



UNIVERSIDAD AUTÓNOMA CHAPINGO

Coordinación de Estudios de Posgrado

Departamento de Fitotecnia

Maestría en ciencias en Biotecnología Agrícola

Análisis fenotípico y filogenético de agaves mezcaleros de la región Valles Centrales de Oaxaca.

Tesis

Que como requisito parcial para obtener el grado de:

**MAESTRO EN CIENCIAS EN BIOTECNOLOGÍA
AGRÍCOLA**

Presenta:

MARÍA PATRICIA VÁSQUEZ MAYA

Bajo la supervisión del Dr. Juan Porfirio Legaria Solano



Chapingo, Estado de México a 13 de febrero del 2023



La presente tesis titulada **“Análisis fenotípico y filogenético de agaves mezcaleros de la región Valles Centrales de Oaxaca”**, fue realizada por la **C. María Patricia Vásquez Maya**, bajo la dirección del Dr. Juan Porfirio Legaria Solano y asesoría del Dr. José Luis Rodríguez De la O y Dr. Juan Martínez Solís, misma que fue aprobada y aceptada como requisito parcial para obtener el grado de:

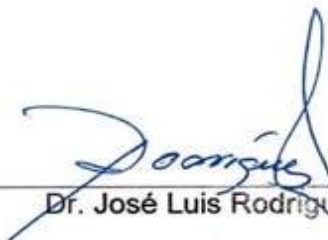
MAESTRO EN CIENCIAS EN BIOTECNOLOGÍA AGRÍCOLA

DIRECTOR:



Dr. Juan Porfirio Legaria Solano

ASESOR:



Dr. José Luis Rodríguez De la O

ASESOR:



Dr. Juan Martínez Solís

DEDICATORIA

Nde na par naa, tn' nesséh na r' nia dixquix naa pur griit' na bu' ná, dixquix Dios pur griit' na buzda lañ galbañ, nesse' h na r' nia dixquixtl griitl, dixquix xñaia cun xtada pur zugaiia laañ guixliiu.

A mi madre, por transmitir su fortaleza y apoyarme en todo momento, porque es el ser más maravilloso que tengo, es un ejemplo y por la que vale la pena luchar.

A mi padre por permitirme entender lo efímero que es la vida, porque cada momento es un aprendizaje.

A mi familia, por su apoyo incondicional y ayudarme a superar cada momento de dificultades, especialmente a Irene y Jaqueline por su apoyo en cada etapa de mi vida, a mis hermanos Silvano, Álvaro y Carlos porque son un ejemplo de superación.

A mi Andrea, por ser una pequeña tan llena de luz y alegría que me enseña a valorar y aprender más de la vida.

A Álvaro por su apoyo incondicional, y apoyarme en todo momento, por darme fortaleza siempre que lo necesito, y alentarme para sacar adelante este trabajo, por ser luz en mi vida y enseñarme a superar cada obstáculo, gracias por tanto amor.

A la maestra Socorro Ramírez y a su familia, por su apoyo en los momentos que necesito cobijo y no sentirme sola al estar lejos de casa.

A Gaby por su apoyo emocional cada que lo requería. A mis compañeras y amigas, por transmitir su calidez. Y a todas las personas restantes simplemente Gracias.

AGRADECIMIENTOS

A Dios, por no haber dejado que me rinda en ningún momento e iluminarme para salir adelante.

A la Universidad Autónoma Chapingo, por haberme dado la oportunidad de continuar mi formación académica y obtener el grado de Maestro en Ciencias.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT), por el apoyo económico otorgado durante el desarrollo de mis estudios de maestría.

Al Dr. Juan Porfirio Legaria Solano, por sus invaluable conocimientos brindados, asesoría y valiosa dirección en la elaboración de esta tesis, así mismo por el tiempo invertido en la revisión y conclusión de esta investigación.

Al Dr. José Luis Rodríguez De la O, por sus invaluable conocimientos brindados, asesoría y por haber mostrado una excelente disposición para revisar este trabajo de investigación.

Al Dr. Juan Martínez Solís, por sus invaluable conocimientos brindados, asesoría y por haber mostrado una excelente disposición para revisar este trabajo de investigación.

Al Departamento de Fitotecnia y a todas las personas que ayudaron de forma directa e indirectamente a la realización de este trabajo de investigación.

A Francisco Pérez de Banhez Mezcal y a la unión de productores agropecuarios del distrito de Ejutla (UPADEC) por su apoyo para llevar a cabo el muestreo en campo.

A la Asociación de Mujeres en el Mezcal junto con todos los productores que hicieron posible este trabajo.

A mis profesores, por sus enseñanzas y experiencias compartidas.

A Compañeros y cada persona que me ayudo en el proceso de cada una de las etapas de este trabajo.

DATOS BIOGRÁFICOS

Datos personales

Nombre completo: Vásquez Maya María Patricia

Fecha de

nacimiento: 26 de agosto de 1995

Lugar de nacimiento: Santiago Apóstol, Ocotlán,
Oaxaca, México.

CURP: VAMP950826MOCSYT01



Desarrollo académico

Bachillerato
2010-2014
Preparatoria Agrícola. Universidad Autónoma
Chapingo. Texcoco, Estado de México, México.

Licenciatura
2014-2018
Ingeniero Agrónomo Especialista en Parasitología
Agrícola, Universidad Autónoma Chapingo. Texcoco,
Estado de México, México.

Maestría
2020-2021
Maestría en Ciencias en Biotecnología Agrícola.
Coordinación de estudios de Posgrado, departamento
de fitotecnia, Universidad Autónoma Chapingo.
Texcoco, Estado de México, México.

CONTENIDO

	Pagina
DEDICATORIA.....	iii
AGRADECIMIENTOS	iv
DATOS BIOGRÁFICOS	v
Índice de cuadros	viii
Índice de figuras	ix
RESUMEN GENERAL	11
ABSTRACT	12
1 INTRODUCCIÓN GENERAL	13
1.1 Objetivo general	14
1.2 Objetivos específicos	14
2. REVISIÓN DE LITERATURA	14
2.1 <i>Agave</i> y la Clasificación taxonómica del género	14
2.2 Distribución geográfica del género <i>Agave</i>	15
2.3 Variabilidad fenotípica en agaves	16
2.4 Métodos multivariados	16
2.4.1 Análisis clúster	17
2.4.1.1 Técnicas jerárquicas	17
2.4.1.2 Técnicas no jerárquicas	18
2.4.1.3 K-medias	18
2.5 Métodos de ordenación	19
2.5.1 Análisis de componentes principales	19
2. 6 Marcadores moleculares	20
2.6.1 Región espaciadora transcrita interna (ITS)	20
2.6.2 Gen de la maturasa matK	20
2.6.3 trnH-psbA	21
2.7 Árboles filogenéticos	21
2.7.1 Tipos de parsimonia	21
2.7.2 Parámetros del árbol	22
2.7.3 BLUP y REML	22
2.8 Análisis filogenético bayesiano	23

2.8.1 Inferencia bayesiana.....	23
2.8.2 Teorema de bayes	23
2.8.3 Probabilidad <i>a priori</i>	24
2.8.4 Cadenas de Márkov Monte Carlo	25
2.8.5 Estudios filogenéticos anteriores del clado <i>Agave</i>	25
2.8.6 Filogenia actual para el género <i>Agave</i>	28
2.9 LITERATURA CITADA.....	29
2 CAPITULO I	35
1. RESUMEN	35
1.1 ABSTRACT	36
2. INTRODUCCIÓN	37
3. Objetivos	38
4. Hipótesis.....	38
5. MATERIALES Y MÉTODOS	38
5.1 Localización del área de estudio	38
5.2 Materiales	38
5.3 Tamaño de la muestra	39
5.4 Diseño experimental	40
5.5 Variables	40
5.6 Análisis estadístico.....	43
5.6.1 Agrupamiento jerárquico sobre componentes principales	44
5.6.2 Mapa de calor	44
5.6.3 K- medias.....	44
5.6.4 Análisis de componentes principales	45
6. RESULTADOS Y DISCUSIÓN	45
6.1 Análisis multivariado	45
6.1.1 Análisis clúster	45
6.1.1.1 Agrupamiento jerárquico.....	45
6.1.1.2 No jerárquica (K-medias)	54
6.1.2 Métodos de ordenación	55
6.1.2.1 Análisis de componentes principales.....	55
7. CONCLUSIONES.....	61

8. LITERATURA CITADA.....	62
3 CAPITULO II	64
1. RESUMEN.....	64
1. 1 ABSTRACT.....	65
2. INTRODUCCIÓN.....	66
3. Objetivos.....	66
4. MATERIALES Y MÉTODOS.....	66
4.1 Secuencias	67
4.2 Alineamiento de las secuencias	70
4.3 Concatenación de las secuencias.....	70
4.4 Construcción del árbol filogenético	70
5. RESULTADOS Y DISCUSIÓN.....	71
6. CONCLUSIONES.....	75
7. LITERATURA CITADA.....	76

Índice de cuadros

Cuadro 1. Identificación y descripción de las poblaciones muestreadas.	39
Cuadro 2. Variables cuantitativas (QN) y cualitativas (QL) evaluadas para el análisis fenotípico de los agaves.....	43
Cuadro 3. Eigenvalores y varianza explicada de 10 componentes.	56
Cuadro 4. Vectores propios para las variables en el ACP.....	56
Cuadro 5. Lista de taxones incluidos en los análisis de inferencia filogenética, siguiendo la clasificación infragénica de Agave propuesto por Gentry (1982) y Álvarez (1995).	67
Cuadro 6. Lista de accesiones del GenBank para las secuencias implementadas, consultadas de la lista de Jiménez <i>et al.</i> (2020) y otros autores.	68
Cuadro 7. Resumen general de los parámetros genéticos para cada locus.	71

Índice de figuras

Figura 1. Características diferenciales de plantas de agave mezcalero e inicio del ciclo de la floración. A. Planta en inicio de floración. B. Forma de la espina terminal: recta (a), curvada (b), filiforme (c). C. Intensidad del color de la hoja: débil (a), medio (b), fuerte (c). D. Forma de la hoja: lineal (a), espatulada (b), deltoide (c), lanceolada (d), oblonga (e), ovada (f). E. Espinas: presentes (a) o ausentes (b). F Forma de las espinas laterales: recta (a), curva (b), ganchuda (c). G. Hábito de crecimiento: acaulescente (a), caulescente (b).	42
Figura 2. Número óptimo de grupos vs variación intra-grupo sobre el método de K-medias aplicadas a la MBD de <i>Agave</i>	45
Figura 3. Dendrograma correspondiente al agrupamiento jerárquico (modo Q) basado en el ligamiento promedio generado a partir de la MBD estandarizada de las especies del género <i>Agave</i>	46
Figura 4. Dendrograma de relaciones entre variables evaluadas correspondiente al agrupamiento jerárquico (modo R) basado en el ligamiento promedio generado a partir de la MBD estandarizada de las especies del género <i>Agave</i> . La distancia se representa en la escala horizontal.	49
Figura 5. Mapa de calor que muestra las relaciones existentes entre las especies pertenecientes al género <i>Agave</i> sp. Con las variables QL y QN evaluadas de la caracterización morfológica en campo.	50
Figura 6. Fenotipo de una planta típica del grupo Tepeztate (<i>Agave marmorata</i>).	51
Figura 7. Grupo potatorum, (a) <i>Agave nussaviorum</i> , (b) <i>Agave potatorum</i> , (c) <i>Agave semanniana</i>	52
Figura 8. Grupo Karwinskii, <i>Agave karwinskii</i> , (a) kuish, (b) barril, (c) marteño. ...	52
Figura 9. Fenotipo de una planta correspondiente al grupo Espadín, <i>Agave angustifolia</i>	53
Figura 10. Fenotipo de una plata correspondiente al grupo Mexicano, <i>Agave rhodacantha</i>	54
Figura 11. K-medias presentadas por medio de un análisis de componentes principales. La variación explicada por cada componente se indica por medio de los	

porcentajes de los ejes. Las especies se muestran en cinco grupos junto con sus centroides.....	55
Figura 12. Porcentaje de varianza explicada de los componentes.....	57
Figura 13. Contribución de cada variable a los componentes principales.	57
Figura 14. Contribución de cada población a los componentes principales. Los valores por arriba de la línea horizontal promedio de color rojo son altamente significativos.	58
Figura 15. Representación de poblaciones y variables de componentes principales sobre el plano de PCA.	59
Figura 16. Grafica de Componente principales dispuestas en 3D.....	61
Figura 17. Filogenia molecular de <i>Agave</i> sensu lato utilizando <i>Yucca</i> como un grupo externo reconstruido a partir de las secuencias ITS y utilizando un análisis de inferencia bayesiana. Los números junto a los nodos indican los valores de probabilidad posteriores	72
Figura 18. Filogenia molecular de <i>Agave</i> usando a <i>Yucca</i> como un grupo externo, reconstruido a partir de las secuencias del conjunto de datos del cloroplasto y usando un análisis de inferencia Bayesiana. Los números al lado de los nodos indican los valores de probabilidad	73

