* y *A. salmiana*. A pesar de que, actualmente, no están categorizadas como especies amenazadas es necesario recopilar más información sobre ellas para prever y mitigar futuros desafíos de conservación, especialmente en el caso de *A. montana*, cuyas poblaciones son más escasas y cuyas flores son usadas intensamente como alimento en la Sierra El Doctor.

Además, este estudio tiene un valor significativo desde una perspectiva académica porque enriquece nuestro conocimiento sobre los procesos de especiación y las dinámicas evolutivas dentro de este género. Comprender cómo se originan y se mantienen las nuevas especies a través de la hibridación y la diversificación genética es fundamental para la biología evolutiva. La identificación de híbridos y la comprensión de la estructura genética nos permiten diseñar e implementar estrategias futuras de conservación efectivas. Esto no solo protege a las especies “puras”, que son esenciales para mantener la integridad genética y la biodiversidad, sino también a los híbridos potencialmente valiosos, que pueden poseer características únicas y adaptativas beneficiosas.

La información obtenida de este estudio podría tener implicaciones prácticas y económicas. En el contexto de la industria del mezcal, conocer las bases genéticas de las especies utilizadas puede llevar a prácticas agrícolas más sostenibles y eficientes, y asegurar la autenticidad y calidad del producto final. Esto, a su vez, beneficia tanto a las comunidades locales involucradas en la producción como a los consumidores que valoran un producto artesanal con un profundo significado cultural.

## **2. Marco teórico**

### **2.1 Hibridación en plantas**

La historia evolutiva de las plantas ha estado influenciada por la hibridación, que se define como la generación de individuos viables a partir del entrecruzamiento entre especies genéticamente diferenciadas (Abbott et al., 2013). Este proceso ha modulado el desarrollo del genoma vegetal en su conjunto, ya que 16 % de los géneros de plantas vasculares muestran indicios de hibridación, en una proporción de nueve especies híbridas por cada 100 especies no híbridas (Zelenin et al., 2016).

En su revisión, Abbott et al. (2013), señalan que las consecuencias de la hibridación pueden variar en dependencia de los taxones involucrados y el grado de divergencia de los mismos, así como de las condiciones ecológicas en las que se produzca la interacción. Por ejemplo, cuanto mayor sea la distancia genética entre las especies involucradas, es probable que se presenten mayores barreras reproductivas que derivan en la imposibilidad del cruce, una baja viabilidad de la descendencia o esterilidad de los híbridos. Esto se debe a que las diferencias genéticas pueden interferir con la correcta segregación y expresión de los genes en los híbridos.

El propio autor continúa explicando que las condiciones ecológicas en las que se produce la interacción entre especies también pueden afectar la probabilidad y la magnitud de la hibridación. Dos especies de plantas con distribución contigua podrían, desde el punto de vista genético, tener la potencialidad de hibridar. Sin embargo, aunque la hibridación sea genéticamente posible, puede ser ecológicamente improbable si, por ejemplo, habitan en

condiciones climáticas diferentes. La diferencia climática en hábitats contiguos podría impedir a los polinizadores establecer un flujo génico entre ellas. Además, ciertos cambios en el entorno, como la fragmentación del hábitat o la introducción de especies invasoras, pueden aumentar la frecuencia de hibridación (McFarlane y Pemberton, 2019). La hibridación es un proceso muy complejo y en todos los casos el patrón resultante que vemos en la actualidad es solo reflejo momentáneo de una interacción en constante evolución.

Torres-Suárez (2016), menciona que entre las ventajas que ofrece la hibridación está la facilitación de cambios evolutivos rápidos al permitir el intercambio de adaptaciones preexistentes. Además, el autor señala que este proceso también promueve la radiación adaptativa en nuevos hábitats cuando las condiciones ecológicas lo permiten. Puede contrarrestar la pérdida de diversidad genética que a menudo se experimenta debido al efecto fundador, generando genotipos novedosos. También es posible que ocurra el enmascaramiento de alelos recesivos nocivos y/o transferencia de alelos beneficiosos de especies localmente adaptadas a especies invasoras o viceversa. Es frecuente que los híbridos tengan una mayor capacidad para sobrevivir, crecer y/o reproducirse en comparación con cualquiera de sus parentales, lo cual se conoce como vigor híbrido (Qiao et al., 2019; Rius & Darling, 2014).

Por otro lado, las consecuencias de la hibridación también pueden ser negativas, en especial en temas relacionados con la conservación de especies amenazadas (Graham et al., 2021). Allendorf (2017), clasifican la hibridación por su causa en natural o antropogénica, considerando esta última como una amenaza para la "integridad genética" de las especies naturales. La hibridación antropogénica puede conducir a la extinción de una o ambas especies parentales, debido a una mayor aptitud biológica de la descendencia híbrida (el anteriormente mencionado vigor híbrido), resultando en un reemplazo del genotipo parental por uno de origen híbrido (inundación génica). Los híbridos F1, aun siendo estériles, podrían ser competidores efectivos, impactando negativamente a las especies parentales u otros miembros de la comunidad (Senanan et al., 2004). Además, puede ocurrir una reducción de la tasa de crecimiento poblacional de las especies parentales debido a la generación de descendencia híbrida inviable (Abbott et al., 2013).



En cuestiones de taxonomía la hibridación es problemática, pues los híbridos en ocasiones no se diferencian fácilmente de los padres y en otras ocasiones forman un gradiente morfológico entre ambos parentales (Baiakhmetov et al., 2021). En dependencia del concepto de especie utilizado para la delimitación taxonómica, ambas situaciones plantean un desafío a la clasificación. Esto finalmente puede tener consecuencias mayores cuando la gestión de la conservación exige un límite taxonómico entre grupos que hibridan (Fitzpatrick et al., 2015).

Como ya se mencionaba al ocurrir hibridación, los descendientes de la primera generación pueden o no ser fértiles (Edelman y Mallet, 2021). Cuando las plantas híbridas conservan la fertilidad, pueden ocurrir dos fenómenos:

1-Pueden establecerse como una nueva especie (Masuelli y Marfil, 2011), lo cual se conoce como especiación híbrida y ocurre cuando los híbridos desarrollan barreras reproductivas frente a las especies progenitoras (Kong y Kubatko, 2021). Esto sucede cuando el proceso de hibridación da lugar a novedades fenotípicas, como variaciones en el tiempo de floración, alteraciones en las estructuras florales y cambios en la fertilidad del polen, que llevan a un aislamiento reproductivo de los híbridos con respecto a las especies progenitoras y, por lo tanto, el híbrido se mantiene como una entidad taxonómica distinta.

2-En el segundo caso, los genes de una especie progenitora logran introducirse en el pool genético del otro progenitor a través del híbrido. Esto significa que la entidad híbrida se fusiona con uno de los padres a través de repetidos retrocruzamientos, lo que se conoce como introgresión (Kong y Kubatko, 2021).

En el caso particular del género *Agave* no es raro encontrar casos de hibridación e introgresión (Roll et al., 2021). Se ha sugerido que estos fenómenos son factibles debido a que las especies de *Agave* tienen orígenes recientes, lo que resulta en un desarrollo incipiente de los mecanismos de aislamiento reproductivo (Rocha et al., 2006). De acuerdo con Jiménez-Barron et al. (2020), la radiación del género *Agave* ocurrió entre 2.68 y 3.55 Ma atrás, por lo que constituyen un grupo relativamente joven.

Existen numerosos ejemplos de híbridos de este género en la literatura. Gentry (1982), presenta múltiples combinaciones híbridas de *Agave*, como *Agave*×*arizonica* (híbrido entre *A. chrysantha* y *A. toumeyana*). Pinkava y Baker (1985) sugieren que *A. arizonica* podría ser un híbrido entre *A. chrysantha* y *A. toumeyana* ssp. *toumeyana*. En el estudio de González-Elizondo et al. (2011), se proponen híbridos resultantes del cruce entre *Agave nickelsiae* y *A. lechuguilla*, *A. nickelsiae* y *A. asperrima*, *A. pintilla* y *A. salmiana*. Además, Villarreal-Quintanilla et al. (2014) proponen a *Agave* × *madrensis* como un posible híbrido. Castorena-Sánchez et al., 1991 registran un retrocruce del híbrido *A. amaniensis* Trel. & Now. x *A. angustifolia* con su parental *A. amaniensis*.

## 2.2 Genómica

La hibridación suele poder inferirse a partir de evidencias morfológicas y ser posteriormente confirmada mediante análisis genético (Boto, 2012). Sin embargo, en ocasiones los híbridos son irreconocibles morfológicamente de los parentales (Mallet, 2005) (o por el contrario desarrollan características morfológicas que los alejan demasiado de los parentales).

Más allá de los cambios morfológicos, estos procesos dejan una huella en el genoma que puede ser detectada por varios métodos como la citogenética o el uso de marcadores moleculares como isoenzimas, microsatélites o secuenciación de genes cloroplásticos. Si bien en el pasado se emplearon estas técnicas, la cantidad limitada de loci genéticos evaluados con ellas puede disminuir la precisión en la asignación de individuos a una especie y limitar la potencial para identificar híbridos con bajos niveles de mezcla (Graham et al., 2021).

En las últimas décadas, las técnicas de análisis genómico basadas en la secuenciación masiva o secuenciación de próxima generación han revolucionado el estudio de la diversidad biológica (López de Heredia, 2016). Su uso ha reducido los costos y el tiempo de los experimentos al secuenciar en paralelo millones de fragmentos de ADN en múltiples individuos (Schloss, 2008; Thareja et al., 2018). Estas tecnologías han transformado la forma en que se investigan las especies silvestres, superando las limitaciones de las técnicas genéticas tradicionales (López de Heredia, 2016).

En general el proceso de secuenciación masiva consta de dos fases: la fase *in vitro*, donde se construyen bibliotecas genómicas específicas y se secuencia el material genético, y la fase *in silico*, que implica el análisis bioinformáticos de los archivos de lecturas generados por las plataformas de secuenciación masiva (López de Heredia, 2016). La fase *in vitro* a su vez puede dividirse en tres pasos: (1) preparación de bibliotecas, (2) inmovilización sobre una superficie bidimensional y amplificación *in vitro* de la biblioteca y (3) secuenciación y captación de señal (López-Causapé et al., 2021).

La preparación de bibliotecas consiste básicamente en fragmentar y marcar las diferentes muestras de ADN cuya secuencia quiere conocerse. Actualmente, existen diferentes estrategias de fragmentación del ADN (física, enzimática o química) cuyo objetivo final es la obtención de una mezcla de fragmentos de un tamaño homogéneo, estando el tamaño deseado determinado por la tecnología de secuenciación empleada. Una vez fragmentado el ADN genómico se procede al marcaje para lo cual se añaden unos oligonucleótidos que funcionan como identificadores únicos. Además de identificar estos oligonucleótidos permiten la posterior fijación de los fragmentos a la superficie bidimensional donde tendrá lugar la reacción de secuenciación (López-Causapé et al., 2021).

Una vez preparadas las bibliotecas, éstas se fijan a una superficie bidimensional donde son amplificadas *in vitro* existiendo también en este punto diferentes estrategias. De todas ellas, la más sencilla es la utilizada por los secuenciadores de Illumina la cual es denominada amplificación puente (bridge amplification). En esta estrategia los cebadores necesarios para el inicio de la reacción de amplificación se encuentran inmovilizados sobre una superficie bidimensional; al ser estos cebadores complementarios a una parte de los identificadores incorporados en la preparación de bibliotecas, las muestras quedan inmovilizadas y en disposición de iniciarse la amplificación clonal en paralelo de todas las muestras. Otras estrategias alternativas incluyen la amplificación en emulsión o la amplificación en nanobolas (rolling-circle amplification).

Finalmente, una vez amplificados los fragmentos de ADN, se produce la reacción de secuenciación para la cual también se han desarrollado diferentes métodos. Uno de estos es la pirosecuenciación, este método utiliza nucleótidos cuya incorporación a la hebra complementaria en síntesis libera pirofosfatos, generando luz detectable para determinar la secuencia de nucleótidos. Otra estrategia utilizada es la denominada secuenciación por

síntesis (SBS, sequencing by synthesis) donde la reacción de síntesis implica la incorporación secuencial y reversible de nucleótidos marcados con fluoróforos distintos que impiden además la adición posterior de más nucleótidos. De esta forma, en cada paso se produce la adición de un único nucleótido a cada uno de los fragmentos en síntesis cuya señal fluorescente es captada y traducida a nucleótidos. Por último, otra aproximación ampliamente utilizada es la denominada secuenciación mediante ligación (SBL; sequencing by ligation) que no utiliza ADN polimerasas y cuya base reside en la utilización de una mezcla de sondas marcadas con fluoróforos que se unen mediante ligasas específicas. Es importante conocer la estrategia química de secuenciación en que se basa el secuenciador ya que determina en gran medida sus características y funcionalidades (López-Causapé et al., 2021).

Respecto a la segunda fase, la fase *in silico*, se debe evaluar la calidad de las lecturas obtenidas y eliminar las lecturas erróneas que se hayan generado durante la secuenciación. Los errores, pueden afectar a la totalidad de la lectura o solamente a una parte específica de las mismas siendo necesario filtrar y limpiar estos errores con herramientas como Prinseq-lite o Trimmomatic (López-Causapé et al., 2021). Dado que en el secuenciador introducimos una mezcla equimolar de un conjunto de muestras, se debe realizar el proceso de “demultiplexado”. Este proceso consiste en la clasificación del total de lecturas generadas en función de los identificadores flanqueantes en las mismas, de tal forma que, a partir de un fichero único de lecturas, se generarán tantos ficheros de lecturas como muestras se hayan secuenciado. O sea, se hace corresponder cada lectura con el individuo al que pertenece.

La disponibilidad de un genoma o transcriptoma de referencia adecuado sobre el cual alinear las lecturas procesadas es clave a la hora de realizar un alineamiento y su posterior análisis de variantes. En el caso de no disponer de un genoma de referencia adecuado para el alineamiento, puede usarse el ensamblado *de novo* (López-Causapé et al., 2021).

Aunque secuenciar genomas completos todavía es costoso, existen alternativas más económicas, como la secuenciación de representación reducida de regiones genómicas específicas, como Genotyping-by-Sequencing (GBS), Restriction site Associated DNA sequencing (RADseq) o Double Digest Restriction-site Associated DNA sequencing (ddRADseq). La representación reducida se refiere a capturar muchos segmentos

dispersos en todo el genoma, pero sólo cubre una fracción de la secuencia total del genoma. Esta técnica se desarrolló para estudios de mapeo genético en plantas de maíz (Elshire et al., 2011) y aunque se ha utilizado principalmente en especies agrónomicamente relevantes, también se ha aplicado en estudios de genética poblacional de especies de plantas silvestres (Chen et al., 2013; Schilling et al., 2014; Ilut et al., 2015; Ratcliffe et al., 2015).

La secuenciación de representación reducida es un método útil para descubrir y genotipificar grandes cantidades de SNP (polimorfismos de un solo nucleótido), que son variaciones en la secuencia de ADN que ocurren en una sola base (Gambini et al., 2016). Una de las ventajas del uso de esta técnica es que permite realizar análisis y comparaciones genómicas entre especies y dentro de ellas. Sin embargo, una de las limitaciones es el manejo del volumen de datos generado, lo cual requiere habilidades bioinformáticas avanzadas y una infraestructura computacional adecuada (López de Heredia y Vázquez-Poletti, 2016).

La técnica de secuenciación de representación reducida también es útil para investigar la hibridación e introgresión debido a la abundancia de SNPs en todo el genoma (en regiones codificantes y no codificantes) (Graham et al., 2021). La creciente disponibilidad de datos genómicos y el uso de herramientas analíticas más sofisticadas han facilitado el desarrollo de varios métodos para la identificación de señales de hibridación (Twyford y Ennos 2012). A pesar de estos avances la detección precisa de la hibridación a partir de datos genómicos sigue siendo una tarea difícil (Kong y Kubatko, 2021). Si la hibridación es reciente los patrones serán claros. Sin embargo, si ya han existido muchas generaciones de retrocruzamientos los loci que evidencian la hibridación se habrán reducido. Por lo tanto, detectar hibridaciones ancestrales resulta más difícil (Hibbins y Hahn, 2021).

### **2.3.1. Genómica de *Agave***

La secuenciación de representación reducida es un método particularmente útil cuando se trabaja con especies con genomas grandes (Ruiz-Mondragón et al., 2022), como *Agave* (Eguiarte et al., 2021). Recientemente se desarrolló una investigación donde se aplicó GBS para examinar la relación entre la asociación genética y la descripción morfológica de diferentes variedades de *A. angustifolia* (Cabrera-Toledo et al., 2022). Otro estudio

reciente se enfocó en analizar la manera en que factores como la distribución geográfica, la historia de vida y las actividades humanas han influido en los patrones de estructura genética y diversidad en *A. angustifolia* var. *pacifica* (Klimova et al., 2022). Además, se realizó un estudio que investigaron los patrones de diversidad genómica en *A. tequilana* en Jalisco, comparando individuos adultos y juveniles y evaluando los niveles de endogamia y conectividad entre los sitios estudiados (Ruiz-Mondragón et al., 2022).

### **2.3 Agaves mezcaleros de la Sierra El Doctor**

El mezcal es una de las bebidas alcohólicas más emblemáticas de México, con una larga tradición en su elaboración por grupos indígenas y campesinos a lo largo de los siglos. Sin embargo, en la segunda mitad del siglo pasado, se ha experimentado una inusual industrialización en su producción, lo que ha favorecido su comercio a nivel internacional. Para regularizar la producción y proteger su autenticidad, se creó el Consejo Regulador del Mezcal, que ha desempeñado un papel central en la implementación de denominaciones de origen (Valiente-Banuet, 2023).

El gobierno mexicano ha declarado las regiones de Guanajuato, Guerrero, Durango, Michoacán, Puebla, Oaxaca, San Luis Potosí y Tamaulipas, como "zonas de denominación de origen" para el mezcal, estableciendo normativas específicas que solo permiten que el mezcal producido en estas áreas y bajo ciertas especificaciones lleve esa denominación. Estos requisitos incluyen ser elaborado con agave mezcalero y seguir métodos de destilación tradicionales y artesanales (Valiente-Banuet, 2023).

El aumento de la demanda de mezcal ha llevado a una sobreexplotación de los agaves silvestres y sus semillas debido a la sobrecolección (Valiente-Banuet, 2023). Actualmente, los agaves mezcaleros de la Sierra El Doctor, que pertenecen a la Sección Salmianae y comprenden las especies *A. gentryi*, *A. salmiana* y *A. montana*, se encuentran clasificados por la Unión Internacional para la Conservación de la Naturaleza (UICN) como de "Preocupación menor" en términos de su estado de amenaza (IUCN, 2019). No obstante, es importante recopilar más información para abordar futuros desafíos de conservación, especialmente ante el posible aumento de la presión de extracción sobre las escasas poblaciones naturales de *A. montana*.

Además de su relevancia en la producción de mezcal, estos agaves tienen otros usos importantes. Por ejemplo, las flores de *A. montana*, conocido localmente como maguey chino, son utilizadas en la Sierra El Doctor para elaborar un plato muy popular, aunque el corte de la inflorescencia antes de la madurez afecta la regeneración de la población de estos agaves en la zona. También, la estética disposición de sus hojas ha llevado a un aumento en su uso con fines ornamentales.

En cuanto a *A. salmiana*, conocido localmente como maguey verde y *A. gentryi*, conocido localmente como maguey silvestre o maguey de cerro, ambos se usan para preparar aguamiel, que puede concentrarse para obtener miel de maguey (Flores-Morales et al., 2021). Los ejes de las inflorescencias maduras, conocidos como "quiotes", son consumidos como golosinas y representan una fuente importante de azúcares y almidón. Además, las flores al igual que en *A. montana*, se utilizan guisadas con huevo, las hojas se emplean como saborizante en la cocción de la barbacoa y su cutícula se usa para envolver diversos guisos (mixiote) (Alanís-Flores y González-Álvarez, 2011). Asimismo, se utiliza el pulque para elaborar salsas y como levadura en el procesamiento de pan. El tallo floral (quiote) de *A. gentryi* se usa para hacer cercas (Magallán, 1998).