RESUMEN GENERAL

Se analizaron 275 individuos de poblaciones correspondientes a las especies *Agave angustifolia*, *Agave karwinskii*, *Agave marmorata*, *Agave rhodacanta*, *Agave potatorum*, *Agave semanniana* y *Agave nussaviorum*, con el propósito de evaluar su variabilidad morfológica, utilizando un análisis estadístico multivariado, análisis de agrupamiento jerárquico, K-medias y análisis de Componentes Principales, mismos que mostraron las relaciones que guardan entre sí las especies de *Agave* estudiadas. *Agave marmorata* fue la especie más lejana o menos relacionada debido a la variable de la textura rugosa de sus hojas. El complejo *potatorum* (que incluye a las especies *Agave potatorum*, *Agave nussaviorum* y *Agave semanniana*) se diferenció por la forma de la espina terminal. *Agave angustifolia* presentó plantas altas y de crecimiento rápido. La especie *Agave karwinskii* fue caracterizada por su hábito de crecimiento y el número de hojas. El grupo mexicano *Agave rhodantha* sus variables no mostraron alta correlación en el análisis del PCA, ni con otro método utilizado. Los complejos varietales *potatorum* y *karwinskii*, fueron los que mostraron mayor variabilidad fenotípica dentro de la familia Asparagáceae. En el análisis de secuencias de ADN (obtenidas de GenBank), para reconstruir relaciones filogenéticas entre especies del género *Agave*, se incluyeron 62 especies de la familia Asparagaceae, de las cuales 46 corresponden al clado *Agave*, seleccionándose también otras 16 especies fuera del género *Agave*, se utilizó al género *Yucca* como grupo externo principal. Para la reconstrucción de los árboles filogenéticos se utilizó el programa de MrBayes. El análisis de las regiones del espaciador transcrito ITS, mostró buena resolución en los clados del árbol filogenético usando el método de inferencia bayesiana y donde los taxones se separan con más resolución y soporte que el árbol construido con genes de cloroplastos, MatK y trnH-PsbA.

Palabras clave: Asparagáceae, análisis multivariado, ITS, matK, trnH-PsbA, filogenia molecular.

Tesis de Maestría en Ciencias, Universidad Autónoma Chapingo.

Autor: Vasquez Maya María Patricia

Director: Dr. Juan Porfirio Legaria Solano

ABSTRACT

A total of 275 individuals from populations corresponding to the species *Agave angustifolia*, *Agave karwinskii*, *Agave marmorata*, *Agave rhodacanta*, *Agave potatorum*, *Agave semanniana* and *Agave nussaviorum* were evaluated with the purpose of evaluating their morphological variability, a multivariate statistical analysis, hierarchical clustering analysis, K-means and Principal Component Analysis (PCA) were used, which showed the relationships among the *Agave* species studied. *Agave marmorata* was the most distantly or least related species, this was attributed to the variable rough texture of its leaves. The potatorum complex (which includes the species *Agave potatorum*, *Agave nussaviorum* and *Agave semanniana*) was differentiated by the shape of the terminal spine. *Agave angustifolia* presented tall and fast-growing plants. The species *Agave karwinskii* were characterized by its growth habit and number of leaves. For the variables of the mexican group (*Agave rhodantha*) did not show a high correlation in the PCA analysis, nor with any other method used. The varietal complexes of *potatorum* and *karwinskii* showed the greatest phenotypic variability within the family of *Asparagaceae*. In the analysis of DNA sequences (obtained from GenBank) was carried out to reconstruct phylogenetic relationships among species of the genus *Agave*. A total of 62 species of the family *Asparagaceae* were included, of which 46 corresponded to the *Agave* clade, and 16 other species outside the genus *Agave* were also selected, of which the genus *Yucca* was taken as the main external group. The MrBayes program was used to reconstruct the phylogenetic trees. The analysis of the ITS transcribed spacer regions showed good resolution in the clades of the phylogenetic tree the Bayesian inference method was used and where taxa was separated with higher resolution and support than the tree constructed with chloroplast genes, *matK* and *trnH-PsbA*.

Keywords: *Asparagaceae*, multivariate analysis, ITS, *matK*, *trnH-PsbA*, molecular phylogeny.

Thesis, Universidad Autónoma Chapingo.

Author: Vásquez Maya María Patricia

Advisor: Dr. Juan Porfirio Legaria Solano

1 INTRODUCCIÓN GENERAL

Los agaves se localizan prácticamente en toda la República Mexicana, son endémicos del continente americano y se distribuyen desde la parte sur de los Estados Unidos de Norteamérica hasta Colombia y Venezuela; la mayor riqueza se encuentra en las zonas áridas, semiáridas y templadas; y se ubican a altitudes que van desde el nivel del mar hasta un máximo de 3000 msnm (García-Mendoza, 2011). Por la gran diversidad de agaves que tiene, México es considerado centro de origen de este género ya que existe una gran variabilidad morfológica en plantas silvestres y domesticadas ligada a la variedad de climas, ambientes y mejoramiento tradicional, encontrándose alrededor de 200 especies para este género (García, 2002), de las cuales un 75% se distribuye en este país. Para el caso específico del estado de Oaxaca, García *et al.* (2004; 2007) indicaron que existen 37 especies registradas, de las que siete son endémicas y la zona con mayor número de especies es la provincia Tehuacán-Cuicatlán, con 15 especies.

Para el caso de mezcal, de acuerdo con cifras del sistema de información Agroalimentaria de consulta (SIACON, 2020) la producción y superficie sembrada de agave mezcalero ha crecido en los últimos diez años. De 2008 a 2018 la producción alcanzó un máximo de 301 790 a 303 382 t, mientras que la superficie sembrada lo hizo de 4 770 a 5 360 ha.

Uno de los retos que se presentan a la hora de estudiar la diversificación de un grupo es comprender los factores que influyen en las tasas de especiación y extinción, ya que los posibles factores involucrados son muchos y es difícil desentrañar cuál o si varios de ellos la están afectando (Zamora-Abrego *et al.*, 2013). El método comparativo permite hacer comparaciones formales y estadísticas entre especies y así analizar información, como caracteres morfológicos, fisiológicos o ecológicos, mediante la incorporación de información sobre las relaciones filogenéticas del grupo de interés (Harvey & Pegel, 1991; Garamszegi, 2014). De esta manera, el método comparativo ha permitido evaluar en qué medida la variación de un personaje se debe a su historia evolutiva o a presiones adaptativas (Morales, 2000; Rezende & Galard, 2003).

Uno de los retos que se presentan a la hora de estudiar la diversificación de un grupo es comprender los factores que influyen en las tasas de especiación y extinción, ya que los posibles factores involucrados son muchos y es difícil desentrañar cuál o si varios de ellos están afectando a la diversificación del género *Agave*.

1.1 Objetivo general

Evaluar la diversidad genética existente en especies de *Agave* cultivadas en la región de Valles Centrales, Oaxaca, utilizando caracteres morfológicos cualitativos y cuantitativos y adicionando un estudio filogenético.

1.2 Objetivos específicos

- Evaluar la diversidad fenotípica de cinco grupos de especies de agave mezcalero, utilizando marcadores morfológicos cualitativos y cuantitativos.
- Realizar una reconstrucción filogenética del género *Agave*, aumentando sustancialmente el muestreo taxonómico y seleccionando los marcadores moleculares apropiados, en relación con estudios previos, con el fin de obtener un mayor nivel de resolución y valores de soporte.

2. REVISIÓN DE LITERATURA

2.1 *Agave* y la Clasificación taxonómica del género

El género *Agave* se ubica en la familia *Agavaceae*, donde se incluyen varias especies de plantas adaptadas a condiciones de aridez; tienen una forma característica de roseta y poseen raíces muy ramificadas, cutícula gruesa, hojas suculentas con estomas hundidos y metabolismo fotosintético tipo CAM. Por otro lado, el crecimiento muy lento de estas plantas, así como sus bajas tasas de reproducción asexual y reproducción sexual limitada por problemas de polinización y viabilidad de las semillas, son factores que hacen a los agaves difíciles de multiplicar masivamente por métodos convencionales. Estos mismos factores limitan las posibilidades de mejoramiento de las especies cultivadas (Domínguez *et al.*, 2008).

Debido a la antigua e intensa relación entre los humanos y los agaves existen variantes con importancia regional, nacional e internacional, de las cuales hay un conocimiento tradicional considerable y una variación morfológica alta (Mora *et al.*, 2011).

Taxonomía

Reino:	Plantae
División:	Angiospermae
Clase:	Monocotyledonea
Orden:	Asparagales
Familia:	Asparagaceae
Subfamilia:	Agavoideae
Género:	<i>Agave</i>
Subgénero:	<i>Agave</i>
Especie:	<i>Agave</i> spp.

(APGIII, 2009).

El género *Agave* pertenece a la subfamilia Agavoide, la cual está integrada dentro de la familia *Asparagaceae*, esta ha sufrido algunas modificaciones en sus clasificaciones. De 1998 a 2009, de acuerdo con las clasificaciones de la APG I y II (Angiosperm Phylogeny Group) el género *Agave* fue considerado dentro de la familia *Agavaceae* (APG, 2009). Con la implementación de la clasificación APG III basada en secuencias moleculares de ADN la familia *Agavaceae* fue removida y desaparecida de la lista de familias, con esto el género *Agave* pasó a ser miembro de la familia *Asparagaceae* dentro de la subfamilia Agavoide (Chase *et al.*, 2009).

2.2 Distribución geográfica del género *Agave*

México es el centro de origen y diversidad natural del género *Agave* mismo que se distribuye principalmente en ecosistemas áridos. Su importancia ecológica radica en el gran número de especies animales y vegetales con las cuales establece asociaciones y en su capacidad de retención de suelos (Esqueda *et al.*, 2011). El éxito de los agaves se basa en su adaptación a climas extremos, con poca demanda de agua, creciendo principalmente en zonas áridas y semiáridas. A pesar de esta

aparente falta de disponibilidad de agua, algunas de estas áreas áridas de los Estados Unidos y México son muy diversas biológicamente. Por ejemplo, las zonas áridas y semiáridas de México incluyen 20% de la flora total del país, que comprende 6000 especies con niveles muy altos de endemismo (60%) (Rzedowski, 1993).

2.3 Variabilidad fenotípica en agaves

Parker *et al.* (2014) realizaron un estudio de variables fenotípicas en seis poblaciones desconocidas de *Agave parryi* var. *huachaensis*, comparándolas con 17 poblaciones de *Agave parryi* previamente determinadas, para establecer si estos caracteres estaban influenciados por la forma de vida silvestre o cultivada de civilizaciones precolombinas. Encontraron que toda la morfología variaba en cierto grado, de manera más significativa en la variedad, además el carácter más explicativo para diferenciar las poblaciones silvestres de las manejadas fue el espacio inter-espigas con mayor variación Inter poblacional. El único carácter que no cambió significativamente en todas las poblaciones fue el grosor de las hojas. Demostrando con esto que las poblaciones con menor variación de formas coinciden con puntos de humanización y establecimiento de comunidades precolombinas. Además, obtuvieron poblaciones poco diversificadas, catalogándolas como potenciales puntos de búsqueda arqueológica. Los cambios morfológicos fueron muy pronunciados en reducción o variación de formas en los valles o partes bajas de Arizona, lugares aptos para el establecimiento de poblaciones antropogénicas y con esto el manejo de especies de *Agave*.

2.4 Métodos multivariados

Los métodos multivariados se aplican en la biología desde el siglo XX. Un aspecto común a las aplicaciones del análisis multivariado es que todas consideran un conjunto de objetos, donde cada objeto es descrito por una serie de atributos o variables. Estos objetos pueden ser conjuntos de individuos, especímenes, taxones, comunidades o cuadrantes geográficos, entre otros. El análisis multivariado intenta encontrar patrones de similitud entre objetos sobre la base de las variables utilizadas. Estos patrones permiten formar grupos cuyos objetos son más similares

entre sí, que con los objetos integrantes de otros grupos. Así mismo, el análisis multivariado busca identificar aquellas variables que permiten discriminar dichos grupos de objetos. Los patrones resultantes del análisis multivariado permiten contrastar hipótesis sobre las relaciones entre los objetos y explicar la causalidad de los agrupamientos, así como también predecir objetos y variables todavía no descubiertos (Palacio *et al.*, 2020).

2.4.1 Análisis clúster

El análisis clúster en castellano se conoce como análisis de conglomerados. Cuando se trabaja con datos de expresión de genes nuestra información es la matriz de expresión. Podemos considerar como observaciones a clasificar los genes. En este caso, la observación son los niveles de expresión observados en todas las muestras, el perfil de expresión del gen sobre las muestras analizadas. Se trata de clasificar las observaciones en grupos de modo que las observaciones de un mismo grupo sean lo más similares que se pueda y que los grupos entre si sean muy distintos. Entre los muchos procedimientos de obtener los grupos a partir de los datos, los más utilizados son dos tipos: procedimientos jerárquicos y métodos de particionamiento (Ayala, 2021).

2.4.1.1 Técnicas jerárquicas

Originan grupos que presentan rangos, en los cuales las unidades de estudio (UE) o grupos de UE subordinados forman parte de un grupo mayor o inclusivo (Palacio, 2020). La idea de estos procedimientos es construir una jerarquía de particiones del conjunto de índices. Sea $\{1, \dots, n\}$ el conjunto de índices que indexan las distintas observaciones. Supongamos que $\{C_1, \dots, C_r\}$ es una partición de este conjunto de índices:

$C_i \subset \{1, \dots, n\}$; son disjuntos dos a dos , $C_i \cap C_j = \emptyset$ si $i \neq j$ con $i, j = 1, \dots, r$ y

$$\bigcup_{j=1}^r C_j = \{1, \dots, n\}$$

Dada una partición del conjunto de índices podemos calcular la matriz $r \times r$ que en la posición (i, j) tiene la disimilitud entre el conjunto C_i y C_j , $d(C_i, C_j)$ según algunos de los procedimientos antes indicados. La representación gráfica muy utilizada para describir los resultados de un clúster jerárquico aglomerativo tiene el nombre de

dendrograma. El dendrograma va mostrando a que valor de la medida de disimilitud se produce la unión de los grupos y simultáneamente que grupos se están uniendo para esta disimilitud. También permite una valoración rápida de cuantos grupos puede haber en el banco de datos (Ayala, 2021).

2.4.1.2 Técnicas no jerárquicas

Originan grupos que no exhiben rangos. Las técnicas de ordenación o escalamiento multidimensional pueden clasificarse en escalamiento multidimensional métrico o no métrico (Legendre & Legendre, 1998). A diferencia de los análisis de agrupamiento los métodos de ordenación son no jerárquicos.

Los procedimientos de aglomeración no jerarquizados también se llaman métodos de aglomeración de K-medias.

2.4.1.3 K-medias

En k-medias el número de grupos se establece *a priori* del análisis. Supongamos que tenemos $C_1, \dots, C_k$ una partición de $\{1, \dots, n\}$. Un modo bastante natural de valorar la calidad del agrupamiento que la partición nos indica sería simplemente considerar la siguiente función.

$$\sum_{i=1}^k \sum_{j \in C_i} dE(x_j, \bar{x}_{C_i})^2, \quad (1)$$

Donde dE denota aquí la distancia euclídea y

$$\bar{x}_{C_i} = \frac{1}{|C_i|} \sum_{j \in C_i} x_j,$$

El procedimiento de agrupamiento de las k-medias simplemente se basa en elegir la partición de los datos de aquella que nos da el mínimo de la función objetivo considerada en ecuación (1) (Ayala, 2021).

Con centros desconocidos, el método K-medias comienza con una división del conjunto de los datos en (x) grupos configurados al azar y posteriormente busca mejorar esta primera clasificación reasignando los elementos al centroide del cluster más cercano, tratando de reducir la distancia media entre cada elemento de un grupo y su centroide (de la Fuente, 2011).

2.5 Métodos de ordenación

Los métodos de ordenación son no jerárquicos, reducen sin gran pérdida de información, el número total de dimensiones (p), y de esa manera facilitan la representación de las UE y sus relaciones en función de las variables empleadas. En un espacio de ordenación, las relaciones entre las UE están reflejadas en la posición en que se disponen en ese espacio, todos los métodos de ordenación producen una representación de las UE en el espacio euclidiano, preservando las distancias originales de la mejor forma posible. Las técnicas de ordenación o escalamiento multidimensional pueden clasificarse en escalamiento multidimensional métrico o no métrico (Palacio *et al.*, 2020).

2.5.1 Análisis de componentes principales

Cuando tomamos medidas sobre genes, personas, objetos, empresas, unidades experimentales de un modo genérico, se tiende a recoger el máximo de variables posibles. En consecuencia, tenemos dimensiones del vector de características X grandes. Una opción consiste en sustituir la dimensión d , por k combinaciones lineales de las mismas. Obviamente pretendemos que k sea mucho menor que d . El objetivo es elegir k de modo que expresen una proporción razonable de la dispersión o variación total cuantificada como la traza de la matriz de covarianza muestral, $tr(S)$, sea X un vector aleatorio de dimensión d con vector de medias μ y matriz de covarianzas Σ . Sea $T = (t_1, t_2, \dots, t_d)$ (los t_i indican la i -ésima columna de la matriz) la matriz ortogonal tal que:

$$T' \Sigma T = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_d),$$

Donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d \geq 0$ son los valores propios de la matriz Σ .

$$Y = T' (X - \mu).$$

Si denotamos la j -ésima componente de Y como Y_j entonces $Y_j = t_j' (X - \mu)$ con $j = 1, \dots, d$. A la variable Y_j la llamamos la j -ésima componente principal de Y . La variable $Z_j = \frac{Y_j}{\sqrt{\lambda_j}}$ es la j -ésima componente principal estandarizada de Y .

Estas componentes tienen algunas propiedades de gran interés. Notemos que el vector t_j tiene longitud unitaria y, por lo tanto, Y_j no es más que la proyección ortogonal de $X - \mu$ en la dirección t_j (Ayala, 2021).

2. 6 Marcadores moleculares

Un marcador molecular o de ADN es la diferencia en la secuencia de nucleótidos de ADN, entre organismos o especies individuales, que está cerca o estrechamente vinculado a un gen objetivo que expresa un rasgo. Por lo general, el gen objetivo, el rasgo expresado o la función biológica y el marcador molecular estrechamente vinculado asociado se heredan juntos (Collar *et al.*, 2005). La estrecha asociación de los marcadores moleculares con un rasgo o gen de una función biológica particular hace que los marcadores sirvan como signos o indicadores prácticos que señalan un locus de un gen en particular y ayudan a la detección o identificación de los rasgos asociados, ya sea que los genes involucrados sean conocidos o desconocidos y si los genes se pueden detectar o no (Reddy *et al.*, 2021).

2.6.1 Región espaciadora transcrita interna (ITS)

En cuanto a los marcadores de DNA, el ITS se refiere a dos regiones no codificantes del DNA nuclear; el ITS1 que se localiza entre los genes ribosomales 18S y 5.8S y el ITS2 entre los genes 5.8S y 26S. Una región espaciadora transcrita interna (ITS) del cistrón ribosómico nuclear es el locus de secuencias más utilizado para investigaciones sistemáticas moleculares de las plantas (Selvaraj *et al.*, 2005).

2.6.2 Gen de la maturasa matK

El gen *matK* del cloroplasto tiene una longitud de 1500 pb y se encuentra dentro del intrón del *trnK* y codifica para la proteína tipo maturasa, que está implicada en el corte y empalme del intrón del Grupo II. Los dos exones del gen *trnK* que flanquean al *matK* se perdieron, dejando el gen intacto en caso de corte y empalme. El gen contiene altas tasas de sustitución dentro de la especie y está emergiendo como candidato potencial para estudiar la evolución y la sistemática de las plantas. El gen *matK* tiene un tamaño ideal, una alta tasa de sustitución, una gran proporción de variación a nivel de ácido nucleico en la posición del primer y segundo codón, una baja proporción de transición/transversión y la presencia de sectores mutacionalmente conservados. Estas características del gen *matK* se explotan para resolver relaciones a nivel de familia y especie (Selvaraj *et al.*, 2008).

2.6.3 trnH-psbA

El fragmento psbA-trnH es una secuencia no codificante entre el gen psbA y el gen trnH en el genoma del ADN del cloroplasto, con una longitud de alrededor de 300 pb. La tasa de evolución de psbA-trnH es mucho más rápida que la del gen matK, pero ligeramente más lenta que la de ITS, que puede usarse para estudios filogenéticos entre grupos de plantas y entre especies. En las plantas, no existe un único código de barras de ADN con un éxito comparable. La mayoría de los códigos de barras de ADN de plantas se encuentran en el genoma del cloroplasto, ya sea dentro de secuencias de codificación (como *rbcLa* y *matK*) o en regiones intergénicas (como *trnH-psbA*) (CBOL Plant Working Group, 2009).

2.7 Árboles filogenéticos

Los árboles filogenéticos, también llamados cladogramas, son representaciones de estructuras jerárquicas que simbolizan las relaciones evolutivas entre los organismos o los genes. Los árboles pueden estar enraizados (cuando se utiliza el grupo externo para polarizar los caracteres) o no enraizados (cuando no se aplica el concepto de polaridad en la interpretación del árbol). Los árboles pueden ser completamente dicotómicos (resueltos), en los que cada nodo está conectado con los nodos, o taxones terminales, o con un nodo y un taxon terminal; o politómicos (parcialmente resueltos), en el que uno o más nodos están conectados con tres o más nodos o taxones terminales. Una politomia puede representar dos situaciones: divergencia simultánea de varios taxones descendientes (por ejemplo radiación adaptativa) o incertidumbre en las relaciones filogenéticas entre los taxones incluidos en la politomia (Palacio *et al.*, 2020).

2.7.1 Tipos de parsimonia

La palabra parsimony en el contexto de árboles filogenéticos fue utilizada por primera vez por Camin y Sokal en 1965, y en español se traduce como parsimonia. El método de parsimonia es de simple aplicación manual cuando los taxones terminales son pocos, pero de una imposible aplicación manual cuando los taxones superan un cierto número. Con más de 30 taxones (aproximadamente) es imposible garantizar una solución óptima, incluso computacionalmente (Palacio *et al.*, 2020).