### 3. Materiales y Métodos

#### 3.1. Área de estudio

La Sierra El Doctor se encuentra en el municipio de Cadereyta, en la zona centro-oriental del estado de Querétaro (99°31'35"- 99°42'00" longitud Oeste, 20°44'10"- 20°56'31" latitud Norte). Tiene una extensión aproximada de 19,000 ha y alcanza más de 3,200 msnm en sus zonas más altas (Bayona y Chávez, 2013). Este macizo montañoso está ubicado a manera de isla ecológica, separada de otros macizos montañosos de la Sierra Madre Oriental por el río Extoraz al Norte, el río Moctezuma al Este y valles al Sur y Oeste. De esta área se extraen minerales metálicos y mármol de alta calidad. La vegetación regional varía desde matorral submontano a bosques *Juniperus*, *Quercus*, *Pinus* y *Abies* (Steinmann et al., 2021). El clima es templado subhúmedo (Cwo) (García,

1964), con una precipitación anual de entre 600 y 1000 mm y una temperatura media anual de entre 10 y 12 °C (INEGI, 2013).

Dentro de La Sierra El Doctor consideraremos tres sitios para el muestreo (Figura 1). El sitio 1 corresponde al área del “Cerro del Espolón”, su altitud va desde 2900 hasta 3100 msnm, presenta bosque de *Abies* con una irradiación global de 60 %, combinado con áreas abiertas y zonas de mexical donde la irradiación global es cercana a 100 %. El sitio 2 corresponde a “Las Minas” con bosque mixto de *Juniperus*, *Quercus*, *Cercocarpus*, una altitud entre 2700-2900 msnm y 70 % de irradiación global. El sitio 3 es conocido localmente como “Cuesta Enxini” o “Cerro de la Cruz” corresponde a un bosque de *Quercus-Juniperus*, con una altitud 2800-2900 msnm y 70 % de irradiación global. Los tres sitios son relativamente próximos (aproximadamente entre dos y cuatro kilómetros), pero presentan diferencias ecológicas ligeras debido, principalmente, a las diferencias en el tipo de vegetación.

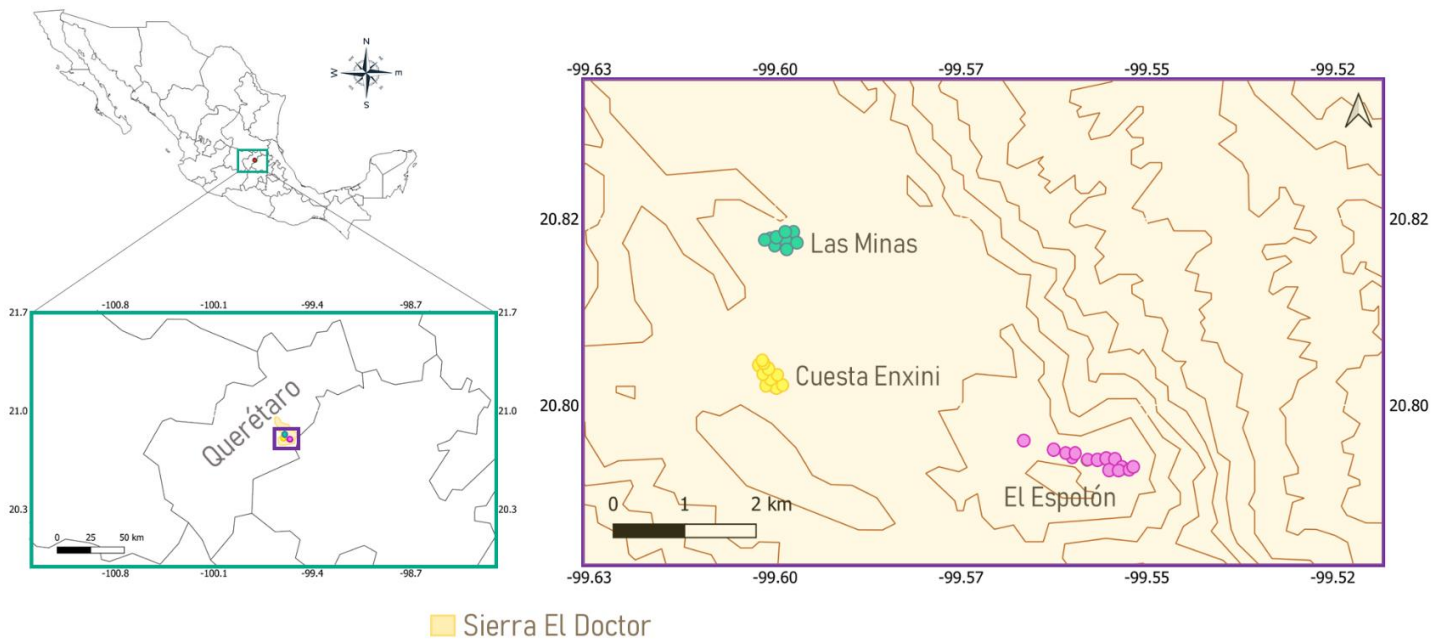


Figura 1. Zonas de muestreo dentro de la Sierra El Doctor, Cadereyta, Querétaro.

### 3.2. Análisis morfométrico

Se recopiló información morfológica de 114 individuos subadultos (individuos desarrollados que aún no alcanzan la etapa reproductiva), incluyendo 27 representantes



de *A. gentryi*, 27 de *A. salmiana*, 30 de *A. montana* y 30 individuos con características morfológicas intermedias (Anexo1). Entre los individuos clasificados como intermedios existió una gran heterogeneidad, pues algunos individuos presentaban afinidad fenotípica con *A. montana*, mientras que otros se asemejan con *A. gentryi* o *A. salmiana*. Estos individuos fueron clasificados como intermedios debido a que exhiben muchas características fenotípicas de una de las especies (*A. montana*, *A. gentryi* o *A. salmiana*), aunque no presentan todas las características morfológicas distintivas de esa especie. Además, muestran características que se combinan con las de otra de las especies mencionadas.

Se midieron de nueve a diez individuos de cada especie por sitio. La ubicación de los individuos se georreferenció y se tuvo en cuenta que los individuos dentro de cada grupo morfológico estuvieran separados al menos cinco metros entre sí, para minimizar la posibilidad de que fueran clones. Se midieron variables morfológicas que han demostrado ser útiles en el estudio y la caracterización agaves (Cabrera-Toledo et al., 2022): altura de la planta (cm), ancho de la roseta (cm), número de hojas, longitud de la hoja (cm), ancho máximo de la hoja (cm), color de la hoja (utilizando la escala de Munsell), longitud de la espina terminal (cm), ancho de la base de la espina terminal (cm), forma de la espina terminal y tipo de margen. Se midieron también variables ecológicas como altitud, tipo de suelo, porcentaje de irradiación global y pendiente del terreno.

Debido a que los datos no cumplieron con el supuesto de normalidad se aplicaron pruebas no paramétricas. Primeramente, para conocer si existían diferencias entre los grupos se realizó PERMANOVA por especie, utilizando un Escalamiento Multidimensional (MDS) para visualizar espacialmente los resultados. Posteriormente se aplicó una prueba Kruskal-Wallis, con la corrección de Bonferroni, para comparar las medianas por grupo, seguido de una prueba de comparaciones múltiples de Dunn. De igual forma se realizó un PERMANOVA comparando los tres sitios para cada especie por separado.

Se llevó a cabo un análisis de Componentes Principales (PCA), para analizar la contribución de las variables cuantitativas (altitud, porcentaje de irradiación global, pendiente, altura de la roseta, ancho de la roseta, número de hojas, longitud de hoja, ancho de la hoja, longitud de la espina terminal, ancho de la base de la espina terminal), además de las variables tipo de margen y forma de la espina terminal que fueron transformadas a

valores numéricos. Posteriormente, se realizó un Factor Analysis of Mixed Data (FAMD), que combina variables categóricas y cuantitativas. Este análisis inicia con el análisis de correspondencias múltiples para explorar las relaciones entre variables categóricas, seguido de un análisis factorial con las variables cuantitativas. Ambos resultados se integran en la visualización final.

Adicionalmente se realizó un análisis de clustering K-Medoids que también incluye datos cualitativos y cuantitativos, con el objetivo de identificar grupos sin clasificación *a priori* de las observaciones. Este análisis fue seleccionado porque a diferencia de K-means, que utiliza la media de las observaciones para representar el clúster, K-Medoids utiliza un medoide. Un medoide es un punto de datos dentro del clúster que minimiza la distancia total a todas las demás observaciones dentro del mismo clúster. Esto hace que K-Medoids sea más robusto ante distribuciones no gaussianas.

Todos los análisis se realizaron en R software, vers. 4.3.2 (R Core Team, 2023), los códigos utilizados aparecen en el Anexo 2. En el marco de estos análisis morfológicos a los individuos con características morfológicas intermedias se les denominó como *Agave* intermedio.

### **3.3. Extracción de ADN**

De los 114 individuos seleccionados para el análisis morfológico se tomaron muestras de hojas de 96 de ellos: 27 con morfología clásica de *A. montana*, 23 con morfología clásica de *A. salmiana*, 21 representantes típicos de *A. gentryi* y 25 de individuos con características intermedias. El tejido fue conservado en sílica gel hasta el momento de la extracción. La extracción de ADN se realizó siguiendo el protocolo de Doyle y Doyle (1987), modificado (Klimova et al., 2022).

Se comprobó la integridad del ADN genómico total utilizando una electroforesis en gel de agarosa al 1 %. La concentración se cuantificó mediante fluorimetría con un Qubit<sup>TM</sup>M. Para cada muestra de ADN, se aplicó una digestión utilizando una combinación de enzimas de restricción sensibles a la metilación (PstI/MspI). La elección de las enzimas estuvo basada en una estandarización anterior para *Agave salmiana* y *Agave lechuguilla* (Ruiz-Mondragon et al., 2022). La estandarización de la muestra y la

preparación de la biblioteca se realizó en el Centro de Biotecnología de la Universidad de Wisconsin (<https://biotech.wisc.edu/>). La secuenciación se llevó a cabo con la plataforma Illumina NovaSeq 2 × 150 (Illumina, Inc., San Diego, CA, EE. UU.).

### 3.4. Análisis Bioinformático

La puntuación media de calidad (Phred) para las 96 muestras secuenciadas fue de 38.60 y el contenido de guanina-citosina (GC) osciló entre 47 % y 48 %. Los datos en bruto se filtraron eliminando los adaptadores y las bases de baja calidad utilizando TRIMMOMATIC (Bolger et al., 2014), con los siguientes parámetros: ILLUMINACLIP (Nextera PE-PE.fa: 2:30:10), LEADING 5, TRAILING 10, SLIDINGWINDOW 4:20, MINLENGTH 75. Con estos valores se alcanzó un both surviving read de 90.93 %.

Las lecturas en formato FASTQ se demultiplexaron, se filtraron y los loci se ensamblaron usando ipyrad v. 0.9.79 (Eaton y Overcast, 2020), con parámetros recomendados para los datos ddRADseq de extremos pareados. Como aún no hay ningún genoma publicado de *Agave* disponible, utilizamos una estrategia de ensamblaje *de novo* y el transcriptoma de *Agave tequilana* (GAHU000000000.1; Gross et al., 2013).

Fueron descartados 4 individuos, correspondientes a dos individuos de *A. salmiana* y dos individuos de *Agave* intermedio debido a su bajo número de lecturas (Cuadro 1). Para el ensamble con transcriptoma resultaron 455,489 SNPs, mientras que para ensamble *de novo* resultaron 954,673 SNPs.

Cuadro 1. Número de individuos muestreados por especie y por sitio, después del filtrado por calidad y datos faltantes.

	sitio de muestreo			total
	Espolón	Minas	Enxini	
<i>A. montana</i>	10	9	8	27
<i>A. gentryi</i>	7	8	6	21
<i>A. salmiana</i>	7	7	7	21
<i>A. intermedio</i>	8	7	8	23

El filtrado de datos final se realizó utilizando VCFtools v.0.1.15 (Danecek et al., 2011). Se generaron archivos vcf para cada una de las especies por separado, para los pares especie-intermedios (que incluye a todos los intermedios), para las especies con los intermedios afines (que solo incluye a los intermedios con mayor similitud morfológica a esa especie) y para el conjunto total de individuos. Solo se retuvieron loci con una profundidad mínima media mayor a 12 por sitio, profundidad máxima media menor a 200 y un máximo de 2 alelos sin InDels (sin inserciones ni deleciones). Se mantuvieron los sitios con no más del 20 % de datos faltantes (máximo 0.8) y frecuencia alélica mínima (MAF) de 0.05 para los sets de datos que solo incluían una especie y MAF= 0.02 para los sets que incluían dos o más especies. Luego se filtraron las variantes que se desviaban significativamente del equilibrio de Hardy-Weinberg ( $p \leq 0.001$  después de la corrección de Bonferroni). Finalmente, para mejorar los efectos de confusión del desequilibrio de ligamiento (LD), se eliminaron los marcadores dentro de la distancia especificada entre sí utilizando Plink v1.9 (Purcell et al., 2007) 100: tamaño máximo de la ventana en variantes: 100, paso entre dos ventanas adyacentes: 10, valor de correlación  $r^2$  máximo entre dos variantes: 0.35.

A los archivos vcf derivados del ensamble *de novo* se le aplicaron los mismos filtros. No obstante, se optó por no seguir utilizando estos datos debido a que los resultados de análisis posteriores no mostraron patrones definidos, lo que sugiere una errónea formación de cluster al momento de la obtención de genotipos. Esto se debe probablemente a errores en el ensamblaje producidos por la alta variabilidad y complejidad intrínseca del set de datos donde los individuos provienen de especies diferentes.

### **3.5. Diversidad y estructura genómica**

Para conocer si los posibles híbridos presentan mayor diversidad que sus parentales estimamos la heterocigosidad esperada ( $H_e$ ), heterocigosidad observada ( $H_o$ ) y coeficiente de endogamia ( $F_{IS}$ ) utilizando los paquetes R v4.3.2 R Core Team, 2023 "adegenet" v2.1 (Jombart y Ahmed, 2011), "hierfstat" v0.4-22 (Goudet y Jombart, 2015) y Genepop v4.6 (Rousset, 2008).

También estimamos la diferenciación por pares ( $F_{ST}$ ) entre las especies y las supuestas poblaciones híbridas, y evaluamos la significancia estadística utilizando la función `gstat.randtest` del paquete "hierfstat" en 10.000 simulaciones.

Se realizó un análisis de coordenadas principales (PCoA) utilizando el paquete "ade4" para examinar el patrón de agrupamiento de los individuos. Se utilizó la variable "especie" como factor de análisis cuando el conjunto de datos contenía múltiples especies, o la variable "sitio" cuando se trataba de una sola especie. Estimamos distancias euclidianas (Edwards, 1971) de las frecuencias genéticas y obtuvimos un dendrograma UPGMA.

Utilizamos el programa Structure v2.3.4 (Pritchard et al., 2000; Falush et al., 2003) para inferir la estructura genética en el conjunto de datos con un modelo probabilístico que asigna a los individuos a diferentes grupos o variantes genéticas. Structure estima la proporción de ascendencia de cada individuo en cada uno de los grupos genéticos inferidos. Probamos de uno a ocho grupos (K) con 15 réplicas para cada una y con un total de 200 000 iteraciones MCMC.

Utilizamos FINERADSTRUCTURE (Malinsky et al., 2018) para asignar individuos a grupos, con un total de 1 000 000 de iteraciones MCMC y un período de burn-in de un millón, tomando muestras cada 10 000 iteraciones. Además, construimos un árbol con 100 000 iteraciones. Los resultados fueron visualizados utilizando 'FINERADSTRUCTUREPLOT.R' (versión R. v. 4.3.2, R Core Team, 2023).

## **4. Resultados**

### **4.1 Morfología**

En el cuadro 2 se presentan los datos morfológicos ya resumidos mostrando las variables que tuvieron diferencias significativas entre especies.

Cuadro 2. Comparación de medias de variables morfológicas entre las especies de agaves mezcaleros de la Sierra El Doctor.

	<i>Agave montana</i>	<i>Agave gentryi</i>	<i>Agave salmiana</i>	<i>Agave intermedio</i>
Altura de la roseta	101,9±28,5	136,5±29,6	175,4±38,3	130,3±31,1
Ancho de la roseta	161,5±38,9	199,6±34,3	260,8±36,9	207,7±54,3
relación longitud/ancho de la roseta	0,6	0,7	0,7	0,6
No. de hojas	51.6±33,0	28,0±7,4	38,1±10,1	49.8±23,9
Longitud de hoja	56,3±13,6	85,6±17,0	109,4±14,0	79,2±22,9
Ancho de la hoja	16,3±5,7	24,8±3,5	26,4±2,9	25,5±5,5
relación longitud/ancho de la hoja	3,0	3,6	4,2	3,1
Longitud de la espina terminal	4,4±1,0	5,5±1,0	6,0±1,0	5,0±1,2
Ancho de la base de la espina terminal	0,8±0,2	0,7±0,2	0,7±0,2	1,1±1,0

En la Figura 2 podemos visualizar espacialmente los resultados del PERMANOVA por especie en un gráfico MDS (Escalamiento Multidimensional). Se destaca la divergencia entre *Agave montana* y las otras especies. En contraste *A. gentryi* y *A. salmiana* no pueden separarse fácilmente con este análisis, mientras que los presuntos híbridos aparecen dispersos en toda la distribución. La Figura 3 muestra los resultados de la prueba Kruskal-Wallis para cada variable que resultó significativa en el PERMANOVA. La prueba de Dunn encontró diferencias entre *A. montana* y las restantes especies, incluido *Agave intermedio*, para todas estas variables. Igualmente, para casi todas las variables no existió una diferencia entre *A. gentryi* y *Agave intermedio* y entre *A. salmiana* y *Agave intermedio*.

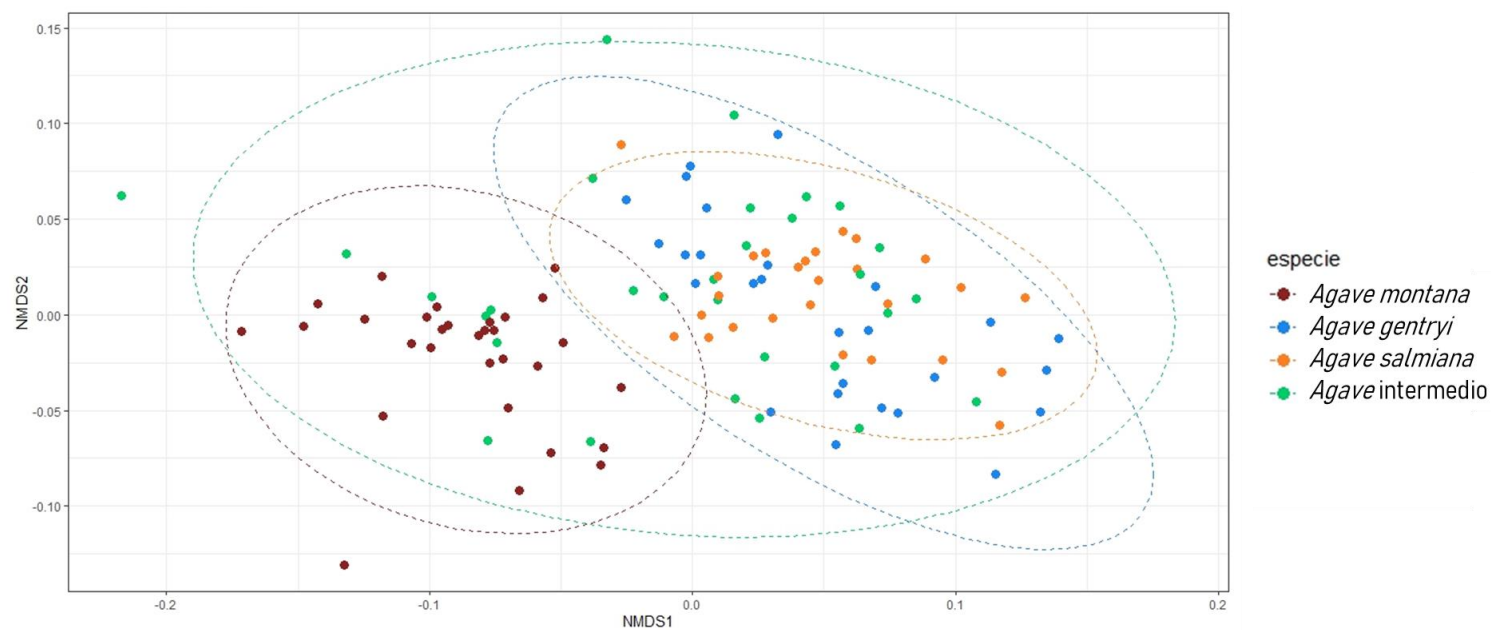
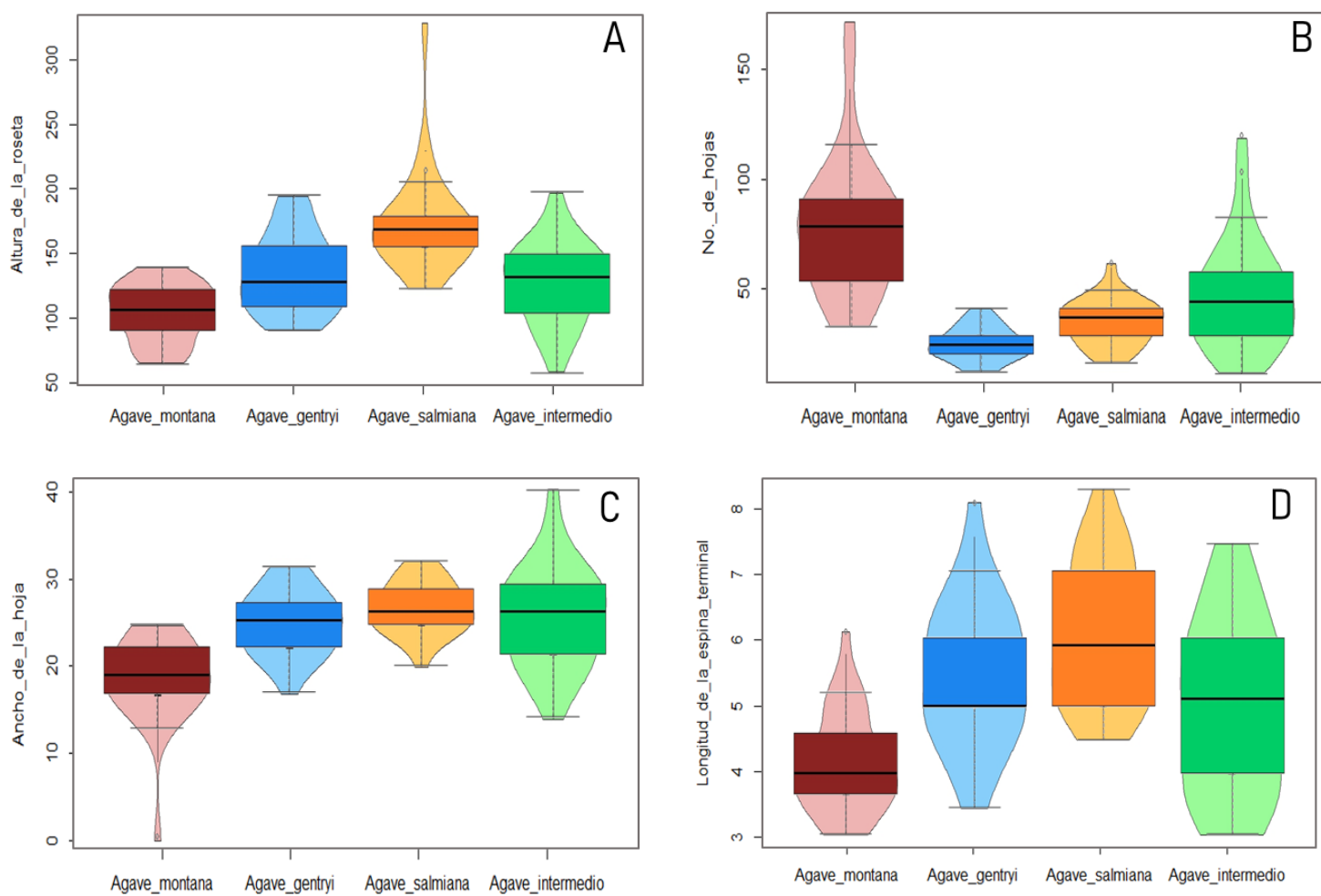


Figura 2. MDS con variables morfológicas cuantitativas por especies



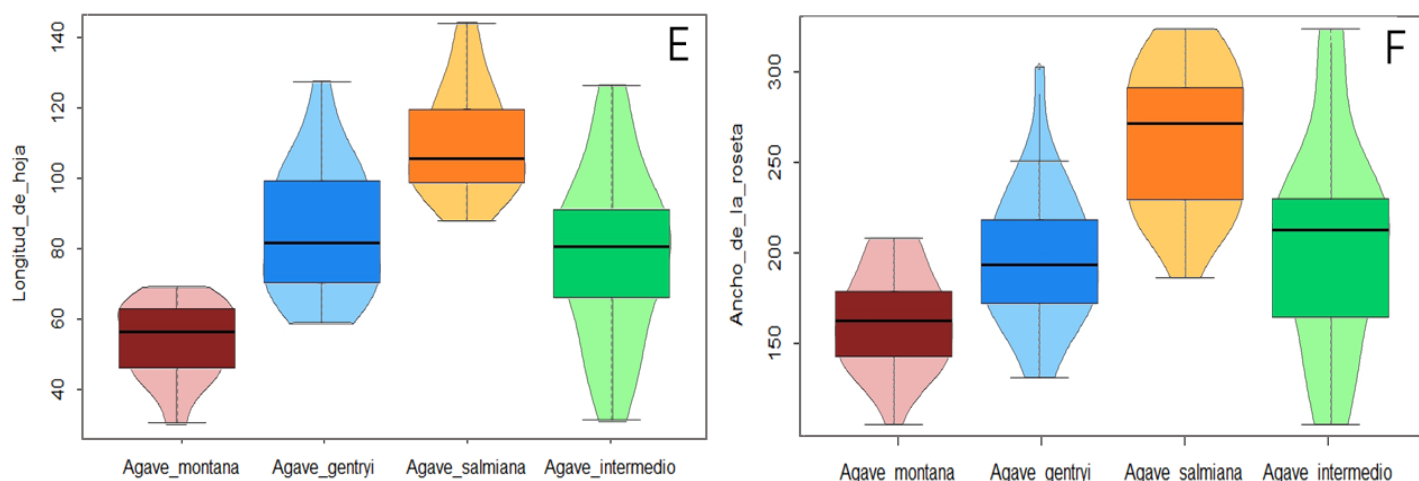


Figura 3: Prueba Kruskal- Wallis de variables morfológicas por especie. A. Altura de la roseta, B. Número de hojas, C. Ancho de la hoja, D. Longitud de la espina terminal, E. Longitud de la hoja, F. Ancho de la roseta.

Se observaron diferencias estadísticamente significativas entre los sitios para *Agave montana* y *A. gentryi* de acuerdo con los resultados del análisis PERMANOVA, mientras que no se encontraron diferencias significativas para *A. salmiana* y *Agave intermedio* (Figura 4). En el caso de *A. montana*, los resultados de las comparaciones múltiples de Dunn entre sitios encontraron diferencias entre Las Minas y Cuesta Enxini en relación al número de hojas, así como entre El Espolón y Las Minas en cuanto al ancho de la roseta. Con respecto a *A. gentryi* existieron diferencias entre El Espolón y Las Minas para las variables ancho de la roseta, altura de la roseta y longitud de la hoja, así como diferencias en el ancho de hoja entre Cuesta Enxini y los otros dos sitios.



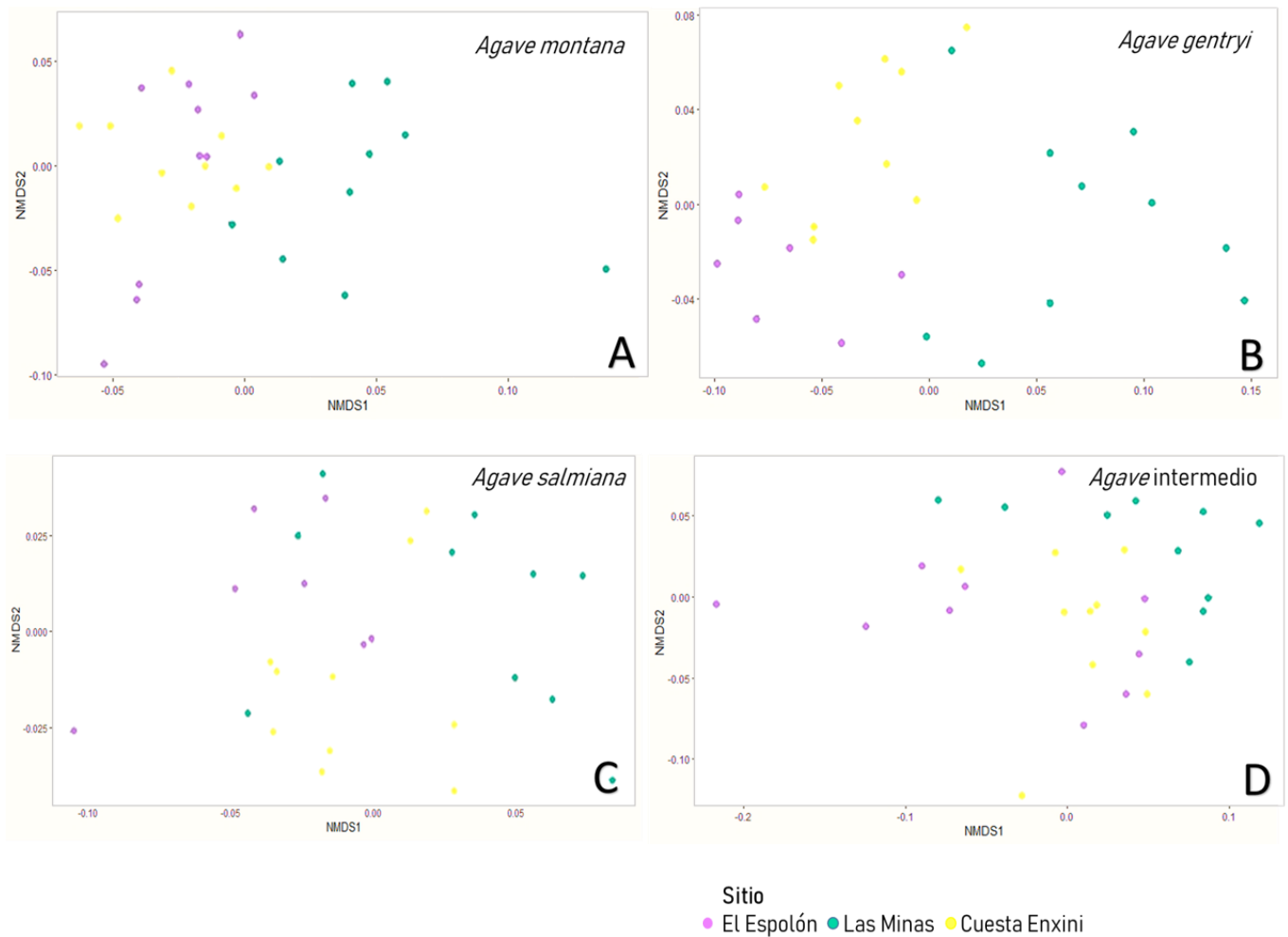


Figura 4. MDS con variables morfológicas cuantitativas por sitio. A. *Agave montana*. B. *Agave gentryi*. C. *Agave salmiana*. D. *Agave intermedio*.

La Figura 5 muestra el análisis de componentes principales. El primer componente logra explicar 37 % de la variación total (Figura 5a). Las variables que más contribuyeron a este primer componente son: longitud de la hoja, ancho de la roseta, forma de la espina terminal, porcentaje de irradiación global y ancho de la hoja (Figura 5b). Cuando representamos a cada individuo en el PCA se puede apreciar que *Agave intermedio* se encuentra en una posición intermedia en el espacio de rasgos morfológicos en comparación con los posibles parentales (Figura 5c).

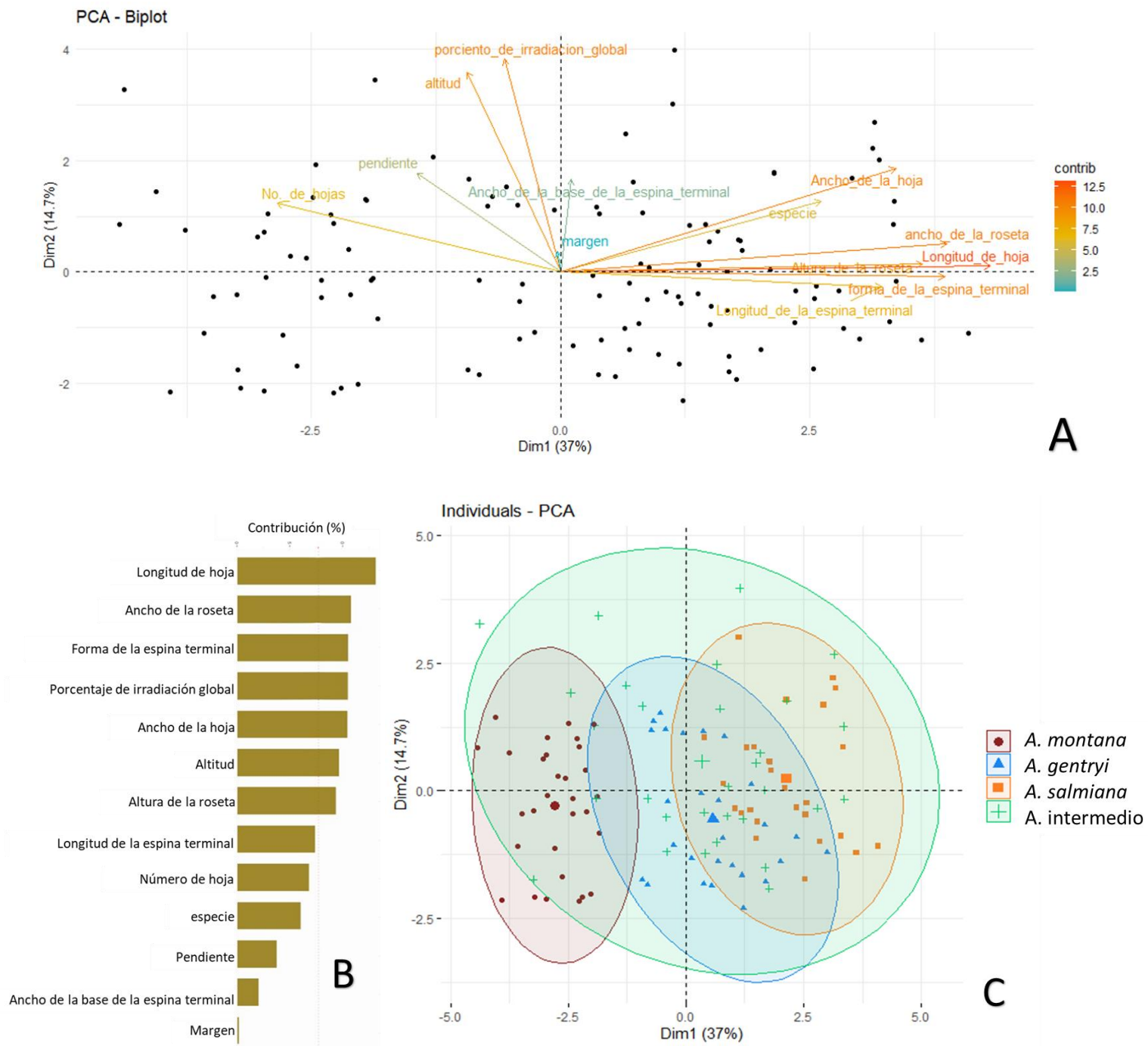


Figura 5. Análisis de componentes principales. A. se muestra la contribución de cada variable. B. contribución de cada variable. C. se muestra el agrupamiento de cada individuo asignado a una especie *a priori*.

La aplicación del Análisis de Correspondencia Múltiple Factorial (FAMD) combina variables cualitativas y cuantitativas (Figura 6) y proporciona una visión detallada del agrupamiento por especie. *Agave montana* aparece más separado del resto, mientras que *A. gentryi* y *A. salmiana* se encuentran más superpuestos. Los híbridos, nuevamente, se dispersan en toda la distribución.

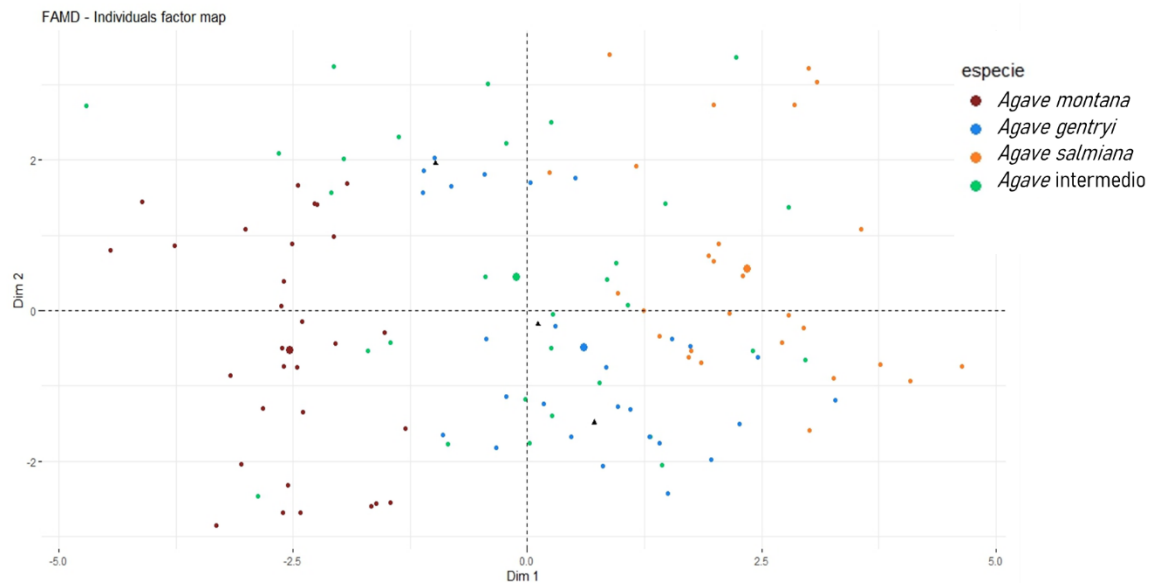
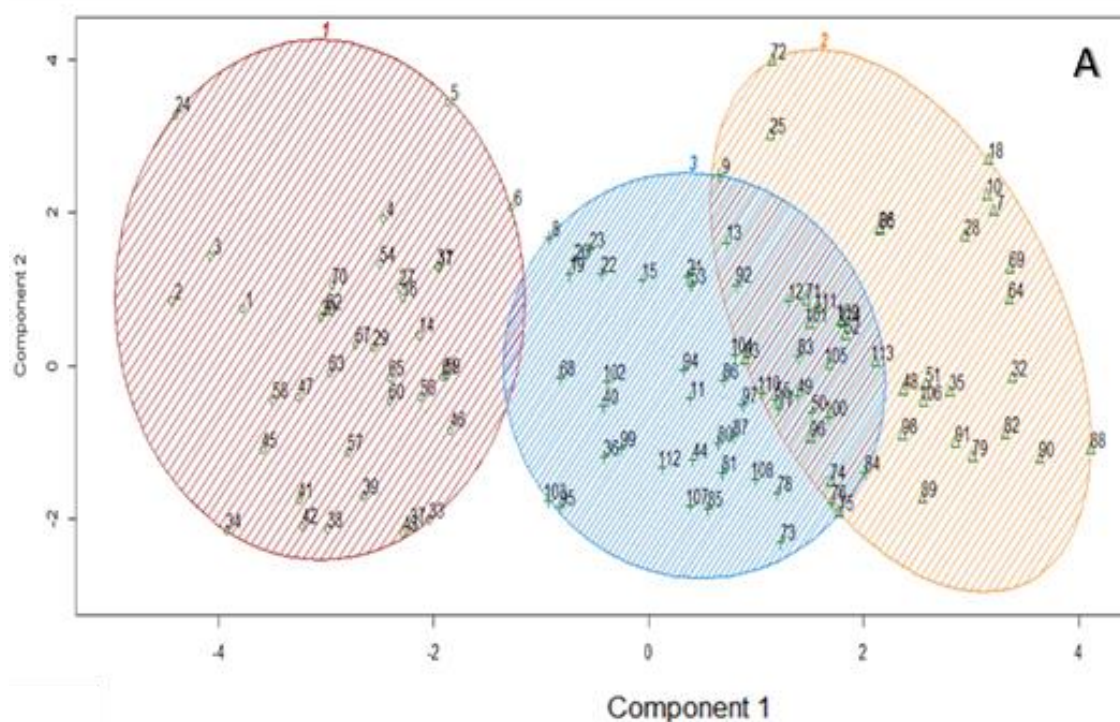


Figura 6. FAMD con datos morfológicos donde se muestra el agrupamiento de cada individuo asignado a una especie *a priori*.