2.7.2 Parámetros del árbol

El árbol hallado se define en función de la cantidad de cambios o pasos entre los estados de los caracteres que sustentan sus relaciones. Cuanto mejor sea el ajuste de los datos al árbol, menor será la cantidad de homoplasias necesarias. La longitud del árbol corresponde a la cantidad de cambios requeridos. El árbol que minimiza la cantidad de homoplasias es el más parsimonioso (Palacio *et al.*, 2020). Podemos decir que hay una homoplasia cuando dos o más taxones comparten caracteres similares o idénticos, y estos no derivan del ancestro común (similares características con distinto origen evolutivo). La homoplasia ha sido rescatada por diversos autores como un concepto importante dentro de la teoría evolutiva. A pesar de las controversias, es imposible distinguir entre los procesos de convergencia, paralelismo y reversión (Hall, 2007).

2.7.3 BLUP y REML

El BLUP se puede interpretar como un método bayesiano, precisa mejor la naturaleza del estimador y sus propiedades que en el caso frecuentista. Desde el punto de vista frecuentista, para que el BLUP sea insesgado hace falta primero usar la definición *ad hoc* de sesgo, pero además hace falta utilizar los parámetros genéticos verdaderos, que al no conocerse no permite asegurar que la estimación sea correcta. Desde el punto de vista bayesiano no existe el insesgamiento por que no se repite el experimento infinitas veces, la estimación es la de máxima probabilidad; además, los parámetros no necesitan ser los verdaderos sino unos parámetros que el genetista utiliza a su criterio para condicionar la función de densidad posterior. Finalmente, los efectos fijos no tienen una naturaleza diferente de los aleatorios; son simplemente efectos que tienen una distribución *a priori* constante en lugar de una normal. La interpretación bayesiana del REML (Máxima verosimilitud restringida) justifica mejor el método de interpretación frecuentista, el REML se estima en un espacio restringido, utilizando unos datos transformados que son siempre menos que los datos originales, lo que supone una pérdida de grados de libertad y puede suponer una pérdida de información. De hecho, en la estimación de la varianza de modelo más simple, en el que los datos se distribuyen de forma N

(μ, σ^2) , se puede demostrar que la estimación REML tiene un riesgo cuadrático mayor de la estimación máximo verosímil (Blasco, 2017).

2.8 Análisis filogenético bayesiano

2.8.1 Inferencia bayesiana

La inferencia bayesiana fue propuesta separadamente por el reverendo Thomas Bayes en 1763 y el marqués de Laplace en 1774, y desarrollada por Laplace a finales del siglo XVIII y principios del XIX. Ciertas dudas filosóficas sobre su uso, y sobre todo serias dificultades de cómputo, hicieron que fuera sustituida desde los años 30 del siglo pasado por la estadística clásica que hoy comúnmente se aplica. Sin embargo, el desarrollo de los computadores ha permitido resolver las dificultades de cómputo de esta escuela de inferencia, y desde los años 90 la escuela Bayesiana ha sido cada vez más aplicada, usándose en la actualidad virtualmente en todos los campos científicos (Blasco, 2017)

2.8.2 Teorema de bayes

La forma esencial de trabajar de la escuela bayesiana consiste en, dados los datos observados en el experimento, describir toda la incertidumbre que puede existir en torno a un parámetro (o a un valor desconocido, como la diferencia entre dos tratamientos), usando como medida natural de la incertidumbre la probabilidad de que el parámetro tome determinados valores. En la estadística clásica esto no es posible. Si una diferencia entre dos tratamientos tiene un *P-valor* del 3%, esto no indica que la probabilidad de que los tratamientos sean diferentes sea del 3%, lo que indica es que, si repitiéramos el experimento, la probabilidad de que una nueva diferencia entre tratamientos obtuviera valores iguales o superiores a los obtenidos serían del 3%, lo que no es lo mismo. De hecho, se ha demostrado que *P-valores* de 0.05, usados habitualmente para establecer diferencias significativas, no indican una evidencia del 95% de que los tratamientos sean diferentes, si no que en esta evidencia significativa nos encontramos con “falsos positivos”, por lo que es muy corriente afirmar que dos tratamientos son diferentes sin evidencias suficientes. Por el contrario, la inferencia Bayesiana da como resultado la probabilidad de que la diferencia entre dos tratamientos sea mayor que cero (es decir que el tratamiento 1

sea superior al tratamiento 2), y afirma que es, por ejemplo, del 97%. (Johnson, 2013; Blasco, 2017).

El Teorema de Bayes se aplica según la fórmula siguiente:

La probabilidad $P(A_i|B)$ viene dada por la expresión:

Sea B un suceso cualquiera del que se conocen las probabilidades condicionales $P(B|A_i)$.

$$P(A_i|B) = \frac{P(B|A_i)(A_i)}{P(B)}$$

Donde:

(A_i) : probabilidades *a priori*.

$P(B|A_i)$: probabilidad de Ben la hipótesis A_i .

$P(A_i|B)$: probabilidades *a posteriori*.

2.8.3 Probabilidad *a priori*

La probabilidad se puede aplicar a los datos de un experimento, por ejemplo, si estamos estimando la heredabilidad de un carácter y decidimos no seleccionar a menos que la heredabilidad sea superior a 0.10, el teorema de Bayes nos permite calcular la probabilidad de que la heredabilidad sea superior a 0.10, basándonos en los datos y de nuestro experimento, según la fórmula siguiente:

$$P(h^2 > 0.1|y) = \frac{P(y|h^2 > 0.1).P(h^2 > 0.1)}{P(y)}$$

Para calcular la probabilidad de que la heredabilidad sea mayor que 0.10 dados nuestros datos (y) tenemos que conocer la probabilidad de nuestros datos y cuando la heredabilidad es mayor que 0.10, pero esto no suele ser un problema porque la distribución de los datos es habitualmente conocida, y muy a menudo es la distribución normal. Por otra parte, la probabilidad de los datos es una constante, nuestros datos tienen una probabilidad determinada. El problema reside en $P(h^2 > 0.1)$, que es la probabilidad de que la heredabilidad sea mayor que 0.10 independientemente de nuestros datos. Esta probabilidad se basa en la información que tenemos antes de hacer el experimento, y no depende por tanto de nuestros datos; es lo que se conoce como probabilidad *a priori* (Blasco, 2017).

2.8.4 Cadenas de Márkov Monte Carlo

La idea de MCMC es proporcionar un conjunto de muestras aleatorias extraídas de una función de densidad de probabilidad, en lugar de usar la expresión matemática de la función. El primer uso de muestras aleatorias de una función de densidad de probabilidad fue propuesto por Student, estadístico de la cervecera Guinness, en un famoso artículo en el que presento la distribución t. El uso de las llamadas cadenas de Markov para encontrar estas muestras aleatorias tiene sus orígenes en el trabajo desarrollado en Los Álamos al final de la Segunda Guerra Mundial para resolver problemas relacionados con armas nucleares, aprovechando la aparición del primer ordenador, que permitía hacer operaciones con armas nucleares, aprovechando la aparición del primer ordenador, que permitía hacer operaciones antes impensables. Mucho más tarde, Gelfand y Smith en 1990 introdujeron esta técnica en el mundo estadístico para obtener distribuciones de probabilidad (Sorensen & Gianola, 2002). La estimación Bayesiana con MCMC provee cadenas con muestras aleatorias de las distribuciones marginales posteriores de cada parámetro o efecto desconocido, a partir de las cuales pueden hacer las inferencias. La longitud de estas cadenas es muy variable, dependiendo de la complejidad del problema. Para una estimación de efectos fijos puede bastar con unos miles de muestras por cadena, pero para estimar parámetros genéticos multivariantes en modelos no lineales pueden ser necesarios centenares de miles incluso millones de muestras. El motivo es que cuando hay muchos parámetros correlacionados en modelos complejos, las muestras consecutivas de las cadenas suelen estar correlacionadas, muestreando la misma zona de la función de densidad (Sorensen & Gianola, 2002).

2.8.5 Estudios filogenéticos anteriores del clado *Agave*

Se observó que diferentes tratamientos taxonómicos diferían ampliamente en los parientes sugeridos de *Agave*, y hubo debate sobre la delimitación de la familia *Agave* o incluso sobre la existencia de las *Agavaceae* (Eguiarte, 1995). La familia *Agave* ahora se considera la subfamilia *Agavoideae* dentro de la gran familia *Asparagaceae*, según el APG III (Angiosperm Phylogeny Group, 2009) y APG IV (Angiosperm Phylogeny Group, 2016).

Eguiarte (1995) fechó el origen de *Agave* y su familia usando un gen del cloroplasto y diferentes secuencias producidas para la filogenia de angiospermas monocotiledóneas (Duvall *et al.*, 1993), concluyó que 13.7 millones de años atrás (Ma) fue la fecha de origen más confiable para las *Agavaceae* según lo delimitado por Dahlgren *et al.* (1985). Es importante mencionar que, si bien esclarecedor, la estimación de las fechas de los relojes moleculares está sujeta a grandes incertidumbres, dados los métodos utilizados, el sesgo heredado de los métodos, los puntos de calibración empleados, así como el muestreo, y por lo tanto debe considerarse como una hipótesis de trabajo sujeta a cambios cuando se disponga de más datos y mejores métodos (Eguiarte *et al.*, 2021).

Bogler y Simpson (1996) utilizaron secuencias de las regiones del gen nuclear ITS1 e ITS2 para definir límites y relaciones entre diferentes linajes relacionados con *Agave*, obteniendo mejor resolución filogenética. En el 2000, Eguiarte y colaboradores, utilizaron secuencias ITS de Bogler y Simpson (1996) y estimaron el origen del género *Agave* en aproximadamente 8 millones de años, usando una reconstrucción de máxima verosimilitud y un reloj molecular estricto. Analizaron un mayor número de secuencias de cloroplastos (*trnL+trnL-trnF*) y *rbcL*, que incluyeron más muestreos para una mejor representación de los diferentes clados relacionados con el agave, en particular de los taxones herbáceos, estimaron también diferencias en las tasas de diversificación (es decir, new species per million years [Myr]) (Good-Avila *et al.*, 2006).

Bogler *et al.* (2006) utilizaron un análisis conjunto de dos regiones de cloroplastos (*ndhF*, *rbcL*) y los datos nucleares de ITS informando nuevamente una resolución pobre dentro de *Agave*, que incluía taxones herbáceos y perennes. Encontraron que *A. striata* y *A. dealbata* E. Morren ex K.Koch. (llamado *A. dasylirioides* Jacobi & C.D. Bouché en ese estudio), fueron los taxones basales del género, como lo detectaron previamente Bogler y Simpson (1996) basándose únicamente en datos de ITS.

Usando marcadores AFLP, para un total de 355 *loci* dominantes y 71 especies de *Agave*, el grupo de June Simpson (Vega *et al.*, 2007) encontró tres grupos bien definidos: uno basal, que incluye cuatro especies de *Manfreda*; otro clado compuesto por *A. striata* y especies relacionadas (*A. dealbata* [= *A. dasylirioides*],

A. petrophila A.García-Mend. & E. Martínez, *A. tenuifolia* Zamudio & E.Sánchez); y el tercer clado formado por el resto de *Agaves* con varios subclados, incluyendo especies de ambos subgéneros, tal como lo definen Berger (1915) y Gentry (1982), es decir, el subgénero *Littaea* (Tagl) Baker, con panículas estrechamente alargadas y tallos cortos, cada uno generalmente con dos flores (Thiede *et al.*, 2019; Thiede, 2020).

En 2016 Heiduk y colaboradores estudiaron la evolución de la fotosíntesis CAM en Agavoideae, y para ello analizaron 272 *loci* nucleares y alrededor de 70 kb del genoma del cloroplasto, todos muestreados por enriquecimiento del locus objetivo. Estudiaron 29 especies de *Agave* y otras especies de Agavoideae, particularmente las del género *Yucca*, totalizando 68 especies (Heiduk *et al.*, 2016).