El análisis de clustering K-Medoids agrupa a individuos que comparten características similares a partir de un conjunto de variables tanto cualitativas como cuantitativas (Figura 7). Esta agrupación no sigue las etiquetas que *a priori* fueron colocadas a los individuos, por lo que cada individuo puede asignarse a un cluster que coincida o no con su etiqueta original. El cluster 1, sombreado en rojo, corresponde a *A. montana*, el cluster 2, sombreado en naranja, correspondiente a *A. salmiana* y el cluster 3, sombreado en azul, corresponde a *A. gentryi* (Figura 7A). En la matriz de confusión (Figura 7b) se observa que todos los individuos con etiqueta de *A. montana* fueron correctamente asignados, mientras que dos individuos etiquetados como *A. gentryi* fueron incluidos en el cluster de *A. salmiana* (cluster 2) y siete individuos etiquetados como *A. salmiana* fueron asignados al cluster de *A. gentryi* (cluster 3, sombreado de azul). Los individuos intermedios, una vez más, están distribuidos en los tres clusters.



Estos dos componentes explican 51.71 % de la variabilidad

	<i>A. montana</i>	<i>A. salmiana</i>	<i>A. gentryi</i>	<i>Agave</i> intermedio
1	30	0	0	7
2	0	20	2	12
3	0	7	25	11

**B**

Figura 7. A. Análisis de clustering K-Medoids con datos morfológicos, k=3. B. Matriz de confusión.

## 4.2 Diversidad y estructura genómica

En el Cuadro 3 se registra el número de SNPs contenido en archivo VCF final derivado del ensamble con transcriptoma para cada especie. En el conjunto de *A. montana* se identificaron más SNPs en comparación con los conjuntos de *A. gentryi* y *A. salmiana*. El conjunto que abarca a todos los individuos tiene menos SNPs que la mayoría de las combinaciones que incluyen intermedios, a pesar de haberse aplicado los mismos filtros

en el análisis. A partir del agrupamiento natural observado en los resultados del análisis de PCA y UPGMA para el conjunto de todos los individuos (figura 8 y 9), se decidió formar combinaciones adicionales: un conjunto compuesto por *A. montana* y sus intermedios afines, y otro compuesto por *A. salmiana*, *A. gentryi* y sus intermedios afines.

Cuadro 3 Número de SNPs obtenidos a partir del filtrado con vcftools.

Set	número de SNPs	N
todos	10,830	92
<i>A. montana</i>	12,039	27
<i>A. gentryi</i>	5,104	21
<i>A. salmiana</i>	8,276	21
<i>A. montana</i> +todos los intermedios	15,056	50
<i>A. gentryi</i> +todos los intermedios	11,677	44
<i>A. salmiana</i> +todos los intermedios	11,079	44
<i>A. montana</i> +intermedios afines	11,285	37
<i>A. gentryi</i> + <i>A. salmiana</i> +intermedios afines	8,093	54

En el dendrograma construido a partir de una matriz de distancias euclidianas se forman dos grandes grupos: uno representado por *A. montana* y los intermedios afines a *A. montana* y otro clado formado por *A. gentryi*, *A. salmiana* y los intermedios afines a ambas especies (Figura 8). Las etiquetas están formadas por las iniciales de la especie, seguido del número del sitio de muestreo (1: Cerro del Espolón, 2: Las Minas, 3: Cuesta

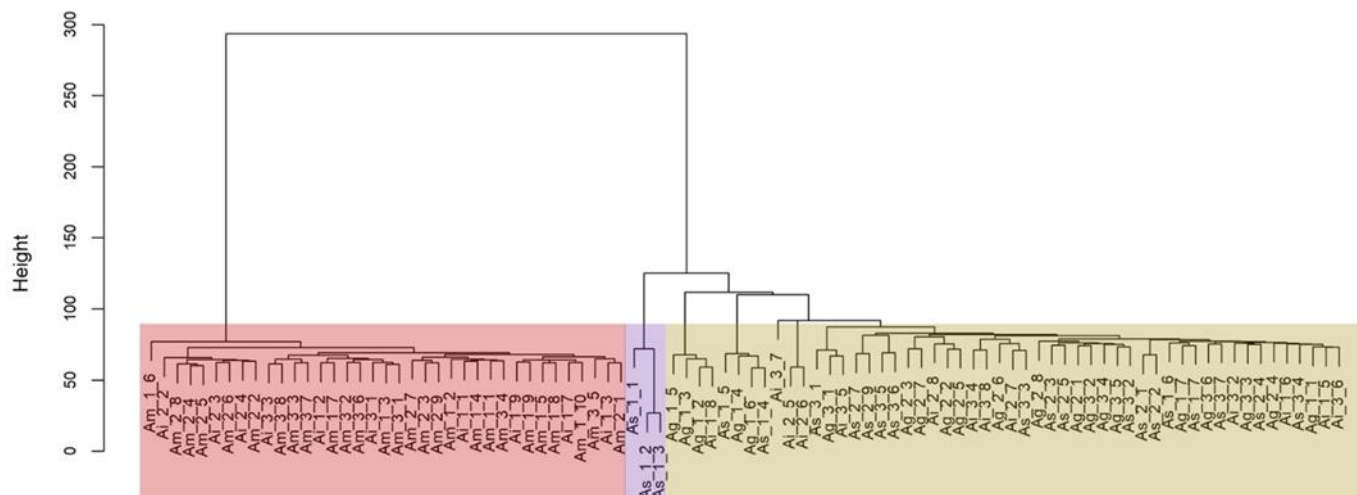


Figura 8. El dendrograma muestra la relación de similitud entre los individuos, donde las ramas más largas representan una mayor distancia entre los individuos. La distancia entre los individuos se calculó utilizando la métrica de distancia euclidiana y se agruparon utilizando el método de agrupamiento jerárquico con el método de Ward. (Am: *Agave montana*, Ag: *Agave gentryi*, As: *Agave salmiana*, Ai: *Agave intermedio*).

Enxini) y finalizan con el número del individuo. El factor sitio de muestreo no parece contribuir de forma importante en el patrón de agrupamiento de los individuos.

El análisis de componentes principales muestra dos grandes clusters que coinciden como en el análisis anterior con *A. montana* y sus intermedios afines y otro grupo formado por *A. salmiana*, *A. gentryi* y sus intermedios afines (Figura 9).

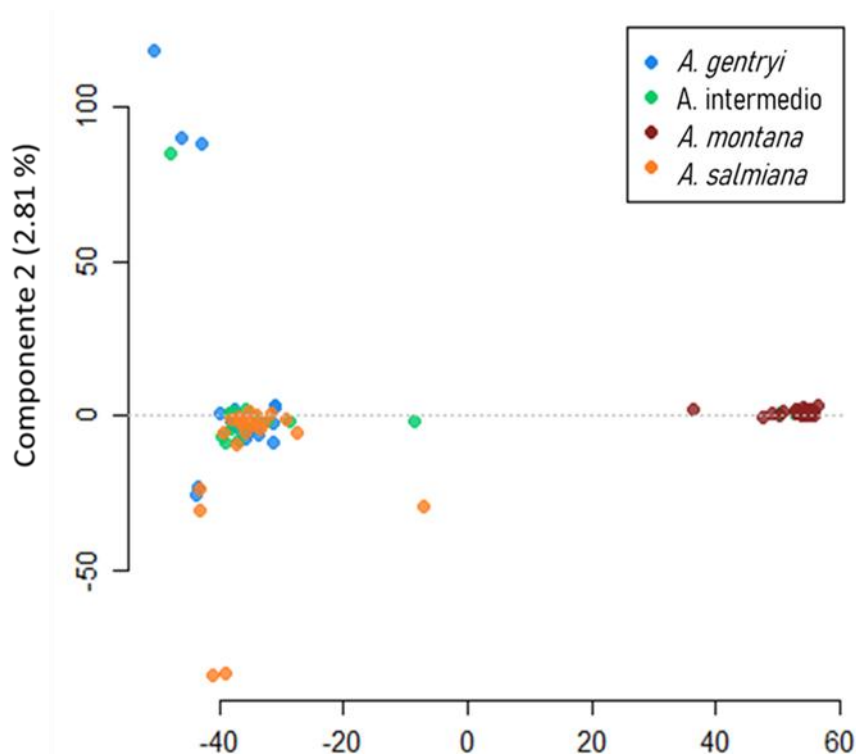


Figura 9. Análisis de componentes principales de los 92 individuos. Los primeros dos componentes explican 11.92 % de la variabilidad de los datos

Al calcular los índices de diversidad obtuvimos que la población de *A. montana*, así como la combinación de *A. montana* con sus intermedios afines presentan un déficit de heterocigotos y un coeficiente de endogamia positivo. Por el contrario, para el caso *A. gentryi*, *A. salmiana* y la combinación de *A. gentryi*, *A. salmiana* y sus intermedios afines hubo un exceso de heterocigotos y un  $F_{IS}$  negativo (cuadro 4).

Cuadro 4. Diversidad genética estimada usando SNPs.  $H_o$ : heterocigosidad observada,  $H_s$ : heterocigosidad esperada,  $F_{IS}$ : coeficiente de endogamia,  $F_{ST}$ : coeficiente de diferenciación genética, N: tamaño de muestra.

índice de diversidad	set completo	<i>A. montana</i>	<i>A. gentryi</i>	<i>A. salmiana</i>	<i>A. montana</i> +intermedios afines	<i>A. gentryi</i> + <i>A.salmiana</i> +intermedios afines
$H_o$	0.143	0.244	0.238	0.233	0.207	0.144
$H_s$	0.135	0.298	0.217	0.210	0.246	0.129
$F_{IS}$	-0.056	0.180	-0.100	-0.109	0.156	-0.114
$F_{ST}$	0.052	0.008	0.010	0.016	-0.001	0.001
N	92	27	21	21	37	54

Mientras mayor sea el valor de  $F_{ST}$  indicará una mayor diferenciación genética entre las poblaciones de las especies comparadas. Los valores de  $F_{ST}$  obtenidos sugieren una mayor diferenciación entre los pares *A. montana* y *A. gentryi*, así como entre *A. montana* y *A. salmiana* (cuadro 5).

Cuadro 5. Coeficiente de diferenciación genética pareado ( $F_{ST}$ )

	<i>A_gentryi</i>	<i>A_intermedio</i>	<i>A_montana</i>	<i>A_salmiana</i>
<i>A_gentryi</i>	NA			
<i>A_intermedio</i>	0.030	NA		
<i>A_montana</i>	0.150	0.052	NA	
<i>A_salmiana</i>	0.004	0.031	0.149	NA

En el análisis de STRUCTURE, para el conjunto de datos de 92 individuos, obtuvimos que el valor de K de mayor probabilidad correspondía con K=2. En este análisis, K representa el número de grupos genéticos distintos en los que se pueden dividir los individuos según sus patrones de variación genética. Es una estimación del número de poblaciones ancestrales o subpoblaciones dentro de la muestra analizada. Este resultado se obtuvo utilizando el método de Evanno et al. (2005), que ayuda a identificar el valor de K que mejor se ajusta a los datos. Para K=2 se agruparon los individuos de *A. montana* con los intermedios afines en una única variante genética sin mezclas (en rojo). Mientras que *A. salmiana*, *A. gentryi* y los intermedios afines (en mostaza) se presentaron como una variante genética única, con un grado mínimo de mezcla de la variante *A. montana* en algunos casos (K=2). Al probar K=3 y K=4, *A. montana* y los intermedios afines permanecieron como una variante única y sin mezclas. Por su parte, *A. gentryi* y *A. salmiana* no se presentaron como variantes genéticas distintas, como se esperaba dado que son especies diferentes, sino que continuaron apareciendo como una variante genética común (Figura 10).

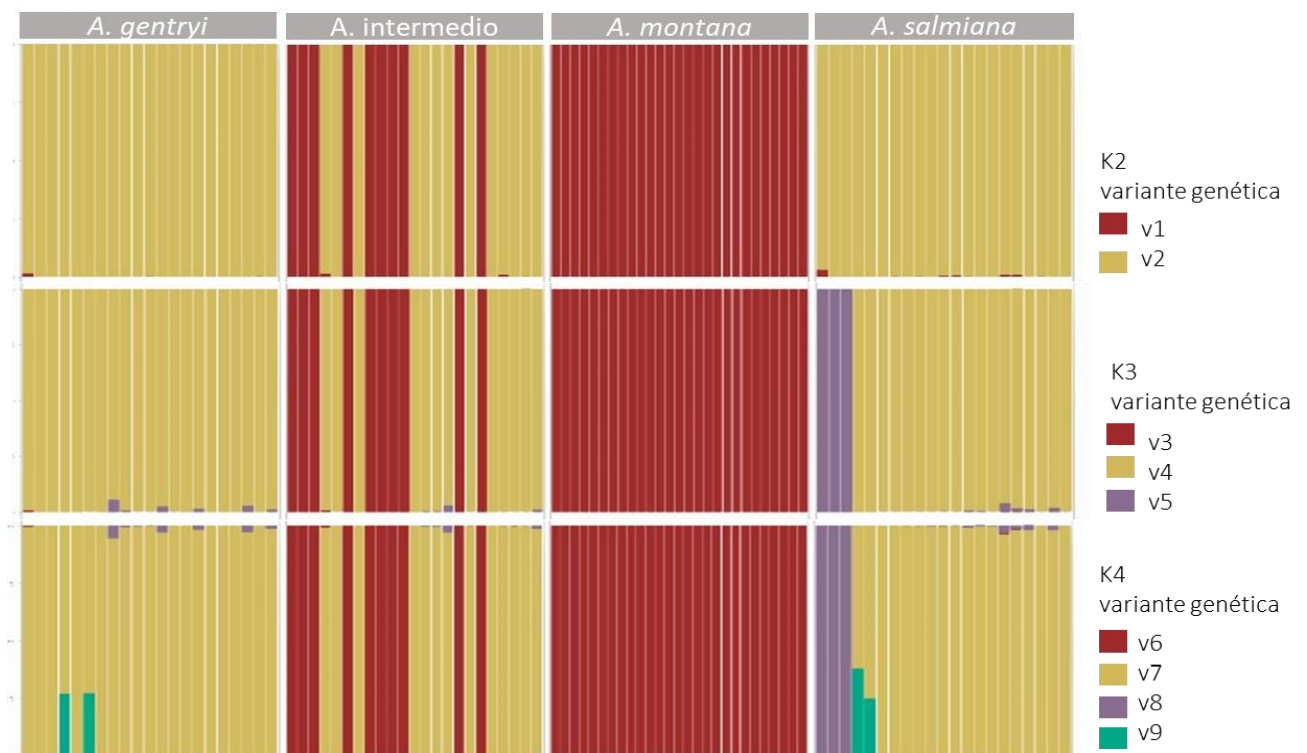


Figura 10. Análisis Stucture (K=2, K=3, K=4), para el conjunto de 92 individuos. El menor valor de k obtenido fue K=2.



El análisis de coancestría realizado mediante FineRadStructur revela la formación de dos grandes clusters. El primer cluster agrupa a *A. montana* y sus intermedios afines, indicando una alta relación de coascendencia dentro de este grupo. El segundo cluster incluye a *A. salmiana*, *A. gentryi* y sus intermedios afines, mostrando igualmente una fuerte relación de coascendencia dentro de este grupo. Estos resultados sugieren que hay dos grupos principales con patrones distintos de ascendencia y posible diferenciación genética entre ellos (Figura 11).

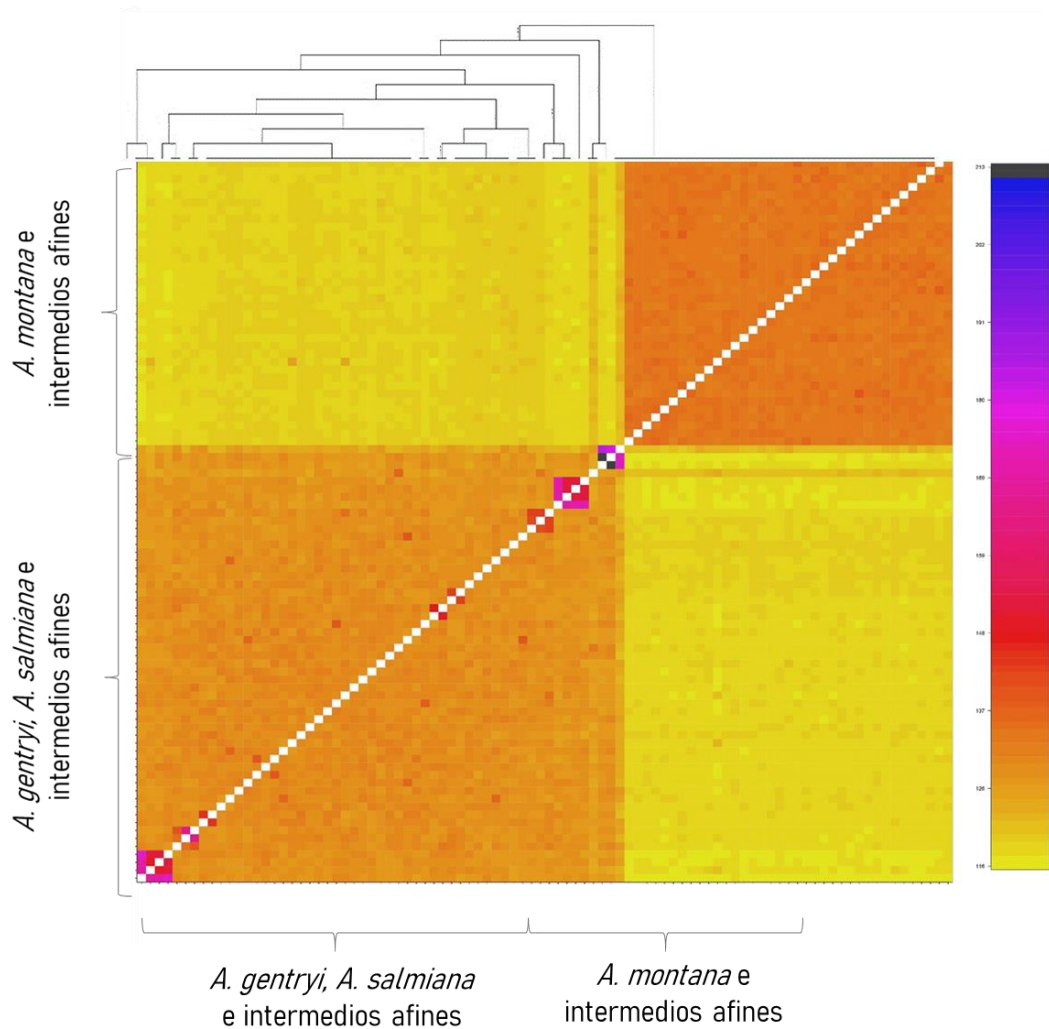


Figura 11. Matriz de coancestría FineRadStructure que incluye a los 92 individuos. Cuanto más oscuro es el cuadrado, mayor es la similitud genética entre un par de individuos. Las etiquetas marcadas con rojo agrupan a los individuos de *A. montana* y los intermedios afines, mientras que las etiquetas mostaza corresponden a *A. gentryi*, *A. salmiana* y sus intermedios afines.

En el análisis de *A. montana*, tanto el PCA, el dendrograma como el análisis de Structure mostraron consistencia entre sí, revelando una mayor proximidad entre los individuos de un mismo sitio de muestreo (Figuras 12, 13).

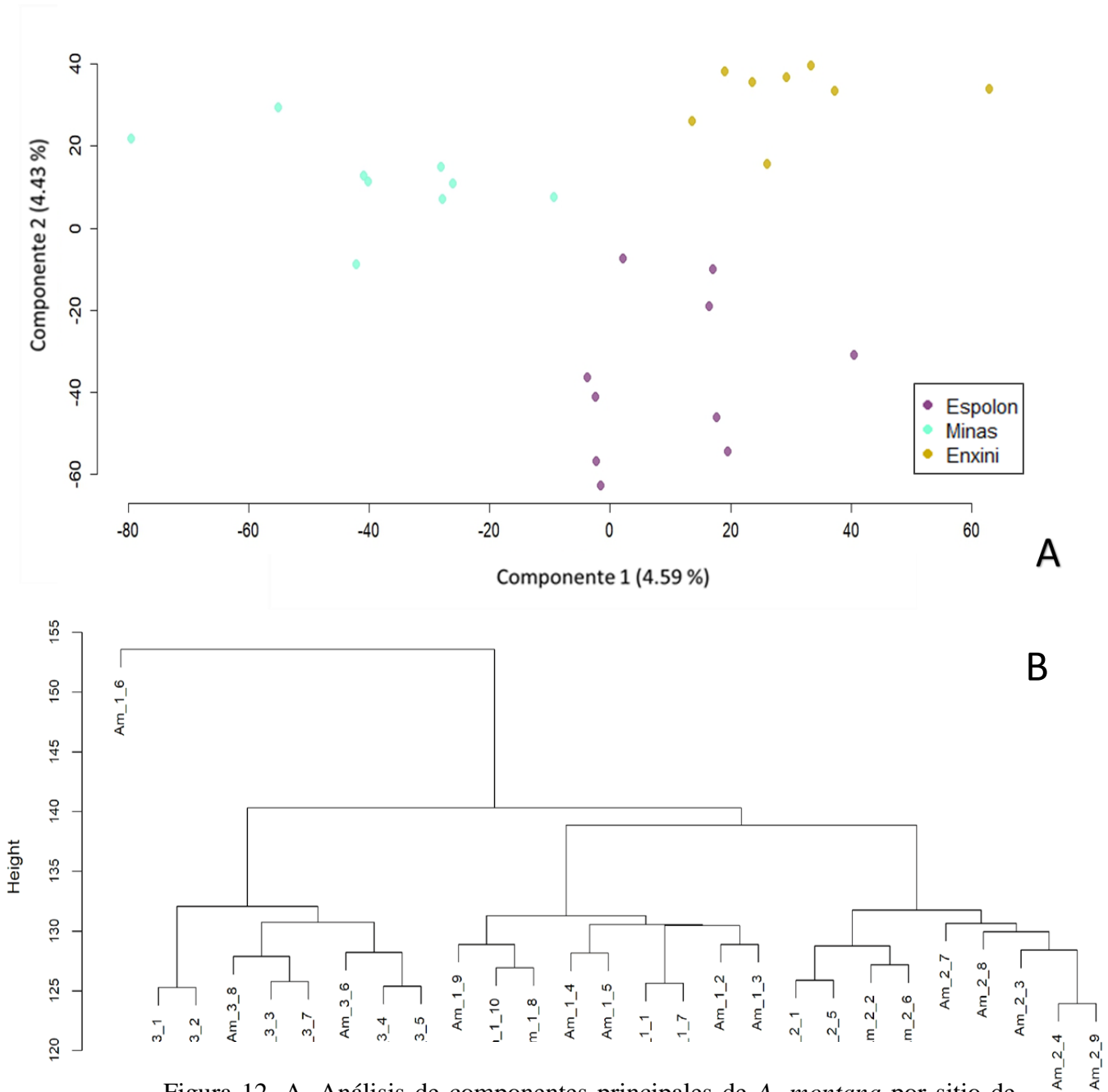


Figura 12. A. Análisis de componentes principales de *A. montana* por sitio de muestreo. B. Dendrograma de los 27 individuos de *A. montana*.

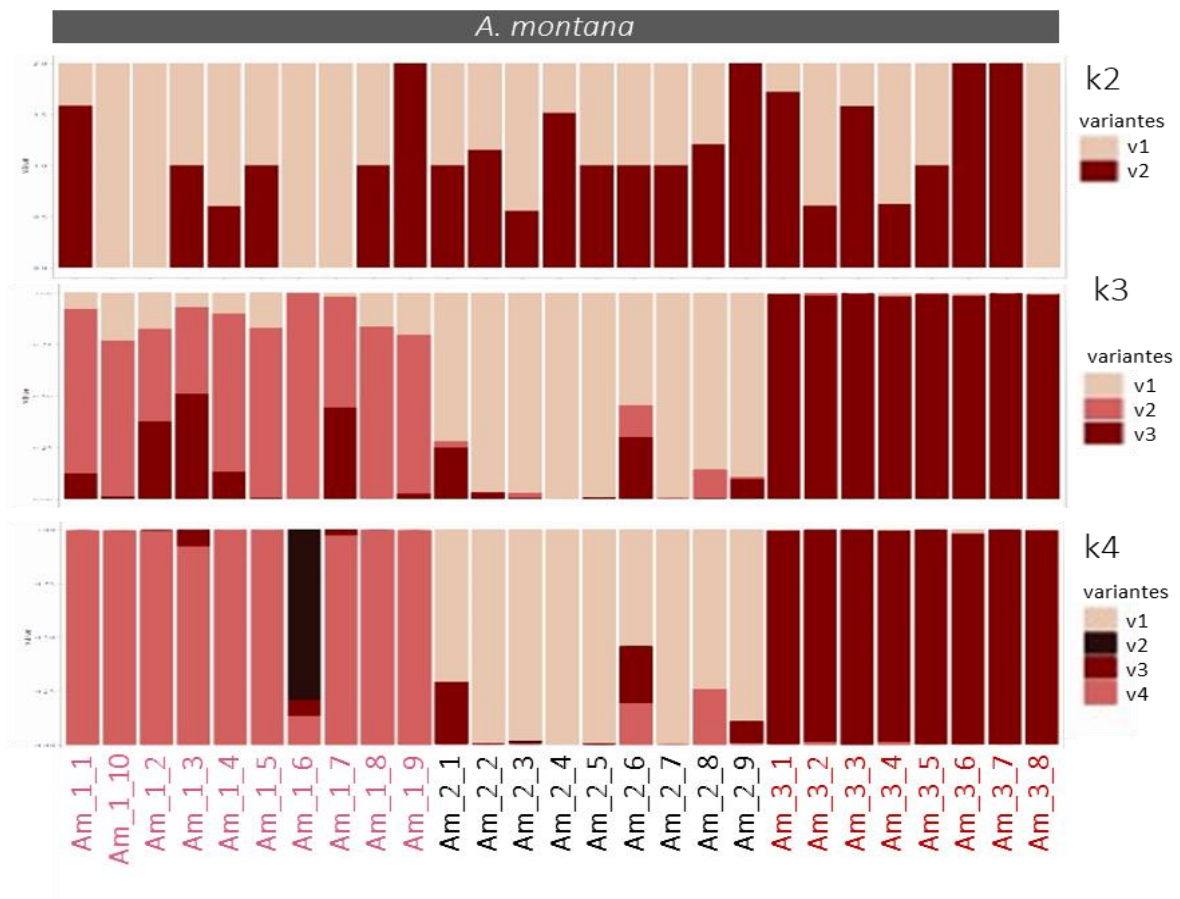
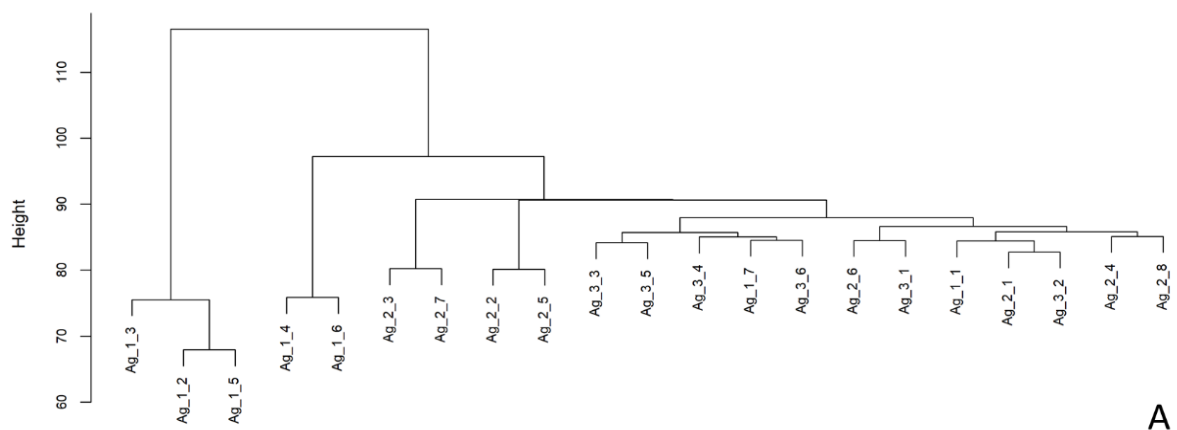


Figura 13. Análisis Structure de los 27 individuos de *A. montana* muestreados.  
El menor valor de k obtenido fue K=1.

Los dendrogramas y PCAs de *A. gentryi* y *A. salmiana* no muestran la misma afinidad entre los individuos de un mismo sitio (Figura 14).



A

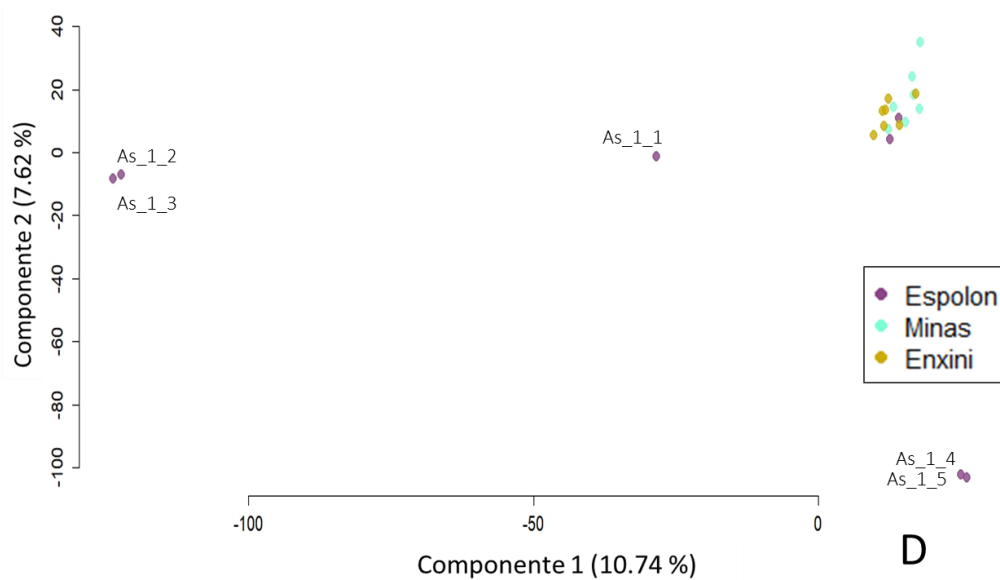
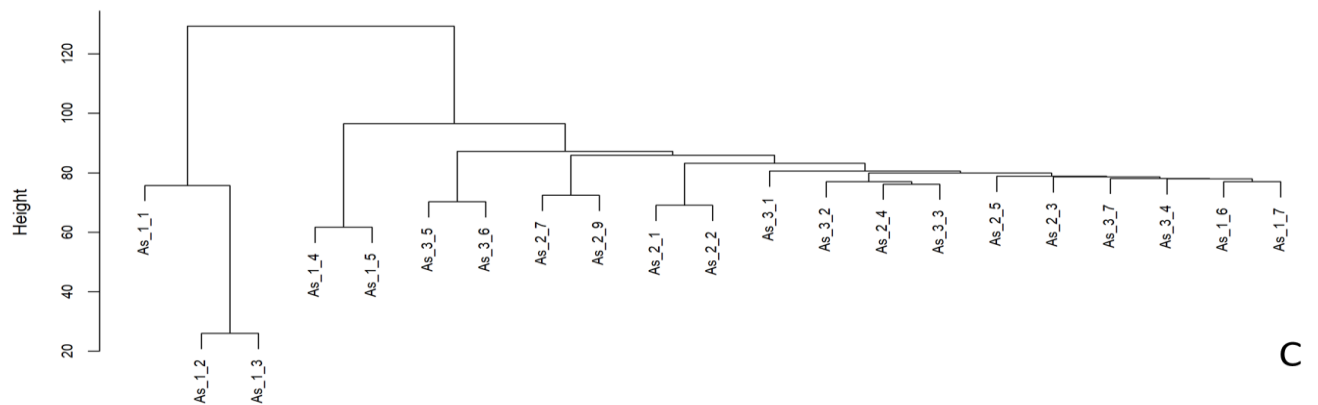
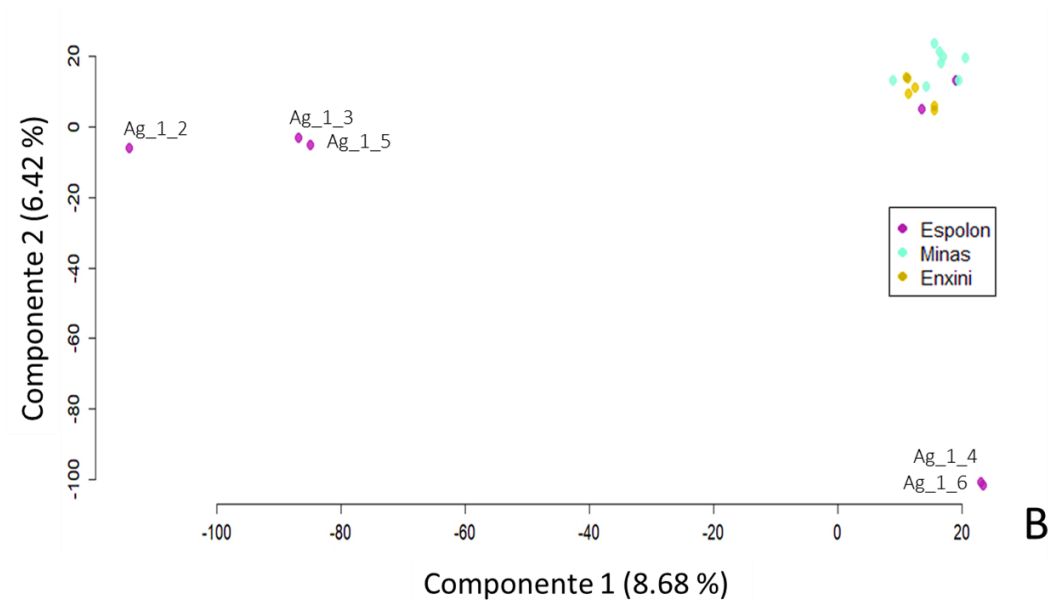


Figura 14. A. Dendrograma de los 21 individuos de *A. gentryi* B. Análisis de componentes principales de *A. gentryi* por sitio de muestreo. C. Dendrograma de los 21 individuos de *A. salmiana* D. Análisis de componentes principales de *A. salmiana* por sitio de muestreo.

Los gráficos de Structure que incluyen a *A. gentryi* y todos los intermedios (Figura 15A) y el conjunto de *A. montana* y todos los intermedios (Figura 15B) muestran, de la misma forma que el análisis con el set de individuos completo, indicios de mezcla de la variante genética de *A. montana* en algunos de los individuos etiquetados como *A. gentryi* o sus intermedios afines.

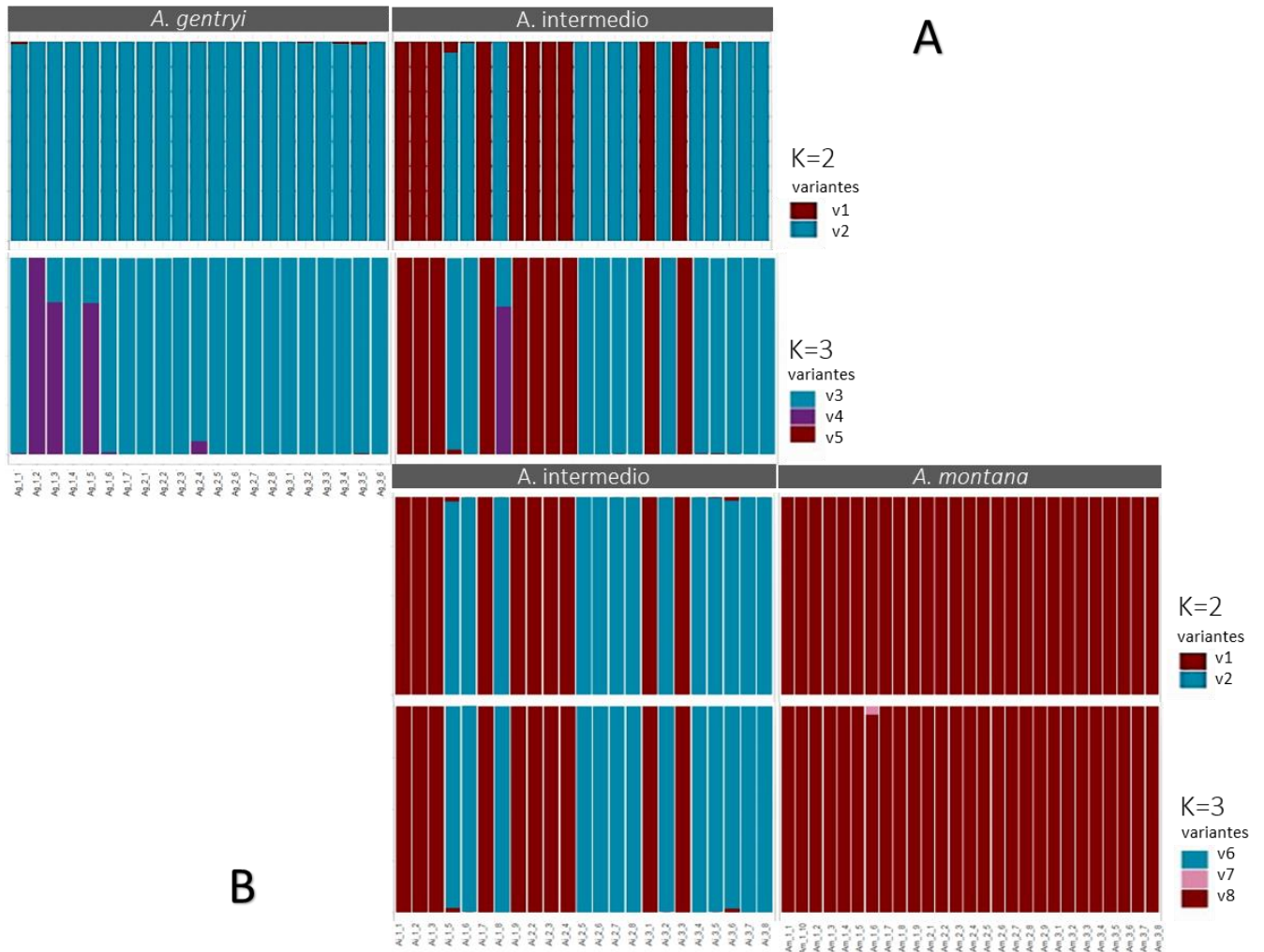


Figura 15. Análisis Structure A. individuos de *A. gentryi* y todos los intermedios. B. individuos de *A. montana* y todos los intermedios. En ambos casos el menor valor de k obtenido fue K=2.

Debido a que los análisis no revelan ninguna mezcla en los individuos intermedios afines a *A. montana* se realizaron análisis con un set formado por todos los individuos de *A. montana* sumado a los intermedios, pero solo estos intermedios afines (Figura 16) para

comprobar si existían los mismos indicios de agrupación de variantes genética por sitio de muestreo como se aprecia en la Figura 13.