Scheinvar *et al.* (2017), en un estudio centrado en *A. lechuguilla* y como herramienta para definir la edad de los principales clados dentro de esta especie, estimaron un reloj molecular utilizando datos de cloroplastos, definiendo una fecha de origen de 8.67 millones de años como la corona edad y 9.89 millones de años como edad del tallo para *Agave*.

En 2019, Flores Abreu y colaboradores analizaron 30 taxones de *Agave* sensu lato con un mejor muestreo de los taxones herbáceos (*Manfreda*, *Polianthes*, *Prochnyanthes*, 15 especies en total) y diferentes grupos externos, utilizando tres regiones de cloroplastos (dos fragmentos del gen *matK* y dos regiones no codificantes, *trnL-trnF* y *trnD-trnT* para un total de 2191 pb para 51 taxones, analizados con datos mitocondriales de los murciélagos filostómidos coevolucionados. Para *Agave* sensu lato reportaron una edad de copa de 2.7-8.5 millones de años, y una edad de tallo de 4.62-12.34 millones de años. Dentro de *Agave* sensu lato, para el grupo que contenía *Prochnyanthes*, la mayoría de los miembros de *Polianthes* y *A. striata*, estimaron una edad de copa de 1.4-5 millones de años, mientras que el grupo que contenía la mayoría de los miembros de *Manfreda* tenía una edad de copa de 0.44-3 millones de años. Se concluyó que *Agave* sensu lato tuvo una tasa de especiación creciente hacia el presente, estimada en torno a 1.48 especies/Myr en la actualidad, mientras que su grupo hermano formado por los géneros *Furcraea* Vent. + *Beschorneria* con solo un total

de 32 especies, mostró una tasa de especiación menor y constante de 0.56 especies/Myr (Flores-Abreu *et al.* 2019).

2.8.6 Filogenia actual para el género *Agave*

Jiménez Barrón *et al.* (2020) analizaron la filogenia utilizando regiones de cloroplastos de 46 especies de *Agave* sensu lato (43 *Agave* sensu stricto + 2 *Manfreda*+1 *Polianthes*) y *Furcraea* (2) + *Beschorneria*(1) + *Yucca*(1) especies como grupos externos. Para los análisis ITS nucleares, utilizaron 81 especies de *Agave* sensu lato (72 especies de *Agave* sensu stricto + 4 *Manfreda* + 4 *Polianthes* + 1 *Prochnyanthes*) más *Beschorneria* (4 especies), *Furcraea* (4), *Yucca* (5), *Hesperoyucca* (Engelm) Baker (1), *Chlorogalum* (Lindl.) Kunth (2), *Camassia* Lindl. (2), *Hesperaloe* Engelm. (3) y *Dasylyrion* Zucc. (3) como grupos externos. Utilizaron cuatro regiones de cloroplasto (*matK*, *trnH-psbA*, *rpl32-trnL* y *rps 16*, 3670 pb en total) y 577 pb de datos nucleares ITS. Aunque la longitud total de las secuencias de cloroplastos concatenadas era grande e incluía las regiones más variables del cloroplasto en el clado, la resolución fue pobre, especialmente considerando los valores de arranque, lo que sugiere una evolución complicada y captura de cloroplastos entre diferentes linajes en el clado. Por ejemplo, los análisis mostraron especies de *Manfreda* dispersas en diferentes lugares de la filogenia *Agave* sensu lato. Resultados similares se han encontrado en otros estudios que utilizan ADN de cloroplastos, como Good-Avila *et al.* (2006), Heyduk *et al.* (2016), Scheinvar *et al.* (2017), y Flores-Abreu *et al.* (2019), sustentando que estos resultados de cloroplastos no están relacionados con errores experimentales (Eguiarte *et al.*, 2021).

En cambio, Jiménez *et al.* (2020), mediante el análisis de secuencias ITS nucleares para este conjunto de especies de *Agave* sensu lato, encontraron una resolución mucho mejor. Estimaron la edad del tallo para *Agave* sensu lato en 9 millones de años y la edad de la copa en 6.18 millones de años y encontraron *A. bracteos* S. Watson ex Engelm y *A. ellemetiana* K. Koch como las dos especies más basales de las analizadas en el género *Agave* sensu lato. Estas especies son interesantes, como Gentry (1982) indicó que pueden ser plantas perennes policárpicas y sugirió la morfología de sus flores ("tépalos separados que surgen de un receptáculo

discoide, en lugar de un tubo", y hojas inermes sin una espina terminal que pudiera justificar su remoción de *Agave* a un género separado.

Dentro del grupo de copa de *Agave* sensu stricto, Jiménez *et al.* (2020) encontró un clado que incluía solo a *Agave pelona* como hermana del resto de *Agave*, dispuesta en tres grandes linajes de corona. *Agave pelona* es una especie interesante, reconocida por Gentry (1982) como única entre los agaves relacionados en la combinación de características, incluidas hojas que carecen de dientes marginales (y de ahí el nombre de *pelona de Gentry*, "calva" en español), de margen blanco, espina larga blanca y flores campanuladas de color rojo oscuro con tépalos recurvados extendidos. Además, la presencia de una inflorescencia "espigada" compacta en la especie es interesante dada su posición filogenética.

2.9 LITERATURA CITADA

- Ayala, G. (Ed.). (2021). *Bioinformática estadística: Análisis estadístico de datos ómicos*. Universidad de Valencia. 453p.
- Berger, A. (Ed.). (1915). *Die Agaven*. (Fischer, Ed.). 288p.
- Bogler, D. J., & Simpson, B. B. (1996). Phylogeny of Agavaceae Based on ITS rDNA Sequence Variation. *American Journal of Botany* 83(9), 1225-1235. <https://doi.org/10.2307/2446206>
- Bogler, D. J., Pires, J. C., & Ortega, J.F. (2006). Phylogeny of Agavaceae Based on ndhF, rbcL, and ITS sequences. *Aliso: A journal of Systematic and Floristic Botany*, 22, 313–328. <https://scholarship.claremont.edu/aliso/vol22/iss1/26>
- Blasco, A. (Ed.). (2017). *Bayesian Data Analysis for Animal Scientists: The Basics*. <https://doi.org/10.1007/978-3-319-54274-4>
- CBOL Plant Working Group. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (31), 12794-12797. <https://doi.org/10.1073/pnas.0905845106>
- Chase, M. W., Reveal, J. L., & Fay, M. F. (2009). A Subfamilial Classification for the Expanded Asparagalean Families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. *Botanical Journal of the Linnean Society* 161, 132-136.
- Collard, B.C.Y., Jahufer M.Z.Z., Brouwer J.B. (2005). An introduction to markers,

- quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142, 169-196.
- Dahlgren, R. M. T., Clifford, H. T., & Yeo, F. T. (1985.) The families of the monocotyledons. Structure, evolution, and taxonomy. *Springer Verlag*. 520pp.
- de la Fuente, Santiago. (Ed.). (2011). *Análisis Conglomerados*. <https://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOMERADOS/conglomerados.pdf>
- Domínguez, R. M. S., González, M. L. J., González, C. R., C. Quiñones, V., Díaz, L. D., Mireles, S. J.. & Pérez, E. M. (2008). El cultivo in vitro como herramienta para el aprovechamiento, mejoramiento y conservación de especies del género *Agave*. *Investigación y Ciencia*, 41, 53-62.
- Duvall, M. R., Clegg, M. T. M., W., Chase W., Clark, D., Kress, W. J., Hills, H. G., Learn, G. H. (1993). Phylogenetic hypotheses for the monocotyledons constructed from rbcL sequence data. *Annals of the Missouri Botanical Garden* 80, 607-619.
- Esqueda V., M., Coronado, A. M., Gutiérrez, S. A. H. & Fragoso, G. T. (2011). *Agave angustifolia Haw. Técnicas para el trasplante de vitroplantas a condiciones de agostadero*. Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación. 20p.
- Eguiarte, L. E. (1995). Hutchinson (Agavales) vs. Huber y Dahlgren (Asparagales): análisis moleculares sobre la filogenia y evolución de las familias Agavaceae sensu Hutchinson dentro de las monocotiledóneas. *Boletín de la Sociedad Botánica de México* 56, 45-56.
- Eguiarte, L. E., Jiménez, B. O. A., Aguirre, P. E., Scheinvar, E., Gámez, N., Gasca-P. J., Castellanos, M. G., Moreno, L. A., & Souza, V. (2021). Evolutionary ecology of *Agave*: distribution patterns, phylogeny, and coevolution (an homage to Howard S. Gentry). *Revista American Journal of Botany*, 108(2), 216-235. <https://doi.org/10.1002/ajb2.1609>
- Flores, A. I. N., Trejo, S R. E., Sánchez, R. L. L., Good, S. V., Magallón, S., García, M. A., & Eguiarte, L. E. (2019). Tempo and mode in coevolution of *Agave* sensu lato (Agavoideae, Asparagaceae) and its bat pollinators, Glossophaginae

- (Phyllostomidae). *Molecular Phylogenetics and Evolution*, 133, 176-188.
<https://doi.org/10.1016/j.ympev.2019.01.004>
- Garamszegi, L. Z. (Ed.). (2014). *Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice*. Springer.
 10.1007/978-3-662-43550-2_1
- García, M. A. (2002). Distribution of Agave (Agavaceae) in Mexico. *Cactus and Succulent Journal*, 74, 177-186.
- García, M. A. (2004). Agaváceas. (García, M. A., Ordoñez, M. & Briones, S. M, Eds.). Biodiversidad de Oaxaca. 165p.
- García, M. A. 2007). *Los agaves de México*. Instituto de Biología, Universidad Nacional Autónoma de México, 1, 14-23.
- García, M. A. (2011). Flora del Valle Tehuacán-Cuicatlán. Agavaceae. *Instituto de Biología, Universidad Nacional Autónoma de México*, 88, 1-95.
- Gentry, H. S. (1982). *Agaves of Continental North America*. Tucson: University of Arizona. 670p.
- Good, A. S. V., Souza, V., Gaut, B. S., & Eguiarte, L. E. (2006). Timing and rate of speciation in Agave (Agavaceae). *Proceedings of the National Academy of Sciences U.S.A.* 103: 9124-9129. <https://doi.org/10.1073/pnas.0603312103>
- Harvey, P. H., and Pagel, M. D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.
- Heyduk, K., McKain, M. R., Lalani, F. & Leebens, M. J.. (2016). Evolution of a CAM anatomy predates the origins of Crassulacean acid metabolism in the Agavoideae (Asparagaceae). *Molecular Phylogenetics and Evolution* 105, 102-113.
- Hall, B. K. (2007). Homoplasy and homology: dichotomy or continuum? *Journal of Human Evolution*, 52(5), 473-479. <https://doi.org/10.1016/j.jhevol.2006.11.010>
- Jiménez, B. O., García, S. R., Magallón, S., García, M. A., Nieto, S. J., Aguirre, P. 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>

- Johnson V, E. (2013). Revised standards for statistical evidence. *Proceedings of the National Academy of Sciences U.S.A.* 110(48), 19313-17.
- Legendre P., Legendre, L. (1998). *Numerical ecology: developments in environmental modelling* (vol. 2). Elsevier Science Publishers. 989pp.
- Mora L. J. L., Reyes A. A., Flores, F. J. L., Peña, V.C. B. & Aguirre, R. J. R. (2011). Variación morfológica y humanización de la sección Salmianae del género Agave. *Agrociencia* 45, 445-467.
- Morales, E. (2000). El método comparativo en ecología vegetal. *Boletín Sociedad Botánica México* 66, 37-52.
- Palacio, F., Apocada ,M., & Crisci, J. (Eds.). (2020). Análisis multivariado para datos biológicos: teoría y su aplicación utilizando el lenguaje R. Fundacion de Historia Natural Felix de Azara. <https://fundacionazara.org.ar/img/libros/analisis-multivariado-para-datos-biologicos/analisis-multivariado-para-datos-biologicos.pdf> 268pp.
- Parker, K. C., Trapnell, D. W., Hamrick, J. L. y 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.
- Reddy, B.V., Reddy, C. V. C.M., Sekhar, A. C., Reddy, P. C. O., Rajasekhar P., & Srinivasulu, K. (2021). Role of molecular based markers methods and their applications in crop improvement. *Plant cell Biothechnology and Molecular Biology*, 22(23-24), 38–54.
<https://www.ikprress.org/index.php/PCBMB/article/view/6137>
- Rezende, E., & Garland, T. (2003). Comparaciones Interespecíficas y Métodos Filogenéticos Estadísticos. Fisiología Ecología y Evolutiva. Chile: Universidad Católica de Chile. 70-98.
- RStudio Team (2020). RStudio: Integrated Development for R. Boston, MA: RStudio, PBC.
- Rzedowski, J. (1993). *Diversity and origins of the phanerogamic flora of México*. (Ramamoorthy, T.p., Bye, R., Lot, A. & Fa. J. Eds.). Oxford University Press. 129-144.