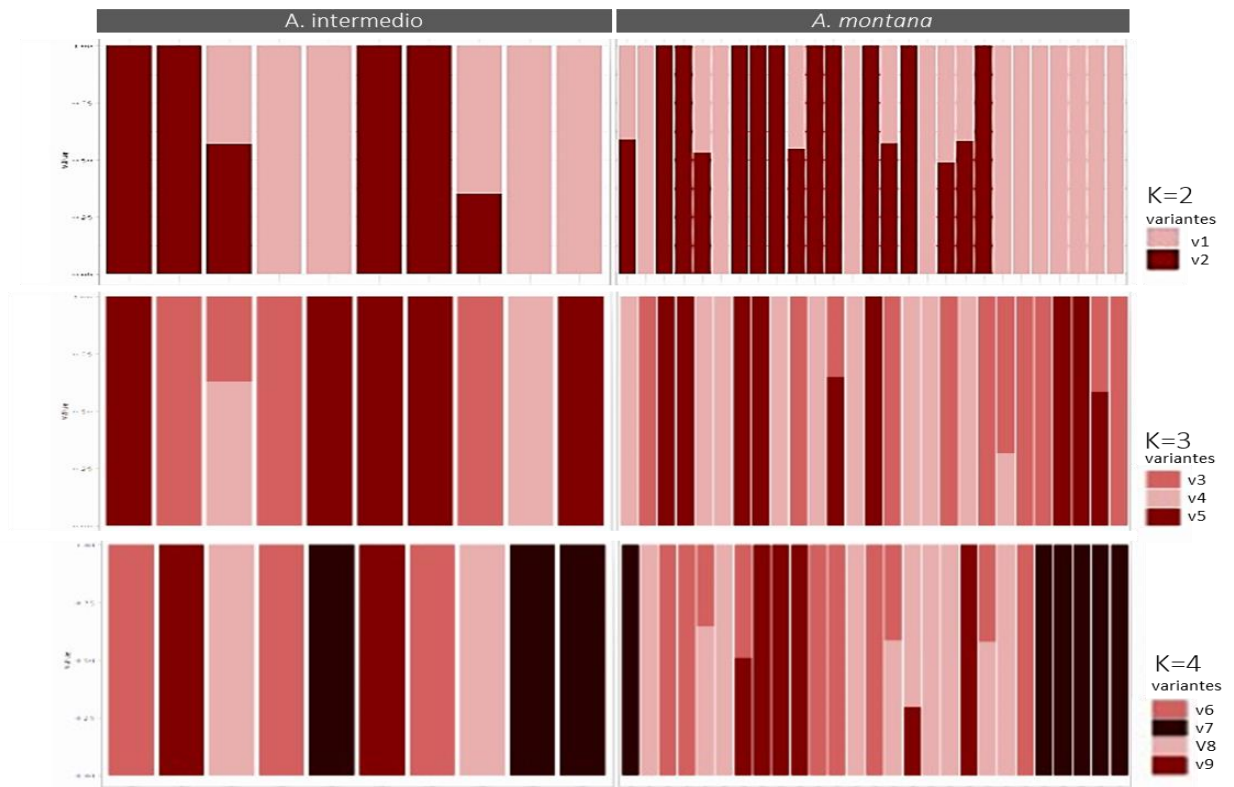


Figura 16. Análisis Structure de *A. montana* e intermedios afines. El menor valor de k obtenido fue K=1

De forma similar realizamos los análisis para el conjunto *A. gentryi*, *A. salmiana* y sus intermedios afines (Figura 17).

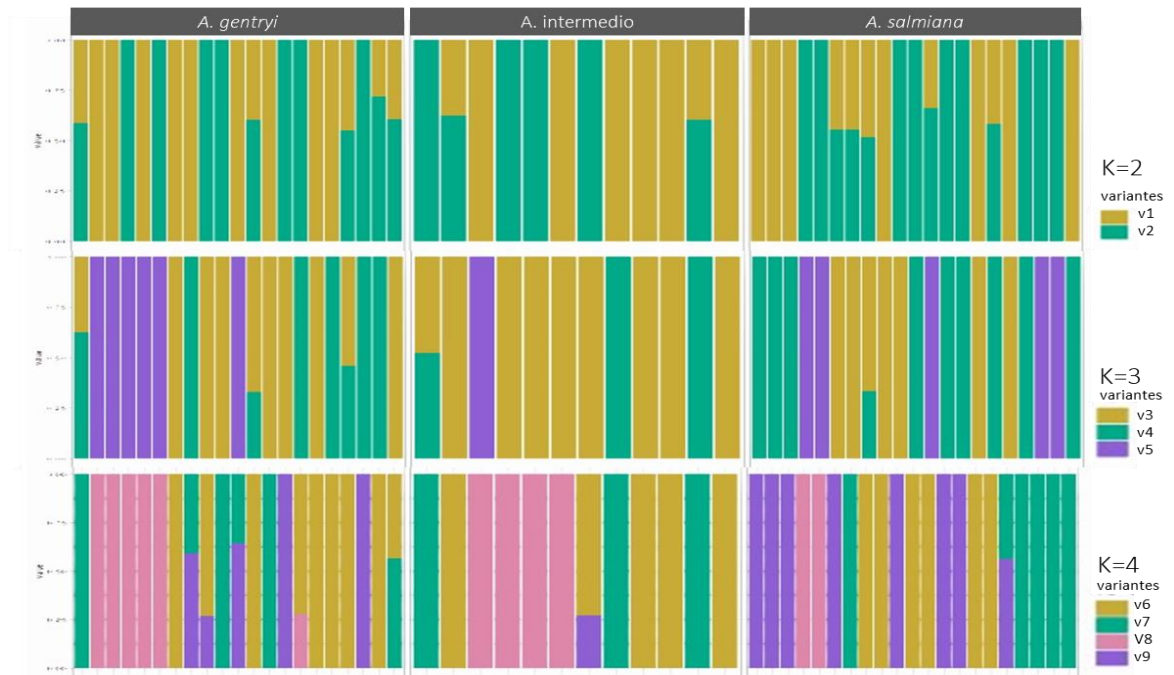


Figura 17. Análisis Structure de *A. gentryi*, *A. salmiana* e intermedios afines. El menor valor de k obtenido fue K=1.

## 5. Discusión

### 5.1 Morfología

Los diferentes análisis morfológicos fueron congruentes en sus resultados y permitieron agrupar, pero no separar completamente cada una de las especies (Figura 2,5,6,7). De acuerdo con las variables medidas, los intervalos morfológicos de *A. gentryi* se superponen con los de *A. montana* y *A. salmiana* (cuadro 2 y Figura 3). Esta superposición podría deberse a que las variables seleccionadas no son adecuadas para distinguir entre grupos. Sin embargo, muchas de estas mismas variables resultaron ser útiles en la

diferenciación morfológica de especies (Márquez-López et al., 2022), e incluso de variedades dentro del complejo *Agave angustifolia* (Cabrera-Toledo et al., 2022).

Es importante aclarar que no nos referimos a que las especies objeto de estudio no sean distinguibles morfológicamente por su apariencia general, sino a que las mediciones de estas variables específicas no permiten una demarcación clara. Esto subraya la necesidad de utilizar varios enfoques que combinen diferentes tipos de datos para poner a prueba nuestra hipótesis.

La elevada varianza de todas las variables (cuadro 2) nos muestra que dentro de cada especie existe una gran plasticidad fenotípica. Esta alta variabilidad intrínseca dentro de las especies del género ya ha sido registrada en numerosas ocasiones (Gentry, 1982; Mora-López et al., 2011; Parker et al., 2014). De manera similar, en un estudio de poblaciones silvestres y cultivadas, García et al. (2022), observaron amplios intervalos en sus mediciones morfológicas de *A. salmiana* (cuadro 6). No encontramos estudios morfológicos publicados de *A. montana* para poder hacer comparaciones.

Cuadro 6. Comparación entre los intervalos de las principales variables morfológicas medidas en el presente estudio y los registrados por García et al., 2022.

	García et al., 2022 (cm)	Datos del presente estudio (cm)
Altura de la roseta	80-200	137-213
Ancho de la roseta	120-250	223-297
Longitud de hoja	70-200	95-123
Ancho de la hoja	16-40	23-29

En el estudio de Jiménez-Barrón et al. (2020), se clasifica *A. gentryi* y *A. salmiana* dentro de un mismo clado, donde *A. salmiana* se sitúa como hermano del clado compuesto por *A. gentryi* + *A. ovatifolia*. Ambas especies, *A. gentryi* y *A. salmiana*, tienen una apariencia



general muy similar, pero se diferencian principalmente en el tamaño, siendo *A. salmiana* más grande y con un mayor número de hojas.

La diferencia de tamaño entre las dos especies podría explicarse por el proceso de domesticación de *A. salmiana*, ya que una característica común de las plantas domesticadas es el aumento de tamaño (Martínez-Ainsworth y Tenaillon, 2016). En el caso de los agaves domesticados para la producción de bebidas alcohólicas, este aumento de tamaño es particularmente notable (Figueredo-Urbina et al., 2017). La poliploidización es una causa frecuente de este aumento de tamaño (Molero-Paredes y Matos, 2008). Gentry (1982) sugiere que *A. salmiana* es el resultado de la poliploidización y domesticación de *A. gentryi*, el cual fue anteriormente conocido como *Agave macroculmis* y luego considerado sinónimo de *A. gentryi* por Ullrich en 1990. García-Mendoza (2007), plantea que *A. salmiana* es hexaploide (6x), mientras que en el estudio de Palomino et al. (2007) encontraron que podía ser tetraploide (4x) o hexaploide (6x). Esto explicaría tanto sus similitudes como sus diferencias en tamaño.

## 5.2 Genómica

El número de SNPs identificados generalmente está relacionado con la variabilidad genética dentro de cada grupo (cuadro 3). En nuestro caso, para *A. montana*, se detectó un mayor número de SNPs (12,039) en comparación con *A. gentryi* (5,104) y *A. salmiana* (8,276). Estas cifras podrían estar explicadas por el número de individuos muestreados (N), en lugar de por la variabilidad genética de cada especie. El total de individuos de *A. montana* analizados fue superior, y esto puede influir en la cantidad de SNPs detectados, ya que un mayor número de individuos puede aumentar la probabilidad de detectar variantes genéticas.

En el conjunto que incluye a los 92 individuos se detectaron menos SNPs que en la mayoría de las combinaciones con intermedios (excepto *A. gentryi* + *A. salmiana* + intermedios afines), a pesar de que fueron aplicados los mismos filtros. Cuando se analizan todos los individuos en conjunto, los filtros y umbrales específicos tienden a detectar SNPs comunes y relevantes para todo el grupo, lo que puede excluir SNPs específicos de subpoblaciones o que no alcanzan la frecuencia mínima necesaria para ser considerados como SNPs válidos en el conjunto completo. El tamaño y la composición

de la muestra pueden influir en la detección de SNPs. Los subgrupos con intermedios afines pueden tener una estructura genética que facilita la identificación de SNPs adicionales debido a la mayor frecuencia de alelos compartidos o a la reducción de la complejidad genética en comparación con el análisis de todos los individuos juntos.

En todos los análisis genómicos encontramos a *A. montana* como un grupo bien definido, separado del resto, en contraste con *A. gentryi* y *A. salmiana* que parecieran tener una identidad genética común (Figuras 8,9,10,11). En este grupo *A. gentryi*-*A. salmiana* encontramos varios individuos con mezcla de la variante genética del grupo de *A. montana* (Figura 10). En *A. montana* no detectamos con nuestros métodos presencia de variantes genéticas de *A. gentryi* o *A. salmiana*.

Villareal (1996), en su artículo de descripción de *A. montana*, propone que la especie más cercana a *A. montana* es *A. gentryi*. Pero lo cierto es que la posición filogenética de *A. montana* es desconocida, ya que no ha sido incluida en las filogenias recientes del grupo. Por lo tanto, no sabemos cuán cercana es a *A. gentryi* y *A. salmiana*, ni el tiempo transcurrido desde la divergencia de sus grupos ancestrales respectivos.

Es posible que *A. montana* haya desarrollado barreras reproductivas efectivas frente a estas otras especies de agaves mezcaleros de la Sierra el Doctor, según indica la ausencia de mezcla que hemos detectado. Sin embargo, esto no se puede confirmar sin un estudio enfocado en su biología reproductiva. Es posible que existan incompatibilidades a pesar de la sincronización de la temporada de floración. Otra posibilidad es que, a pesar de que haya compatibilidad entre el polen de otro agave mezcalero y el estigma de *A. montana*, las semillas resultantes sean inviables. Un estudio en curso (Uribe et al., en preparación) encontró que aproximadamente 60 % de las semillas de *A. montana* en la zona son inviables, siendo completamente blancas en contraste con las semillas negras que contienen un embrión. Para esclarecer estas cuestiones se recomienda desarrollar un estudio de biología reproductiva de estas especies que incluya experimentos de cruce controlado.

En cuanto a *A. gentryi* y *A. salmiana*, los análisis genómicos indican una falta de diferenciación (Figura 10) y los análisis morfológicos muestran una superposición en los intervalos de muchas variables. Como mencionábamos anteriormente existe una

posibilidad de que una sea poliploide de la otra. Si esto fuera cierto implica que es poco probable encontrar grandes diferencias en los SNPs puesto que no ha pasado tiempo suficiente desde su separación para acumular diferencias. Estaríamos en presencia del mismo genoma duplicado.

Además, es válido señalar que esta posible poliploidía de *A. salmiana* no se tuvo en cuenta en los diferentes análisis genómicos. De igual forma se desconoce la ploidía de *A. montana* y *A. gentryi*, por lo que los resultados deben ser analizados con cuidado teniendo en cuenta este posible sesgo.

Los análisis de componentes principales (PCA) de las especies por separado (Figuras 12A, 14B, 14D), que comparan los tres sitios de muestreo, mostraron patrones consistentes con los descritos por las comparaciones de  $F_{ST}$  por pares y la agrupación de STRUCTURE. De acuerdo con otros estudios sobre agaves, la diferenciación genética suele ser baja entre poblaciones (Eguiarte et al., 2013, 2021; Klimova et al., 2022). Por ejemplo, Klimova et al. (2022) reportaron un  $F_{ST}$  de 0.005 en *A. angustifolia*, mientras que Eguiarte et al. (2013) encontraron un  $F_{ST}$  de 0.057 en *Agave cocui* utilizando isoenzimas, en localidades separadas por 90-250 km. Aunque los valores de  $F_{ST}$  en nuestro estudio son similares, no son directamente comparables, ya que los estudios mencionados analizaban diferencias entre poblaciones. En contraste, nuestro estudio compara sitios que no deben considerarse poblaciones distintas, dado que la distancia máxima entre ellos es de 4.4 km, lo que en la práctica los hace parte de una misma población de cada especie.

Este patrón de baja diferenciación dentro de una especie entre sitios muy próximos se mantiene por los polinizadores más importante, como los murciélagos nectarívoros del género *Leptonycteris*, que tienen la capacidad de volar cientos de kilómetros para alimentarse (Eguiarte et al., 2021). Sin embargo, la gran cantidad de inflorescencias disponibles durante la floración, cuando los tres agaves mezcaleros de la Sierra El Doctor florecen simultáneamente, resulta en la abundancia de néctar disponible. Esta abundancia probablemente hace innecesario que los murciélagos vuelen entre sitios, incluso si están separados por solo dos a cuatro kilómetros. Esto podría explicar el caso de *A. montana*, donde hay indicios de la prevalencia de variantes genéticas diferentes en cada sitio (Figura

13, correspondiente a  $K=3$  y  $K=4$ ). Morfológicamente también hubo diferencias significativas cuando analizamos la población de *A. montana* por sitio.

Sin embargo, se esperaría que este patrón también se repitiera en *A. gentryi* y *A. salmiana*, pero no encontramos variantes genéticas asociadas al sitio de muestreo en estas especies. Como mencionamos, *A. montana* parece tener mecanismos que evitan la mezcla con las otras especies. En cambio, *A. gentryi* y *A. salmiana* aparentemente carecen de estos mecanismos y pueden establecer flujo entre ellas, además de ocasionalmente recibir flujo génico de *A. montana*. Este flujo génico posiblemente impide la estructuración poblacional en *A. gentryi* y *A. salmiana*.

*Agave salmiana* en la Sierra El Doctor es cultivada, y es muy probable que los individuos muestreados hayan sido plantados en el cerro por los habitantes de las comunidades cercanas. Estos individuos provienen de una reproducción asexual, en este caso por hijuelos, lo que en la práctica significa que son clones. La reproducción asexual impide el surgimiento en la descendencia de nuevos genotipos combinados y normalmente se asocia con una menor heterocigosidad. Sin embargo, el bajo valor de  $F_{IS}$  obtenido en *A. salmiana* indica un exceso de heterocigotos. Esta idea puede parecer contradictoria, pero es importante considerar que el grado de heterocigosidad de los descendientes depende exclusivamente de la heterocigosidad de la planta madre.

En poblaciones con reproducción sexual, pero con muy baja diversidad genética, los descendientes pueden ser homocigóticos, mientras que en una población altamente diversa y heterocigótica con reproducción asexual se puede obtener descendencia con alta heterocigosidad. Los  $F_{IS}$  negativos no son infrecuentes en las plantas y se han reportado en varias especies manejadas, como *Agave angustifolia*, *A. tequilana* y *A. rhodacantha*, con valores de  $F_{IS}$  que van desde -0.8420 a 0.1326 (Cabrera-Toledo et al., 2022). En otros estudios, como el de *A. tequilana* (2022), se encontró que a pesar de la reproducción clonal, el  $F_{IS}$  era negativo y había un exceso de heterocigotos.

Además, no debemos olvidar que, aunque estos *A. salmiana* fueron posiblemente plantados, nada asegura que las plantas madres hayan provenido de la misma población. *A. salmiana* tiene una gran diversidad genotípica a lo largo del país y se reconocen al

menos cuatro variedades (Figueredo-Urbina et al., 2021). Por lo tanto, es posible que la alta diversidad genotípica de las plantas madres contribuya al bajo  $F_{IS}$  observado.

El coeficiente de endogamia positivo observado para *A. montana* (cuadro 4) podría indicar la ausencia de flujo génico, sugiriendo cierto grado de aislamiento reproductivo, tanto respecto a otros agaves mezcaleros de la Sierra El Doctor, como de otras poblaciones de *A. montana*. La población de *A. montana* registrada más cercana se encuentra a más de 600 km en la Sierra de Tamaulipas. Este aislamiento puede llevar a la endogamia, aumentando la homocigosidad dentro de la población. Es importante destacar que es poco probable que exista clonalidad entre los individuos de *A. montana* muestreados, ya que no se observaron hijuelos ni bulbillos, a diferencia de lo que se encontró en *A. salmiana* y *A. gentryi*. La ausencia de estos métodos de reproducción asexual en *A. montana* refuerza la idea de que el aislamiento reproductivo y la endogamia podrían ser las causas principales del coeficiente de endogamia positivo, en lugar de la clonación.

Otro factor que podría estar influyendo en el déficit de heterocigotos y  $F_{IS}$  positivo de *A. montana* es el corte de las inflorescencias antes de su total maduración. Los pobladores locales usan las flores tiernas como alimento, por lo que buena parte de los individuos maduros no alcanzan a contribuir genéticamente en la próxima generación. Esta circunstancia impacta directamente sobre el tamaño efectivo de la población, disminuyéndolo. Al limitar la reproducción sexual de la especie, es probable que, debido a la deriva genética, algunos alelos ya se hayan perdido o se perderán en el futuro si esta práctica continúa.

### **5.3 Integración morfo-genómica**

La variabilidad morfológica y la plasticidad fenotípica de estas especies podrían hacer que las características morfológicas sean inadecuadas para detectar mezclas o cuantificar el grado de hibridación en este caso. El reconocimiento de los individuos que evidenciaron cierto grado de introgresión es casi imposible desde el punto de vista morfológico. En algunos casos etiquetamos como presuntos híbridos a individuos en los

que no detectamos luego señales de mezcla con los SNPs. De forma contraria, etiquetamos como “puros” a individuos que luego en los análisis genómicos mostraron cierto nivel de mezcla. También es posible que haya presente otros individuos con mezcla en nuestra muestra pero que no hayan podido ser detectados con nuestros métodos. A pesar de estas limitaciones, los datos morfológicos representan información complementaria importante. Sin embargo, y retomando nuestra hipótesis, la morfología por sí sola no explica toda la variabilidad observada, sugiriendo que otros factores, como la hibridación, pueden estar en juego.

Además, reiteramos que una pregunta de investigación como la planteada en este estudio necesariamente debe ser abordada desde múltiples enfoques. Además del morfológico y genómico, aquí utilizados, podrían ser útiles el enfoque de biología reproductiva y el citogenético.