- Scheinvar, E., Gámez, N., Castellanos, M. G., Aguirre, P. E., & Eguiarte, L. E. (2017). Neogene and Pleistocene history of *Agave lechuguilla* in the Chihuahuan Desert. *Journal of Biogeography* 44, 322-334. <https://doi.org/10.1111/jbi.12851>
- Selvaraj, D., Sarma, R. K., & Sathishkumar, R. (2008). Phylogenetic analysis of chloroplast matK gene from Zingiberaceae for plant DNA barcoding. *Bioinformation* 3(1), 24-27. <https://doi.org/10.6026/97320630003024>
- SIACON. (2020). Base de datos del servicio de información agroalimentaria y pesquero (SIAP). Mexico. <https://www.gob.mx/siap/documentos/siacon-ng-161430>
- Sorensen, D. & Gianola, D. (2002). Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics: *Springer Science and Business Media*; 60, 539. <http://dx.doi.org/10.1007/b98952>
- The Angiosperm Phylogeny Group. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, 161, 105-121. <https://doi.org/10.1111/j.1095-8339.2009.00996.x>
- The Angiosperm Phylogeny Group. (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, 1-20. <https://doi.org/10.1111/boj.12385>
- Thiede, J., Smith, G. F., & Eggli, U. (2019). Infrageneric classification of *Agave* L. (Asparagaceae: Agavoideae/Agavaceae): a nomenclatural assessment and updated classification at the rank of section, with new combinations. *Bradleya* 37, 240–264. <https://doi.org/10.25223/brad.n37.2019.a22>
- Thiede, J., & Eggli, U. (2020). *Agave* (Eggli, U., & Nyffeler, R. Eds.). Monocotyledons. Illustrated handbook of succulent plants. *Springer*. 9-19. https://doi.org/10.1007/978-3-662-56486-8_111
- Vega, K. C., Díaz, Q. C. E., Nava, C. A., García, M. A., & Simpson, J. (2007). Análisis AFLP del género *Agave* refleja la clasificación taxonómica basada en caracteres morfológicos y otros métodos moleculares. (Colunga, G.M., Larqué,

S., Eguiarte, L.E., & Zizumbo, V. Eds.). En ancestral hay futuro: del tequila, los mezcales y otros agaves, *Centro de Investigaciones Científicas de Yucatán, Mérida*. 23-39.

Zamora , A. J. G., Manríquez, M. N. L., Ortiz, Y. C. E., & Ortega, L. A. M. (ed.). (2013). Uso de técnicas moleculares como herramienta para conservar la diversidad biológica, en *Biología Molecular Aplicada a la Producción Animal y la Conservación de Especies Silvestres*. H. A. López. Universidad Nacional de Colombia. 313-386.

2 CAPITULO I

ANÁLISIS FENOTÍPICO DE AGAVES MEZCALEROS DE LA REGIÓN VALLES CENTRALES DE OAXACA

1. RESUMEN

Los agaves mezcaleros, son considerados actualmente de alto valor económico en el estado de Oaxaca. Debido a la demanda de producción de mezcal, por lo tanto, la presente investigación pretende destacar la importancia de la diversidad morfológica existente dentro y entre especies en *Agave* cultivadas. Se evaluaron 275 individuos de las especies *Agave angustifolia*, *Agave karwinskii*, *Agave marmorata*, *Agave rhodacanta*, *Agave potatorum*, *Agave semanniana*, *Agave nussaviorum*, a través de un análisis multivariado, se determinó la variabilidad fenotípica y las relaciones existentes entre las mismas. Por medio de un análisis cluster se realizaron dendrogramas correspondientes al agrupamiento modo Q y modo R, formándose 7 grupos. Por el método de k- medias se formaron 5 grupos. Mediante un Análisis de Componentes Principales (PCA) se determinó que los dos primeros componentes principales explican el 55.99% de la variación total. Para los CP1 y CP2 las variables con mayor contribución fueron las relacionadas a la forma de la hoja (Fh), tamaño de la espina lateral (Tel), número de hojas (Nh), altura de planta (H), uniformidad del tamaño de la espina (UTE) y la forma de la espina terminal (FET).

Palabras clave: Asparagáceae, análisis multivariado, ITS, matK, trnH-PsbA, filogenia molecular.

Tesis de Maestría en Ciencias, Universidad Autónoma Chapingo.

Autor: Vásquez Maya María Patricia

Director: Dr. Juan Porfirio Legaria Solano

PHENOTYPIC ANALYSIS OF MEZCAL AGAVES FROM THE CENTRAL VALLES REGION OF OAXACA

1.1 ABSTRACT

The mezcal agaves are currently considered of high economic value in the state of Oaxaca. Due to the demand for mezcal production, the present research aims to highlight the importance of morphological diversity within and among cultivated *Agave* species. A total of 275 individuals of the species *Agave angustifolia*, *Agave karwinskii*, *Agave marmorata*, *Agave rhodacanta*, *Agave potatorum*, *Agave semanniana*, *Agave nussaviorum* were evaluated by multivariate analysis to determine the phenotypic variability and the existing relationship among them. Through a cluster analysis, dendograms corresponding to the Q-mode and R-mode grouping were formed, forming 7 groups. By using the k-means method, 5 groups were formed. By using means of a Principal Component Analysis (PCA) it was determined that the first two principal components explained 55.99% of the total variation. For CP1 and CP2, the variables with the greatest contribution were those related to leaf shape (Fh), lateral spine size (Tel), number of leaves (Nh), plant height (H), spine size uniformity (UTE) and terminal spine shape (FET).

Keywords: morphological diversity, multivariate analysis, PCA, Q-mode, R-mode.

Thesis, Universidad Autónoma Chapingo.
Author: Vásquez Maya Maria Patricia
Advisor: Dr. Juan Porfirio Legaria Solano

2. INTRODUCCIÓN

En México, la diversidad del agave no está distribuida uniformemente a pesar de la presencia de poblaciones nativas y cultivadas en la mayor parte del país. El agave puede ser muy diverso en algunas áreas, donde pueden coexistir 20 o más especies, como en la región de Tehuacán-Cuicatlán (García-Mendoza, 2002, 2004, 2011).

El estado de Oaxaca tiene el mayor número de especies, albergando 43 especies de *Agave*, lo cual no es sorprendente dado que este estado incluye el Valle de Tehuacán-Cuicatlán (que tiene el mayor número de especies de *Agave*) y es conocido por su alta biodiversidad vegetal. Otros estados con una gran cantidad de especies de agave son el estado de Jalisco, en el centro oeste de México, con 38 especies, y los estados del centro este de Hidalgo, Veracruz y San Luis Potosí, que albergan 32 especies cada uno (Villaseñor, 2016).

Mediante el análisis multivariado se pueden estudiar integralmente las mediciones de atributos o características de los elementos de una población. Ofrece la posibilidad de cuantificar la intensidad de la influencia o asociación de variables independientes en el modelo matemático concreto, o bien, tomadas como punto de partida para investigar determinado fenómeno sobre una variable dependiente matemáticamente (Vazquez & de Bernard, 1988). Actualmente las agaves dentro del género *Agave* son un producto importante económicamente debido principalmente a sus características industriales, por lo que se encuentran en alta demanda. Bajo estas circunstancias se vuelve de vital importancia generar información científica en la cual apoyarse para su aprovechamiento y conservación. El análisis de datos, junto con estudios de caracterización morfológica y molecular, son herramientas que ayudan a gestionar eficientemente las accesiones dentro de un banco de germoplasma (Barrera-Guzmán *et al.*, 2020).

El análisis de la variación genética entre y dentro de las poblaciones es de interés fundamental para su conservación y para la realización de programas de mejoramiento genético, por que proporciona estimadores de la extensión de la variación genética disponibles, contribuye al monitoreo de germoplasma y permite predecir ganancias genéticas potenciales (Moreno-González & Cubero, 1993). El

proyecto contempla desarrollar información que dará bases para el conocimiento del género *Agave*, en su diversidad genética.

3. Objetivos

- Evaluar la variabilidad morfológica de diferentes poblaciones de *Agave* dentro de la región de Valles Centrales de Oaxaca, utilizando variables cuantitativas y cualitativas.
- Hacer un análisis robusto utilizando un análisis multivariado, para conocer que variables son más significativas para diferenciar a las especies y cuales presentan mayor variabilidad fenotípica.

4. Hipótesis

Ho: no se manifiestan diferencias morfológicas entre especies de *Agave* sp. ni existe variabilidad dentro y entre las poblaciones de estas especies de *Agave* cultivadas.

Ha: es posible detectar diferencias morfológicas entre especies de *Agave* sp. y existe variabilidad dentro y entre las poblaciones de estas especies de *Agave* cultivadas.

5. MATERIALES Y MÉTODOS

5.1 Localización del área de estudio

El área de estudio se localiza en zonas productoras del cultivo de *Agave* de Valles Centrales del estado de Oaxaca. Se consideraron los distritos característicos de producción del cultivo y producción de Mezcal; entre los que se encuentran Ejutla, Ocotlán, Tlacolula y Zimatlán. Se realizó muestreos en diferentes parcelas de *Agave* de estos municipios (Cuadro1).

5.2 Materiales

Se llevó a cabo un análisis fenotípico de poblaciones correspondientes a cinco especies de *Agave*. Las especies siguientes conformaron la unidad experimental (UE): *Agave angustifolia* (Espadin, nombre común), *Agave karwinskii* (Tobasiche, Kuish, Marteño, Barril, nombres comunes), *Agave potatorum*, *Agave semanianna*, *Agave nussaviorum* (Tobala, nombre común), *Agave rhodacantha* (Mexicano, nombre común) y *Agave marmorata* (Tepeztate, nombre común) (Cuadro 1). Para la toma de datos se utilizó flexómetros, cintas métricas, reglas y un vernier. Para

determinar las coordenadas y altura de los sitios muestreados se hizo uso de un GPS.

5.3 Tamaño de la muestra

Se evaluaron cinco especies de *Agave* y cinco poblaciones por especie, con diez individuos por población más un testigo. Los individuos estudiados se eligieron según las características siguientes: 1) estar en etapa de desarrollo de maduración, es decir, con escapo floral o próximos a emitirlos; y 2) tener una edad no menor de 5 años; dado que una vez alcanzada su maduración se expresan de manera más visibles sus características morfológicas.

Cuadro 1. Identificación y descripción de las poblaciones muestreadas.

P	Nombre de la especie	Condición y nombre común	Número individuos/población	Distrito	Localización de la parcela	Coordenadas geográficas	Alt.
1	<i>Agave angustifolia</i>	Cultivado, espadín	11 individuos	Ejutla, Oax.	San Agustín Amatengo	16°29'59"N, 96°46'36"O	1380
2	<i>Agave angustifolia</i>	Cultivado, espadín	11 individuos	Ejutla, Oax.	Agua del espino	16°35'25"N, 96°48'31"O	1422
3	<i>Agave angustifolia</i>	Cultivado, espadín	11 individuos	Ocotlán, Oax.	santa Catarina minas	16°47'23"N, 96°36'16"O	1630
4	<i>Agave angustifolia</i>	Cultivado, espadín	11 individuos	Ocotlán, Oax.	San Dionisio	16°44'55"N, 96°40'11"O	1530
5	<i>Agave angustifolia</i>	Cultivado, espadín	11 individuos	Ocotlán, Oax.	San Dionisio	16°29'59"N, 96°46'36"O	1500
6	<i>Agave karwinskii</i>	Cultivado, kuish	11 individuos	Ejutla, Oax.	San Agustín Amatengo	16°30'59"N, 96°46'32"O	1440
7	<i>Agave karwinskii</i>	Cultivado, barril	11 individuos	Ejutla, Oax.	Agua del espino	16°35'45"N, 96°48'45"O	1457
8	<i>Agave karwinskii</i>	Cultivado, marteño	11 individuos	Ejutla, Oax.	La compañía	16°35'40"N, 96°48'49"	1380
9	<i>Agave karwinskii</i>	Cultivado, tobasiche	11 individuos	Ocotlán, Oax.	santa Catarina minas	16°47'5"4.7", 96°36'55"	1560
10	<i>Agave karwinskii</i>	Cultivado, barril	11 individuos	Ocotlán, Oax.	santa Catarina minas	16°47'5"4.7", 96°36'55"	1562
11	<i>Agave semanniana</i>	Cultivado, tobala	11 individuos	Ocotlán, Oax.	San Dionisio	16°44'57", 96°40'5"	1530
12	<i>Agave potatorum</i>	Cultivado, tobala	11 individuos	Ocotlán, Oax.	santa catarina minas	16°46'41"N, 96°36'58"O	1630
13	<i>Agave nussaviorum</i>	Cultivado, tobala	11 individuos	Tlacolula, Oax.	San Dionisio Ocotepec	16°48' N, 96°24' O	1670
14	<i>Agave semanniana</i>	Cultivado, tobala	11 individuos	Tlacolula, Oax.	San Baltazar	16°46'N, 96°29'O	1548
15	<i>Agave potatorum</i>	Cultivado, tobala	11 individuos	Tlacolula, Oax.	San Dionisio	16°48' N, 96°24' O	1670

16	<i>Agave rhodacantha</i>	Cultivado, mexicano	11 individuos	Ocotlán, Oax.	santa catarina minas	16°47'23"N, 96°36'16"O	1630
17	<i>Agave rhodacantha</i>	Cultivado, mexicano	11 individuos	Ocotlán, Oax.	santa catarina minas	16°47'23"N, 96°36'16"O	1630
18	<i>Agave rhodacantha</i>	Cultivado, mexicano	11 individuos	Tlacolula, Oax.	San Baltazar Guelavila	16°46'N 96°29'O	1548
19	<i>Agave rhodacantha</i>	Cultivado, mexicano	11 individuos	Tlacolula, Oax.	San Dionisio	16°48' N 96°24' O	1670
20	<i>Agave rhodacantha</i>	Cultivado, mexicano	11 individuos	Tlacolula, Oax.	San Baltazar Guelavila	16°46'N 96°29'O	1548
21	<i>Agave marmorata</i>	Cultivado, tepeztate	11 individuos	Ocotlán, Oax.	Santiago Apóstol	16°47'58"N 96°42'56"O	1440
22	<i>Agave marmorata</i>	Cultivado, tepeztate	11 individuos	Ocotlán, Oax.	Santa Ana Zegache	16°48'42"N 96°43'48"O	1480
23	<i>Agave marmorata</i>	Cultivado, tepeztate	11 individuos	Tlacolula, Oax.	San Baltazar Guelavila	16°46'N 96°29'O	1548
24	<i>Agave marmorata</i>	Cultivado, tepeztate	11 individuos	Tlacolula, Oax.	Tlacolula	16°43'N 96°33'O	1600
25	<i>Agave marmorata</i>	Cultivado, tepeztate	11 individuos	Tlacolula, Oax.	Matatlán	16°51'48"N 96°22'58"O	1718

P: población; Alt.: altitud (m.s.n.m)

5.4 Diseño experimental

Se utilizó un diseño experimental completamente al azar con 25 tratamientos (poblaciones) y 11 repeticiones (individuos). La unidad experimental consistió en un individuo de cada población durante su etapa madura.

El modelo estadístico corresponde a:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

Donde:

i= 1,..., t; t=25 tratamientos.

j= 1..., n; n= 11 repeticiones por tratamiento.

Y_{ij} = respuesta observada con el i-ésimo tratamiento en la j-ésima repetición

μ = media general

τ_i = efecto de i-esimo tratamiento

ε_{ij} = error experimental

5.5 Variables

En cuanto a las variables evaluadas e integración de la matriz básica de datos (MBD), los datos utilizados en este estudio consistieron en 20 variables morfológicas

(Cuadro 2, Figura 1) registradas para las 275 UE, obtenidas del manual “Guía técnica para la descripción varietal de Agave” editado por el Servicio Nacional de Inspección y Certificación de Semillas (SNICS).

El conjunto de caracteres se comprende de: 10 variables cuantitativas continuas correspondientes a: H: altura de la planta en centímetros (cm) (midiendo de la base del tallo hasta la punta de la espina terminal de la hoja), DIAM: diámetro de la roseta (cm), Nh: número de hojas, Ah: ancho de la hoja (cm), Lh: longitud de la hoja (cm), Tel: tamaño de la espina lateral (mm o cm), Nel: número de espinas laterales en 10 cm, Del: distancia entre las espinas laterales en 10 cm, Let: longitud de la espina terminal (mm o cm), CIF: ciclo a inicio de la floración.

10 variables cualitativas codificadas de la siguiente forma: FET: forma de la espina terminal (recta: 0, curvada: 2, filiforme: 3, polifurcada:4); ICh: intensidad del color de la hoja (débil:3, medio: 5, fuerte: 7); VT: visibilidad del tallo (ausente: 0, presente: 1); Fh: forma de la hoja (lineal: 1, espatulada: 2, deltoide: 3, lanceolada:4, oblonga: 5, ovada:6); Txh: textura de la hoja (lisa: 1, rugosa: 2); ELh: espinas laterales de la hoja (ausente: 0, presente: 1); FEL: forma de las espinas laterales (recta: 1, curva: 2, ganchuda: 3, filifera:4); PEL: perfil de la espina lateral (monofurcada: 1, bifurcada: 2, trifurcada: 3, polifurcada: 4); UTE: uniformidad del tamaño de la espinas laterales (homogénea: 1, heterogénea: 9); HC: hábito de crecimiento (acaulescente:1, caulescente: 2).

En la Figura 1 se muestran plantas de agave donde se pueden observar algunos de estos caracteres diferenciales y el inicio del ciclo de floración.

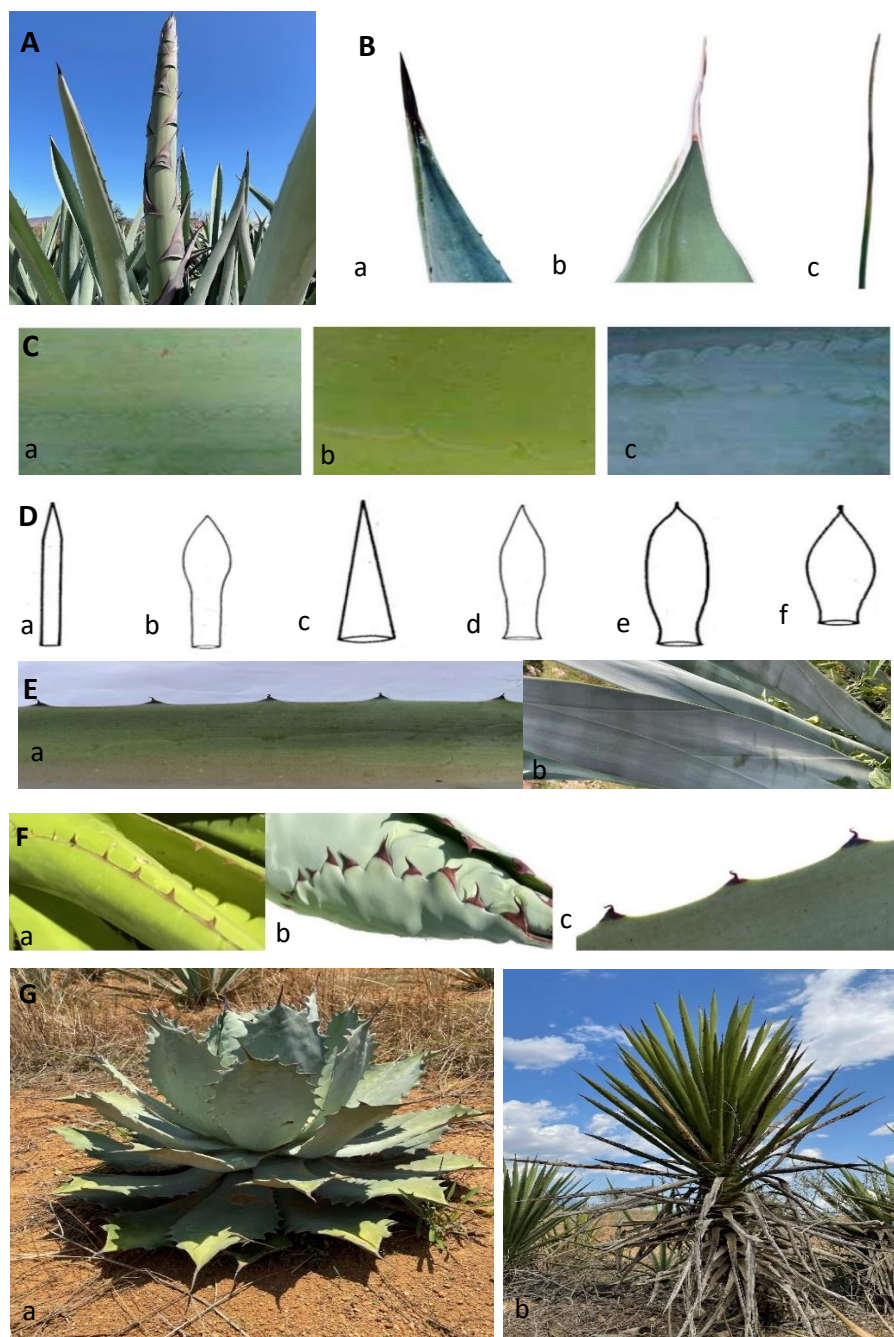


Figura 1. Características diferenciales de plantas de agave mezcalero e inicio del ciclo de la floración. A. Planta en inicio de floración. B. Forma de la espina terminal: recta (a), curvada (b), filiforme (c). C. Intensidad del color de la hoja: débil (a), medio (b), fuerte (c). D. Forma de la hoja: lineal (a), espatulada (b), deltoide (c), lanceolada (d), oblonga (e), ovada (f). E. Espinas: presentes (a) o ausentes (b). F. Forma de las espinas laterales: recta (a), curva (b), ganchuda (c). G. Hábito de crecimiento: acaulescente (a), caulescente (b).

5.6 Análisis estadístico

Se realizó un análisis descriptivo de 25 poblaciones cultivadas de agave, en cada población hubo un testigo (una planta madura de entre 5-7 años con mayor vigor que compita con los individuos dentro la población). Se utilizaron 20 descriptores morfológicos y agronómicos propuestos en la Guía técnica para la descripción varietal de agave (*Agave spp.*), del SNICS (Servicio Nacional de Inspección y Certificación de Semillas), todo representado por 20 variables: 10 cuantitativas y 10 cualitativas (Cuadro 2).