Finalmente, todas las interpretaciones aquí plasmadas se basan en datos locales, están limitadas a un pequeño espacio dentro de la amplia distribución de las tres especies. La panorámica ampliada, donde se incluyan otras zonas donde las tres especies coinciden como Tamaulipas, o lugares donde no sean simpátricas podrían cambiar nuestra interpretación y revelar patrones. Incluir áreas adicionales en el estudio también permitiría validar o refutar hipótesis planteadas a partir de los datos locales como la ausencia de mezcla en *A. montana*.

## 6. Conclusiones

1- Se logró caracterizar morfológicamente las poblaciones estudiadas. Sin embargo, debido a la variabilidad morfológica y la plasticidad fenotípica de las especies, se encontró que las características morfológicas no son adecuadas para detectar mezclas o cuantificar el grado de hibridación en este caso específico.

2- Se logró caracterizar genéticamente las poblaciones. *A. montana* presenta un déficit de heterocigotos y un alto coeficiente de endogamia, mientras que *A. gentryi* y *A. salmiana* presentan

un exceso de heterocigotos y bajo coeficiente de endogamia. Las poblaciones se diferencian poco entre sí. Aparentemente, existe un flujo genético de *A. montana* hacia *A. gentryi* y *A. salmiana*, pero no a la inversa. También parece existir un flujo bidireccional entre *A. gentryi* y *A. salmiana*.

3-Toda esta evidencia nos lleva a rechazar la hipótesis alternativa de que toda la variabilidad observada es solo producto de plasticidad fenotípica. Todo apunta a procesos de hibridación en curso.

## 7. Referencias

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., & Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26, 229–246.

Alanís-Flores, G. J., & González-Álvarez, M. (2011). Formas de uso de los magueyes (*Agave* spp.) en Nuevo León, México. *Revista Salud Pública y Nutrición*, 5, 287-299.

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology*, 26(3), 420-430. <https://doi.org/10.1111/mec.13948>

Angiosperm Phylogeny Group [APG IV]. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, 181(1), 1–20. <https://doi.org/10.1111/boj.12385>

Baiakhmetov, E., Ryzhakova, D., Gudkova, P. D., et al. (2021). Evidence for extensive hybridisation and past introgression events in feather grasses using genome-wide SNP genotyping. *BMC Plant Biology*, 21, 505. <https://doi.org/10.1186/s12870-021-03287-w>

Bayona, A., & Chávez, R. G. (2013). Cuantificación de las áreas afectadas durante la construcción de un acueducto en una zona de alta diversidad biológica en Querétaro, México; hábitat prioritario para la conservación de cactáceas y suculentas amenazadas y endémicas. *Cat Suc Mex*, 58(2), 49–63.

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

Boto, L. (2012). Evolución reticulada. *Evolución*, 7(2), 73-83.

Cabrera-Toledo, D., Mendoza-Galindo, E., Larranaga, N., Herrera-Estrella, A., Vásquez-Cruz, M., & Hernández-Hernández, T. (2022). Genomic and Morphological Differentiation of Spirit Producing *Agave angustifolia* Traditional Landraces Cultivated in Jalisco, México. *Plants*, 11, 2274. <https://doi.org/10.3390/plants11172274>

Castorena-Sánchez, I., Escobedo, R. M., & Quiroz, A. (1991). New cytotaxonomical determinants recognized in six taxa of *Agave* in the sections *Rigidae* and *Sisalanae*. *Can. J. Bot.*, 69, 1257–1264. <https://doi.org/10.1139/b91-163>

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

Colunga-García Marín, P., Zizumbo-Villarreal, D., Torres, I., Casas, A., Figueredo-Urbina, C. J., Rangel-Landa, S., et al. (2017). Los agaves y las prácticas mesoamericanas de aprovechamiento, manejo y domesticación. In A. Casas, J. Torres-Guevara, & F. Parra (Eds.), *Domesticación en el Continente Americano. Investigación para el Manejo Sustentable de Recursos Genéticos en el Nuevo Mundo* (pp. 273–308). México: UNAM-UNALM.

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

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

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, 36, 2592–2594.



Edelman, N. B., & Mallet, J. (2021). Prevalence and adaptive impact of introgression. *Annual Review of Genetics*, 55, 265–283. <https://doi.org/10.1146/annurev-genet-021821-020805>

Edwards, A. W. F. (1971). Distance between populations on the basis of gene frequencies. *Biometrics*, 27, 873–881. <https://doi.org/10.2307/2528824>

Eguiarte, L. E., Aguirre-Planter, E., Aguirre, X., Colín, R., González, A., Rocha, M., & Souza, V. (2013). From isozymes to genomics: population genetics and conservation of *Agave* in México. *The Botanical Review*, 79(4), 483–506. DOI: 10.1007/s12229-013-9123-x.

Eguiarte, L. E., Jiménez Barrón, O. A., Aguirre-Planter, E., Scheinvar, E., Gámez, N., Gasca-Pineda, J., Castellanos-Morales, G., Moreno-Letelier, A., & Souza, V. (2021). Evolutionary ecology of *Agave*: Distribution patterns, phylogeny, and coevolution (an homage to Howard S. Gentry). *American Journal of Botany*, 108, 216–235.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., & et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6(5), e19379. <https://doi.org/10.1371/journal.pone.0019379>

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.

Figueredo-Urbina, C. J., Casas, A., & Torres-García, I. (2017). Morphological and genetic divergence between *Agave inaequidens*, *A. cupreata* and the domesticated *A. hookeri*. Analysis of their evolutionary relationships. *PLoS ONE*, 12(11), e0187260. <https://doi.org/10.1371/journal.pone.0187260>

Figueredo-Urbina, C. J., Álvarez-Ríos, G. D., García-Montes, M. A., & Octavio-Aguilar, P. (2021). Morphological and genetic diversity of traditional varieties of agave in Hidalgo State, Mexico. *\*PLOS ONE*, 16\*(7), e0254376. <https://doi.org/10.1371/journal.pone.0254376>

Fitzpatrick, B. M., Ryan, M. E., Johnson, J. R., Corush, J., & Carter, E. T. (2015). Hybridization and the species problem in conservation. *Current Zoology*, 61(1), 206–216.  
<https://doi.org/10.1093/czoolo/61.1.206>

Flores-Morales, A., Chávez-Avila, V. M., & Jiménez-Estrada, M. (2021). Evaluación de una alternativa de propagación de maguey pulquero (*Agave salmiana*) variedad púa larga. *Revista Mexicana de Agroecosistemas*, 8(1), 46-58.

Gambini, J., Gimeno-Mallench, L., Inglés, M., Olasso, G., Abdelaziz, K. M., Avellana, J. A., Belenguer, Á., Cruz, R., Mas-Bargues, C., Borrás, C., & Viña, J. (2016). Identificación de polimorfismos de nucleótido simple en centenarios. *Revista Española de Geriatria y Gerontología*, 51(3), 146–149. <https://doi.org/10.1016/j.regg.2015.09.006>

García, E. (1964). Modificaciones al Sistema de Clasificación Climática de Köppen (para adaptarlo a las condiciones de la República Mexicana). México: Instituto de Geografía, UNAM.

García, F. P., Méndez, J. P., Muñoz, E. J., Sandoval, O. A. A., & Laguna, R. R. (2022). Taxonomic, physical and morphological characterization of four species of agave with potential for the production of cellulose fibers from the leaves. *South Florida Journal of Development*, 3(1), 1277–1301. <https://doi.org/10.46932/sfjdv3n1-099>

García-Mendoza, A. (2007). Los Agaves de México. *Ciencias*, 87, 14-23.

García-Mendoza, A. J., Franco Martínez, I. S., & Sandoval Gutiérrez, D. (2019). Cuatro especies nuevas de *Agave* (Asparagaceae, Agavoideae) del sur de México. *Acta Botánica Mexicana*, 126, e1461. <https://doi.org/10.21829/abm126.2019.1461>

Gentry, H. S. (1982). *Agaves of continental North America*. Tucson, Arizona, USA: The University of Arizona Press.

González-Elizondo, M. S., Elizondo, M., Enríquez, I., Reséndiz-Rojas, L., Flores, J. A., & Retana, F. (2011). El complejo *Agave victoriae-reginae* (Agavaceae). *Acta Botánica Mexicana*, 95, 268.

Goudet, J., & Jombart, T. (2015). Package “hierfstat”, Version 1.5-11. Estimation and Tests of Hierarchical F-Statistics. Available online:  
<https://cran.rproject.org/web/packages/hierfstat/index.html>

Graham, C. F., Eberts, R. L., Goncin, U., & Somers, C. M. (2021). Spontaneous hybridization and introgression between walleye (*Sander vitreus*) and sauger (*Sander canadensis*) in two large reservoirs: Insights from genotyping by sequencing. *Evolutionary Applications*, 14(4), 965–982.

Gross, S. M., Martin, J. A., Simpson, J., Abraham-Juarez, M. J., Wang, Z., & Visel, A. (2013). *De novo* transcriptome assembly of drought-tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genomics*, 14(1), 563. <https://doi.org/10.1186/1471-2164-14-563>

Hibbins, M. S., & Hahn, M. W. (2021). Phylogenomic approaches to detecting and characterizing introgression. *EcoEvoRxiv*. <https://doi.org/10.32942/osf.io/uahd8>

INEGI. (2013). Conjunto de Datos Vectoriales de Uso de Suelo y Vegetación. Escala 1:250,000. Serie 5, Capa unión. Instituto Nacional de Estadística, Geografía e Informática.

Ilut, D. C., Sanchez, P. L., Costich, D. E., Friebe, B., Coffelt, T. A., Dyer, J. M., Jenks, M. A., & Gore, M. A. (2015). Genomic diversity and phylogenetic relationships in the genus *Parthenium* (Asteraceae). *Industrial Crops and Products*, 76, 920-929.

IUCN. (2019). The IUCN Red List of Threatened Species. Version 2019-3. Available at: [www.iucnredlist.org](http://www.iucnredlist.org).

Jacques-Hernández, C., Herrera-Pérez, O., & Ramírez-De León, J. A. (2007). El Maguey Mezcalero y la agroindustria del mezcal en Tamaulipas. In P. Colunga-GarcíaMarín, L. Eguiarte, A. Larqué, & D. Zizumbo-Villarreal (Eds.), *En lo ancestral hay futuro: Del tequila, los mezcales y otros agaves* (pp. 287–317). Mexico: Centro de Investigaciones Científicas de Yucatán CONACYT-Comisión Nacional de Biodiversidad-Instituto Nacional de Ecología.

Jiménez-Barron, O., García-Sandoval, R., Magallón, S., García-Mendoza, A., Nieto-Sotelo, J., Aguirre-Planter, E., & Eguiarte, L. E. (2020). Phylogeny, Diversification Rate, and Divergence Time of *Agave* sensu lato (Asparagaceae), a Group of Recent Origin in the Process of Diversification. *Frontiers in Plant Science*, 11, 536135.  
<https://doi.org/10.3389/fpls.2020.536135>

Jombart, T., & Ahmed, I. (2011). New Tools for the Analysis of Genome-Wide SNP Data. *Bioinformatics*, 27, 3070–3071.

Klimova, A., Ruiz Mondragón, K. Y., Molina Frenan, F., Aguirre-Planter, E., & Eguiarte, L. E. (2022). Genomic Analyses of Wild and Cultivated Bacanora Agave (*Agave angustifolia* var. *pacifica*) Reveal Inbreeding, Few Signs of Cultivation History and Shallow Population Structure. *Plants*, 11, 1426. <https://doi.org/10.3390/plants11111426>

Kong, S., & Kubatko, L. S. (2021). Comparative Performance of Popular Methods for Hybrid Detection using Genomic Data. *Systematic Biology*, 70(5), 891–907. <https://doi.org/10.1093/sysbio/syaa092>

López-Causapé, C., González Candelas, F., Tomás Carmona, M., & Oliver Palomo, A. (2021). Aplicaciones de las técnicas de secuenciación masiva en la Microbiología Clínica. En E. Cercenado Mansilla & R. Cantón Moreno (Eds.), *Procedimientos en Microbiología Clínica* (pp. 146–149). Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica (SEIMC). ISBN: 978-84-09-31350-1

López de Heredia, U. (2016). Las técnicas de secuenciación masiva en el estudio de la diversidad biológica. *Munibe, Cienc. Nat.*, 64, 7-31. Donostia/San Sebastián.

López de Heredia, U., & Vázquez-Poletti, J. L. (2016). RNA-seq analysis in forest tree species: bioinformatics problems and solutions. *Tree Genetics & Genomes*, 12, 30. <https://doi.org/10.1007/s11295-016-0995-x>

Magallán, F. (1998). *Las Agaváceas de Querétaro* (Tesis de licenciatura). Escuela de Biología, Universidad Autónoma de Querétaro, Querétaro, Qro., México.

Malinsky, M., Trucchi, E., Lawson, D., & Falush, D. (2018). RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution*, 35,\* 10.1093/molbev/msy023.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20(5), 229-237.

- Márquez-López, R. E., Santiago-García, P. A., & López, M. G. (2022). *Agave fructans* in Oaxaca's emblematic specimens: *Agave angustifolia* Haw. and *Agave potatorum* Zucc. *Plants*, 11(14), 1834. <https://doi.org/10.3390/plants11141834>
- Martínez-Ainsworth, N. E., & Tenaillon, M. I. (2016). Superheroes and masterminds of plant domestication. *Comptes rendus biologies*, 339(7-8), 268–273. <https://doi.org/10.1016/j.crv.2016.05.005>
- Masuelli, R. W., & Marfil, C. F. (2011). Variabilidad epigenética en plantas y evolución. *Journal of Basic & Applied Genetics*, 22.
- McFarlane, S. E., & Pemberton, J. M. (2019). Detecting the true extent of introgression during anthropogenic hybridization. *Trends in Ecology and Evolution*, 34(4), 315–326. <https://doi.org/10.1016/j.tree.2018.12.013>
- Mora-López, J., Reyes-Agüero, J., Flores-Flores, J., Peña-Valdivia, C., & Aguirre-Rivera, J. (2011). Variación morfológica y humanización de la sección *Salmaniae* del género *Agave*. *Agrociencia*, 45, 465-477.
- Palomino, G., Martinez, J., & Mendez, I. (2007). Variación inter e intraespecífica en especies de *Agave* por citometría de flujo y análisis de sus cromosomas. En P. Colunga-García Marín, A. Larqué, L. Eguiarte, & D. Zizumbo-Villareal (Eds.), *En lo ancestral hay futuro: del tequila, los mezcales y otros agaves* (pp. [números de páginas, si están disponibles]). CICY, CONACYT, CONABIO, INE y SEMARNAT. Mérida, Yucatán, México.
- Parker, K. C., Trapnell, D. W., Hamrick, J. L., & Jodgson, W. C. (2014). Genetic and morphological contrast between wild and anthropogenic populations of *Agave parryi* var. *huachucensis* in south-eastern Arizona. *Annals of Botany*, 113, 939-952.
- Pérez-Zavala, M. d. L., Hernández-Arzaba, J. C., Bideshi, D. K., & Barboza-Corona, J. E. (2020). *Agave*: a natural renewable resource with multiple applications. *Journal of the Science of Food and Agriculture*, 100, 5324–5333. <https://doi.org/10.1002/jsfa.10586>
- Pritchard, J. K., Stephens, P., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.

Pinkava, D. J., & Baker, M. A. (1985). Chromosome and Hybridization Studies of *Agave*. *Desert Plants*, 7(2), 93–100.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>.

Qiao, H., Liu, W., Zhang, Y., Zhang, Y. Y., & Li, Q. Q. (2019). Genetic admixture accelerates invasion via provisioning rapid adaptive evolution. *Molecular Ecology*, 28, 4012–4027.

R Core Team. (2023). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: <http://www.R-project.org>

Ratcliffe, B., El-Dien, O. G., Klápště, J., Porth, I., Chen, C., Jaquish, B., & El-Kassaby, Y. A. (2015). A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* × *glauca*) using unordered SNP imputation methods. *Heredity*, 115, 547–555.

Rius, M., & Darling, J. A. (2014). How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, 29, 233–242.

Rocha, M., Good-Ávila, S. V., Molina-Freaner, F., Arita, H. T., Castillo, A., García-Mendoza, A., Silva-Montellano, A., Gaut, B. S., Souza, V., & Eguiarte, L. E. (2006). Pollination biology and adaptive radiation of Agavaceae, with special emphasis on the genus *Agave*. *Aliso*, 22, 329–344.

Roll, C. M., Hodgson, W. C., & Salywon, A. M. (2021). Notes on a hybrid colony of *Agave parryi* var. *huachucensis* x *Agave palmeri* (Agavaceae) in southeastern Arizona. *Phytoneuron*, 25, 1–7.

Rousset, F. (2008). GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, 8, 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931>.

Ruiz Mondragon, K. Y., Aguirre-Planter, E., Gasca-Pineda, J., Klimova, A., Trejo-Salazar, R.-E., Reyes Guerra, M. A., Medellin, R. A., Piñero, D., Lira, R., & Eguiarte, L. E. (2022).

Conservation genomics of *Agave tequilana* Weber var. *azul*: low genetic differentiation and heterozygote excess in the tequila agave from Jalisco, Mexico. *PeerJ*, 10, e14398. DOI: 10.7717/peerj.14398.

Schloss, J. A. (2008). How to get genomes at one ten-thousandth the cost. *Nature Biotechnology*, 26, 1113-1115.

Senanan, W., Kapuscinski, A. R., Na-Nakorn, U., & Miller, L. M. (2004). Genetic impacts of hybrid catfish farming (*Clarias macrocephalus* x *C. gariepinus*) on native catfish populations in central Thailand. *Aquaculture*, 235(1-4), 167–184.

Steinmann, V. W., Arredondo-Amezcu, L., Hernández-Cárdenas, R. A., & Ramírez-Amezcu, Y. (2021). Diversity and Origin of the Central Mexican Alpine Flora. *Diversity*, 13, 31. <https://doi.org/10.3390/d13010031>.

Thareja, G., Mathew, S., Mathew, L. S., Mohamoud, Y. A., Suhre, K., & Malek, J. A. (2018). Genotyping-by-sequencing identifies date palm clone preference in agronomics of the State of Qatar. *PLoS ONE*, 13(12), e0207299. <https://doi.org/10.1371/journal.pone.0207299>

Torres-García, I., Casas, A., Vega, E., Martinez-Ramos, M., & Delgado-Lemus, A. (2015). Population Dynamics and Sustainable Management of Mescal Agaves in Central Mexico: *Agave potatorum* in the Tehuacán-Cuicatlán Valley. *Economic Botany*, 69, 26–41. DOI: 10.1007/s12231-014-9295-2.

Torres-Suárez, O. L. (2016). Evolución reticulada por hibridación en corales. *Bulletin of Marine and Coastal Research*, 43(2). <https://doi.org/10.25268/bimc.invemar.2014.43.2.2>

Treviño-Carreón, J., Mora-Olivo, A., Carreón-Pérez, A., & Valiente-Banuet, A. (2011). Descubriendo el valor de los magueyes tamaulipecos. *Ciencia UAT*, 5(3), 34–40.

Twyford, A., & Ennos, R. A. (2012). Next-generation sequencing as a tool for plant ecology and evolution. *Plant Ecology and Diversity*, 5(4), 411-413. <https://doi.org/10.1080/17550874.2012.754513>

Ullrich, B. 1990. *Agave macroculmis* Todaro en *Agave gentryi* Ullrich spec. nov. Succulenta 69(10): 210-214.

Valiente-Banuet, A. (2023). Mezcal Boom and Extinction Debts. In: Jones, R.W., Ornelas-García, C.P., Pineda-López, R., Álvarez, F. (eds), Mexican Fauna in the Anthropocene. Springer, Cham. [https://doi.org/10.1007/978-3-031-17277-9\\_14](https://doi.org/10.1007/978-3-031-17277-9_14).

Vega-Ramos, K. L., Gutiérrez-Coronado, M. A., & Gómez-Leyva, J. F. (2022). Marcadores moleculares ISSR para identificación de hijuelos de agaves. Revista Fitotecnia Mexicana, 45(1), 75–81. <https://doi.org/10.35196/rfm.2022.1.75>.

Villarreal Q., J. A. (1996). Una nueva especie de *Agave* subgénero *Agave* (Agavaceae) de México. Sida, Contributions to Botany, 17(1), 191–195.

Villarreal-Quintanilla, J. A., Ramírez, A., Estrada, E., González, D. U., & Jasso, D. (2014). *Agave* × *madrensis* (Asparagaceae), a putative hybrid from the Sierra Madre Oriental, Mexico. Botanical Research Institute of Texas, 8(2), 441–447.

Zelenin, A. V., Rodionov, A. V., Bolsheva, N. L., Badaeva, E. D., & Muravenko, O. V. (2016). Genome: Origins and evolution of the term. Molecular Biology, 50, 542–550. <https://doi.org/10.1134/S0026893316040178>.



## 8. Anexos

Anexo 1: Información geográfica sobre los individuos estudiados morfológicamente. En violeta aparecen destacados los individuos que fueron incluidos en el estudio genómico.

especie	sitio	latitud	longitud	altitud	especie	sitio	latitud	longitud	altitud
<i>A. montana</i>	Espolón	20 47.550	99 33.486	2968	<i>A. montana</i>	Enxini	20.48.174	99.35.973	2904
<i>A. montana</i>	Espolón	20 47.635	99 33.748	3005	<i>A. montana</i>	Enxini	20.48.157	99.35.964	2910
<i>A. montana</i>	Espolón	20 47.613	99 33.646	3100	<i>A. intermedio</i>	Enxini	20.48.146	99.35.959	2912
<i>A. intermedio</i>	Espolón	20 47.605	99 33.647	3131	<i>A. montana</i>	Enxini	20.48.138	99.35.970	2922
<i>A. intermedio</i>	Espolón	20 47.585	99 33.613	3133	<i>A. montana</i>	Enxini	20.48.140	99.35.986	2932
<i>A. intermedio</i>	Espolón	20 47.589	99 33.586	3112	<i>A. salmiana</i>	Enxini	20.48.135	99.35.999	2941
<i>A. salmiana</i>	Espolón	20 47.555	99 33.493	3086	<i>A. montana</i>	Enxini	20.48.131	99.35.999	2943
<i>A. intermedio</i>	Espolón	20 47.548	99 33.485	3087	<i>A. intermedio</i>	Enxini	20.48.140	99.36.006	2944
<i>A. intermedio</i>	Espolón	20 47.548	99 33.485	3083	<i>A. montana</i>	Enxini	20.48.146	99.36.015	2947
<i>A. salmiana</i>	Espolón	20 47.548	99 33.485	3089	<i>A. intermedio</i>	Enxini	20.48.153	99.36.020	2945
<i>A. intermedio</i>	Espolón	20.47.559	99.33.457	2887	<i>A. intermedio</i>	Enxini	20.48.161	99.36.023	2943
<i>A. salmiana</i>	Espolón	20.47.558	99.33.370	3080	<i>A. montana</i>	Enxini	20.48.168	99.36.021	2940
<i>A. intermedio</i>	Espolón	20.47.558	99.33.371	3076	<i>A. salmiana</i>	Enxini	20.48.171	99.36.029	2941
<i>A. montana</i>	Espolón	20.47.551	99.33.300	3075	<i>A. intermedio</i>	Enxini	20.48.173	99.36.042	2944
<i>A. gentryi</i>	Espolón	20.47.557	99.33.295	3076	<i>A. gentryi</i>	Las Minas	20.42.097	100.26.561	2749
<i>A. montana</i>	Espolón	20.47.560	99.33.294	3079	<i>A. intermedio</i>	Las Minas	20.49.303	99.35.870	2790
<i>A. intermedio</i>	Espolón	20.47.545	99.33.270	3077	<i>A. intermedio</i>	Las Minas	20.49.302	99.35.875	2826
<i>A. intermedio</i>	Espolón	20.47.519	99.33.239	3075	<i>A. gentryi</i>	Las Minas	20.49.299	99.35.882	2857
<i>A. gentryi</i>	Espolón	20.47.508	99.33.201	3081	<i>A. intermedio</i>	Las Minas	20.49.303	99.35.907	2883
<i>A. gentryi</i>	Espolón	20.47.508	99.33.194	3084	<i>A. gentryi</i>	Las Minas	20.49.307	99.35.922	2894
<i>A. gentryi</i>	Espolón	20.47.489	99.33.135	3031	<i>A. gentryi</i>	Las Minas	20.49.319	99.35.939	2906
<i>A. gentryi</i>	Espolón	20.47.498	99.33.175	3082	<i>A. intermedio</i>	Las Minas	20.49.328	99.35.940	2913
<i>A. gentryi</i>	Espolón	20.47.475	99.33.164	3079	<i>A. gentryi</i>	Las Minas	20.49.326	99.35.941	2914
<i>A. intermedio</i>	Espolón	20.47.474	99.33.158	3079	<i>A. salmiana</i>	Las Minas	20.49.319	99.35.941	2914
<i>A. salmiana</i>	Espolón	20.47.489	99.33.181	3076	<i>A. gentryi</i>	Las Minas	20.49.295	99.35.938	2911
<i>A. salmiana</i>	Espolón	20.47.483	99.33.190	3071	<i>A. gentryi</i>	Las Minas	20.49.299	99.35.921	2917
<i>A. montana</i>	Espolón	20.47.487	99.33.186	3077	<i>A. gentryi</i>	Las Minas	20.49.292	99.35.903	2917
<i>A. salmiana</i>	Espolón	20.47.532	99.33.298	3075	<i>A. gentryi</i>	Las Minas	20.49.278	99.35.898	2916
<i>A. montana</i>	Espolón	20.47.513	99.33.289	3075	<i>A. gentryi</i>	Las Minas	20.49.293	99.35.892	2912
<i>A. montana</i>	Espolón	20.47.506	99.33.291	3074	<i>A. salmiana</i>	Las Minas	20.49.291	99.35.896	2913
<i>A. montana</i>	Espolón	20.47.478	99.33.316	3083	<i>A. salmiana</i>	Las Minas	20.49.311	99.35.982	2828
<i>A. intermedio</i>	Las Minas	20.49.304	99.35.830	2722	<i>A. salmiana</i>	Las Minas	20.49.306	99.35.992	2846
<i>A. montana</i>	Las Minas	20.49.328	99.35.840	2725	<i>A. salmiana</i>	Las Minas	20.49.303	99.35.994	2861

<i>A. montana</i>	Las Minas	20.49.322	99.35.836	2889	<i>A. gentryi</i>	Espolón	20.49.178	99.35.515	3094
<i>A. intermedio</i>	Las Minas	20.49.347	99.35.861	2892	<i>A. intermedio</i>	Enxini	20.47.748	99.34.047	2846
<i>A. intermedio</i>	Las Minas	20.49.350	99.35.853	2885	<i>A. gentryi</i>	Enxini	20.48.759	99.36.014	2856
<i>A. montana</i>	Las Minas	20.49.349	99.35.854	2880	<i>A. gentryi</i>	Enxini	20.48.243	99.36.025	2878
<i>A. montana</i>	Las Minas	20.49.370	99.35.860	2873	<i>A. salmiana</i>	Enxini	20.48.238	99.36.027	2884
<i>A. montana</i>	Las Minas	20.49.373	99.35.856	2871	<i>A. intermedio</i>	Enxini	20.48.257	99.36.039	2881
<i>A. intermedio</i>	Las Minas	20.49.374	99.35.848	2866	<i>A. gentryi</i>	Enxini	20.48.263	99.36.048	2889
<i>A. intermedio</i>	Las Minas	20.49.374	99.35.847	2876	<i>A. gentryi</i>	Enxini	20.48.278	99.36.050	2886
<i>A. montana</i>	Las Minas	20.49.383	99.35.871	2880	<i>A. gentryi</i>	Enxini	20.48.281	99.36.061	2891
<i>A. montana</i>	Las Minas	20.49.383	99.35.871	2876	<i>A. intermedio</i>	Enxini	20.48.288	99.36.067	2894
<i>A. intermedio</i>	Las Minas	20.49.390	99.35.879	2864	<i>A. gentryi</i>	Enxini	20.48.293	99.36.071	2893
<i>A. montana</i>	Las Minas	20.49.391	99.35.879	2882	<i>A. gentryi</i>	Enxini	20.48.295	99.36.073	2895
<i>A. montana</i>	Las Minas	20.49.390	99.35.918	2899	<i>A. salmiana</i>	Enxini	20.48.304	99.36.082	2897
<i>A. montana</i>	Las Minas	20.49.390	99.35.917	2905	<i>A. intermedio</i>	Enxini	20.48.296	99.36.080	2900
<i>A. salmiana</i>	Las Minas	20.49.333	99.35.996	2907	<i>A. salmiana</i>	Enxini	20.48.313	99.36.092	2899
<i>A. salmiana</i>	Las Minas	20.49.335	99.36.003	2908	<i>A. gentryi</i>	Enxini	20.48.307	99.36.092	2901
<i>A. salmiana</i>	Las Minas	20.49.340	99.36.008	2911	<i>A. gentryi</i>	Enxini	20.48.314	99.36.093	2898
<i>A. salmiana</i>	Las Minas	20.49.341	99.36.015	2913	<i>A. salmiana</i>	Enxini	20.48.316	99.36.108	2902
<i>A. salmiana</i>	Las Minas	20.49.332	99.36.027	2913	<i>A. salmiana</i>	Enxini	20.48.327	99.36.118	2901
<i>A. salmiana</i>	Espolón			3105	<i>A. intermedio</i>	Enxini	20.48.335	99.36.117	2897
<i>A. montana</i>	Espolón	20.47.717	99.33.992	3080	<i>A. gentryi</i>	Enxini	20.48.329	99.36.116	2896
<i>A. salmiana</i>	Enxini	20.47.752	99.34.165	2807	<i>A. salmiana</i>	Enxini	20.48.223	99.36.035	2910
<i>A. montana</i>	Enxini	20.48.216	99.35.998	2902	<i>A. salmiana</i>	Enxini	20.48.235	99.36.054	2915
<i>A. montana</i>	Enxini	20.48.208	99.35.978	2891					
<i>A. montana</i>	Enxini	20.48.182	99.35.972	2901					

## Anexo 2. Códigos R

### Análisis morfológicos

- SHAPIRO-WILK

```
library(readr)
```

```
datos <- read.csv("datos.csv")
```

```
nombres_columnas <- colnames(datos)
```

```
for (columna in nombres_columnas) { resultado_prueba <- shapiro.test(datos[, columna]);
cat("Variable:", columna, ", p-valor:", resultado_prueba$p.value, "\n") }
```

#### ▪ PRUEBAS DE DUNN

```
library(dunn.test)
```

```
datos <- read.csv("datos.csv")
```

```
datos$especie <- factor(datos$especie, levels = c("Agave_montana", "Agave_gentryi",
"Agave_salmiana", "Agave_intermedio"))
```

```
# Obtener los datos de cada grupo
```

```
grupo_1 <- datos$Altura_de_la_roseta[datos$especie == "Agave_montana"]
```

```
grupo_2 <- datos$Altura_de_la_roseta[datos$especie == "Agave_gentryi"]
```

```
grupo_3 <- datos$Altura_de_la_roseta[datos$especie == "Agave_salmiana"]
```

```
grupo_4 <- datos$Altura_de_la_roseta[datos$especie == "Agave_intermedio"]
```

```
mis_datos <- list(grupo_1, grupo_2, grupo_3, grupo_4)
```

```
comparaciones_dunn <- dunn.test(mis_datos)
```

```
colores <- c("brown4", "dodgerblue2", "chocolate1", "springgreen3")
```

```
boxplot(Altura_de_la_roseta ~ especie, data = datos, xlab = "Especie", ylab = "Altura de la
roseta", col = colores)
```

```
signif <- comparaciones_dunn$p.adjusted < 0.05
```

```
if (any(signif)) {
```

```
  comparisons <- comparaciones_dunn$comparison[signif, ]
```

```
  segments(
```

```
    x0 = as.numeric(comparisons[, 1]),
```

```
    x1 = as.numeric(comparisons[, 2]),
```

```
    y0 = par("usr")[3],
```

```
    y1 = par("usr")[3],
```

```
col = "red", lty = 2, lwd = 2

)

}
```

- Permanova y MDS todos

```
datos <- read.csv("datos.csv")

datos$especie <- factor(datos$especie, levels = c("Agave_montana", "Agave_gentryi",
"Agave_salmiana", "Agave_intermedio"))

permanova_result <- adonis2(datos[, c("altitud", "porciento_de_irradiacion_global",
"pendiente", "Altura_de_la_roseta", "ancho_de_la_roseta", "No._de_hojas",
"Longitud_de_hoja", "Ancho_de_la_hoja", "Longitud_de_la_espina_terminal",
"Ancho_de_la_base_de_la_espina_terminal", "margen", "forma_de_la_espina_terminal")] ~
especie, data = datos)

summary(permanova_result)

permanova_result$aov.tab

# Realizar MDS

library(vegan)

datos_numericos <- datos[, sapply(datos, is.numeric)]

mds_result <- metaMDS(datos_numericos)

coordenadas_mds <- scores(mds_result, display = "sites")

datos_mds <- data.frame(coordenadas_mds, especie = datos$especie)

library(ggplot2)

mds_plot <- ggplot(datos_mds, aes(x = NMDS1, y = NMDS2, color = especie)) +

  geom_point(size = 3) +

  ggtitle("MDS Plot - PERMANOVA") +

  theme(legend.position = "right",
```

```

panel.background = element_rect(fill = "white")) +

scale_color_manual(values = c("Agave_montana" = "brown4", "Agave_gentryi" =
"dodgerblue2", "Agave_salmiana" = "chocolate1", "Agave_intermedio" = "springgreen3")) +

stat_ellipse(aes(group = especie, color = especie), type = "norm", level = 0.95, linetype =
"dashed") +

theme_bw()

print(mds_plot)

```

- Permanova y MDS por especie

```

datos <- read.csv("ruta_archivo.csv")

datos$localidad <- factor(datos$localidad)

colores <- c("Espolon" = "orchid4", "Minas" = "aquamarine1", "Enxini" = "gold3")

permanova_result <- adonis2(datos[, c("altitud", "porciento_de_irradiacion_global",
"pendiente", "Altura_de_la_roseta", "ancho_de_la_roseta", "No._de_hojas",
"Longitud_de_hoja", "Ancho_de_la_hoja", "Longitud_de_la_espina_terminal",
"Ancho_de_la_base_de_la_espina_terminal", "margen", "forma_de_la_espina_terminal")] ~
localidad, data = datos)

library(vegan)

datos_numericos <- datos[, sapply(datos, is.numeric)]

mds_result <- metaMDS(datos_numericos)

coordenadas_mds <- scores(mds_result, display = "sites")

datos_mds <- data.frame(coordenadas_mds, localidad = datos$localidad)

library(ggplot2)

mds_plot <- ggplot(datos_mds, aes(x = NMDS1, y = NMDS2, color = localidad)) +

mds_plot <- ggplot(datos_mds, aes(x = NMDS1, y = NMDS2, color = localidad)) +
geom_point(size = 3) + ggtitle("MDS Plot - PERMANOVA") + theme(legend.position =
"right", panel.background = element_rect(fill = "white")) + scale_color_manual(values =
colores) + theme_bw()

```

```
print(mds_plot)
```

- PCA

```
library(tidyverse)
```

```
data <- read.csv("ruta_archivo.csv ")
```

```
library(textshape)
```

```
rownames(data) <- data$IDnombrelocalidad
```

```
data <- data[-which(names(data) == "IDnombrelocalidad")]
```

```
data_std <- scale(data)
```

```
pca_result <- prcomp(data_std)
```

```
summary(pca_result)
```

```
plot(pca_result)
```

```
biplot(pca_result)
```

```
print(pca_result)
```

```
str(pca_result)
```

```
library(factoextra)
```

```
library(FactoMineR)
```

```
# crear un gráfico de barras de la contribución de cada componente a la varianza
```

```
fviz_eig(pca_result)
```

```
individuos_pca <- pca_result$x
```

```
dim(individuos_pca)
```

```
data$especie <- factor(data$especie)
```

```
colores <- c("brown4", "dodgerblue2", "chocolate1", "springgreen3")
```

```
fviz_pca_ind(pca_result, geom = "point", col.ind = data$especie, palette = colores, addEllipses  
= TRUE)
```

```

    ■ FAMD
install.packages("FactoMineR")

install.packages("factoextra")

library(FactoMineR)

library(factoextra)

datos <- read.csv("ruta_archivo.csv ")

datos$especie <- as.factor(datos$especie)

resultado_famd <- FAMD(datos, ncp = 5, graph = FALSE)

print(resultado_famd)

colores <- c("brown4", "dodgerblue2", "chocolate1", "springgreen3")

fviz_famd_ind(resultado_famd,

               repel = TRUE,

               geom = "point",

               habillage = "especie",

               palette = colores) +

labs(title = "FAMD - Individuals factor map",

      x = "Dim 1",

      y = "Dim 2")

# Graficar los resultados del FAMD para las variables

fviz_famd_var(resultado_famd,

               repel = TRUE,

               palette = colores) +

labs(title = "FAMD - Variables factor map",

```

```
x = "Dim 1",  
y = "Dim 2")
```

- Clustering

```
library(cluster)  
  
library(fpc)  
  
library(factoextra)  
  
datos <- read.csv("ruta_archivo.csv ")  
  
datos_estandarizados <- scale(datos)  
  
k <- 3  
  
resultados_kmedoids <- pam(datos_estandarizados, k)  
  
colores <- c("red3", "chocolate1", "dodgerblue2")  
  
cluster_colors <- colores[resultados_kmedoids$clustering]  
  
plot(datos_estandarizados, col = cluster_colors)  
  
points(resultados_kmedoids$medoids, col = 1:k, pch = 16, cex = 2)  
  
silhouette_plot <- fviz_silhouette(resultados_kmedoids, palette = colores)  
  
print(silhouette_plot)  
  
cluster_profile <- clusplot(datos_estandarizados, resultados_kmedoids$clustering,  
                             color = TRUE, shade = TRUE, labels = 2, lines = 0, col.clus = colores)  
  
matriz_confusion <- table(resultados_kmedoids$clustering, datos$especie)  
  
print(matriz_confusion)
```

## **Diversidad y estructura genómica**

- PCA

```
vcf_obj <- read.vcfR("ruta_archivo.vcf")
```



```

Agaves.ind <- vcfR2genind(vcf_obj)

Ind_labels <- read.table("ruta_barcodes.txt",sep="\t",header=T)

pop(Agaves.ind) <- Ind_labels$sitio

popNames(Agaves.ind)

Agaves.fstat <- genind2hierfstat(Agaves.ind)

Agaves.genpop <- genind2genpop(Agaves.ind)

X <- scaleGen(Agaves.ind, NA.method="mean")

pca1.ALL <- dudi.pca(X,scannf=FALSE,scale=FALSE)

summary(pca1.ALL)

Cols<-c("orchid4", "aquamarine1", "gold3")

color_pallete_function <- colorRampPalette(colors =Cols,space = "Lab")

num_colors <- nlevels(pop(Agaves.ind))

Colores.cucu <- color_pallete_function(num_colors)

png("PCA_1.png")

plot(pca1.ALL$li,col=transp(Colores.cucu[pop(Agaves.ind)],0.8),pch=20,cex=2,xlab="Axis.1
XXX%",ylab="Axis.2 XXX%",bty="n",cex.axis=1.25,cex.lab=1.5)

abline(h=0,lty=3,col="gray75")

legend("topright",legend=levels(pop(Agaves.ind)),xpd=TRUE,pt.cex=2,text.font=1,cex=1.25,co
l=Colores.cucu,pch=20)

dev.off()

```

- HETEROCIGOSIDAD OBSERVADA Y ESPERADA

```

BasicSS <- basic.stats(Agaves.ind)

BasicSS

Div_Raw.Hs <-as.data.frame(BasicSS$Hs)

Div_Raw.Ho <-as.data.frame(BasicSS$Ho)

```

```

png("Div_pop.png")

par(mfrow=c(1,2))

barplot(colMeans(Div_Raw.Hs,na.rm=T),las=2,main="HE")

barplot(colMeans(Div_Raw.Ho,na.rm=T),las=2,main="HO")

dev.off()

```

- FST

```

Agaves_fst.mat <-pairwise.neifst(Agaves.fstat)

Agaves_fst.mat

```

- DENDROGRAMA

```

Agaves.dist <-dist(Agaves.ind@tab)

Agaves.clust<-hclust(Agaves.dist,method="ward.D2")

png("Inds_Dist.png")

plot(Agaves.clust, cex = 1.2, hang = 0.05)

dev.off()

```