Cuadro 2. Variables cuantitativas (QN) y cualitativas (QL) evaluadas para el análisis fenotípico de los agaves.

CARACTERÍSTICAS QN	Código
Planta: altura	H
Planta: diámetro de la roseta	DIAM
Planta: número de hojas	Nh
Hoja: ancho	Ah
Hoja: longitud	Lh
Hoja: tamaño de la espina lateral en 10 cm	Tel
Hoja: número de espinas laterales en 10 cm	Nel
Hoja: distancia entre las espinas laterales	Del
Hoja: longitud de la espina terminal	Let
Planta: ciclo de inicio de floración	CIF
CARACTERÍSTICAS QL	
Hoja: forma de la espina terminal	FET
Hoja: intensidad del color	Ich
Tallo: visibilidad	VT
Hoja: forma	Fh
Hoja: textura	Txh
Hoja: espinas laterales (P o A)	ELh
hoja: forma de espinas laterales	FEL
Hoja: perfil de la espina lateral	PEL
Hoja: uniformidad en el tamaño de las espinas laterales	UTE
Planta: hábito de crecimiento	HC

El análisis estadístico se realizó a través de una técnica descriptiva de análisis estadístico multivariada, esta técnica estadística es de gran utilidad puesto que la interpretación del resultado puede hacerse de manera sencilla a través de gráficas. Con este procedimiento se puede evidenciar de manera perceptible el grado de relación de las especies presentes.

Se estimó el coeficiente de Pearson y se elaboró una matriz de correlaciones mediante el paquete estadístico Rstudio. Con ello se hizo un análisis de componentes principales para las variables cuantitativas y apreciar los agrupamientos de poblaciones.

Para los métodos que se utilizaran de aquí en adelante se eliminó una variable Elh: espinas laterales de la hoja (Presente o ausente), por ser una variable no binaria, como todas las especies tienen presencia de espinas en la hoja los datos se vuelven iguales y no dan ninguna información.

5.6.1 Agrupamiento jerárquico sobre componentes principales

El primer paso consistió en aplicar un método de ordenación, y se analizó la variación explicada por los ejes principales.

Se aplicó el método de ligamiento promedio sobre los scores resultantes del CA, se especificó en 5 el número de agrupamientos, para posteriormente modificar el objeto, que define el número de grupos a graficar, para visualizar el dendrograma correspondiente. Se graficó e identificó a las UE sobre los ejes principales del CA, según el grupo al que pertenecen.

Finalmente se graficó un biplot simétrico para identificar asociaciones entre UE y variables

5.6.2 Mapa de calor

Con la finalidad de analizar la interacción entre variables y UE se creó un mapa de calor sobre la base de la MBD estandarizada de *Agave* de los 5 grupos.

5.6.3 K- medias

Utilizando la MBD estandarizada (modo Q). Como configuración inicial se representó las K-medias, suponiendo 5 grupos, esto con el objetivo de que el programa asignara cada una de las especies a alguno de los 5 grupos, buscando la configuración con la menor variación intra-grupo, una vez realizado este procedimiento con la función k-medias se identifica puntualmente el grupo en el que se encuentra cada especie.

Se evaluó *a posteriori* el número óptimo de grupos dando como resultado la Figura 2. en la cual es posible apreciar que al establecer 5 grupos se logra la mayor

proporción del descenso de la variación, alcanzando un valor óptimo de esta misma, por lo que esta cantidad queda constituida para la elaboración final de las k-medias.

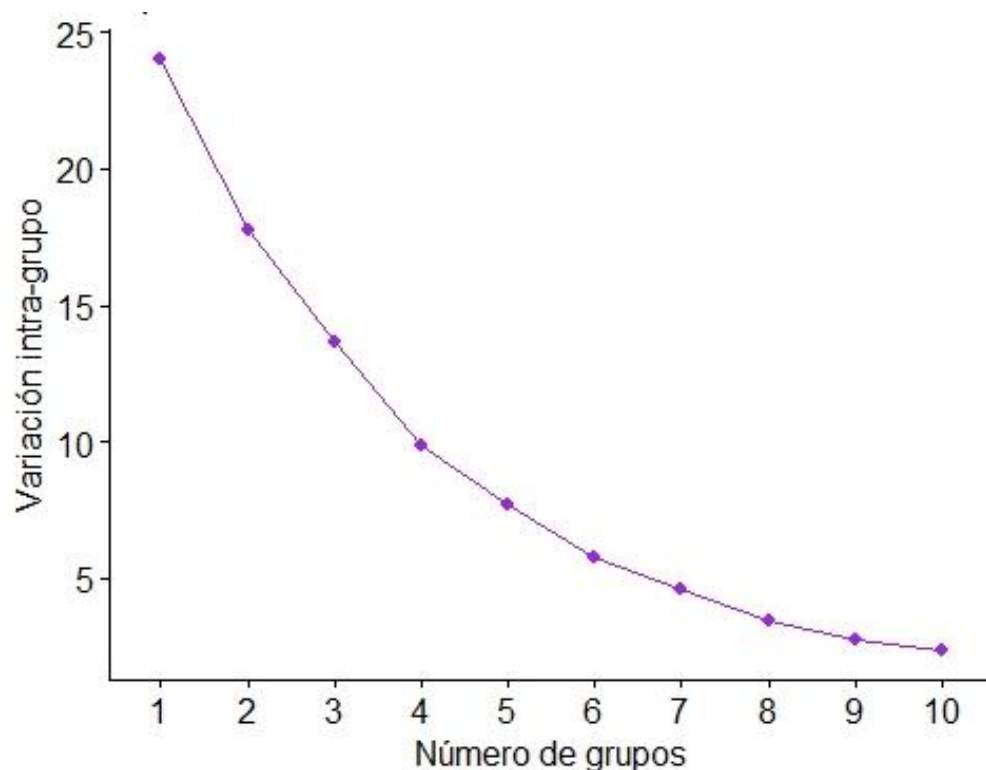


Figura 2. Número óptimo de grupos vs variación intra-grupo sobre el método de K-medias aplicadas a la MBD de Agave.

5.6.4 Análisis de componentes principales

También en este estudio se llevó a cabo un análisis de componentes principales (ACP), bajo el modelo de componentes principales se genera una gráfica en dos ejes con los componentes con mayor porcentaje de varianza.

6. RESULTADOS Y DISCUSIÓN

6.1 Análisis multivariado

6.1.1 Análisis clúster

6.1.1.1 Agrupamiento jerárquico

El dendrograma correspondiente al agrupamiento jerárquico (modo Q) se muestra en la Figura 3, este representa una estructura escalonada lo cual sugiere que los individuos analizados presentan diferencias marcadas entre cada una de las UE.

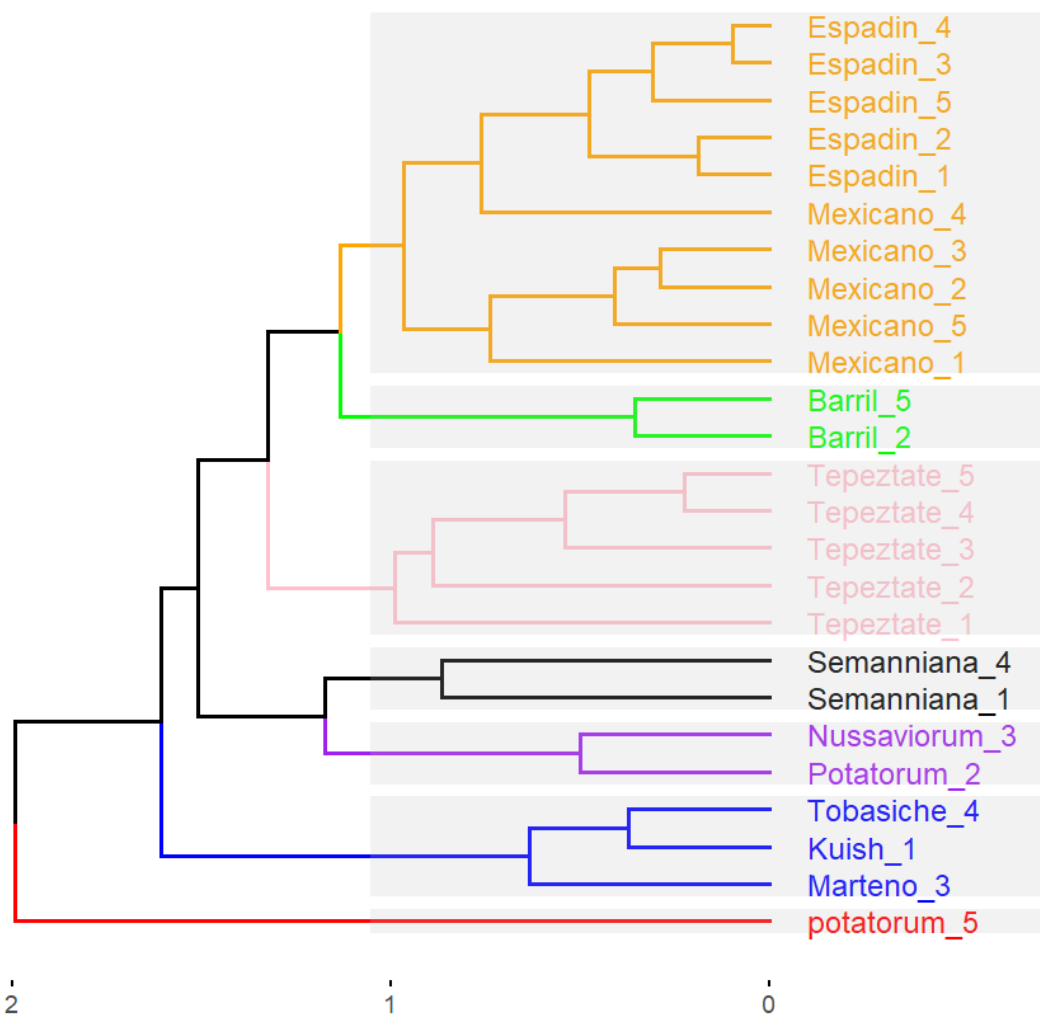


Figura 3. Dendrograma correspondiente al agrupamiento jerárquico (modo Q) basado en el ligamiento promedio generado a partir de la MBD estandarizada de las especies del género *Agave*.

El dendrograma resultante del agrupamiento jerárquico (modo Q) basado en el ligamiento promedio generado a partir de la MBD estandarizada de las especies del género *Agave* se constituyó de 7 grupos, lo que sucede a un nivel de distancia aproximado de 2. El primer grupo, presente en el área inferior del dendrograma se integró exclusivamente de la especie *Agave potatorum*. El segundo grupo correspondió al grupo Karwinskii (*nombre común: tobasiche, kuish, marteño, barril, barril*), representando a tres tipos de *Agave karwinskii* (*tobasiche, kuish, marteño*), mismos que presentan semejanzas a la distancia marcada. El tercer grupo se compone de dos especies relacionadas correspondientes al grupo potatorum

(*Agave potatorum* y *Agave nussaviorum*) con una distancia de 1.5. El cuarto grupo se estructuró de *Agave semanniana*, y está muy relacionado con el grupo tres de *potatorum* a una distancia de 1.4. La quinta agrupación correspondió a Tepeztate (*Agave marmorata*). El sexto grupo corresponde al grupo Karwinskii, pero se integró exclusivamente de tipos Barril (*A. karwinskii*) con una distancia de 1.2, separándose de *tobasiche*, *kuish* y *martelño*. El séptimo grupo presente en el área superior del dendrograma, se integró de 10 poblaciones que taxonómicamente son parecidas, correspondientes a las especies *Agave angustifolia* y *Agave rhodacantha*, a una distancia de 0.90. En este grupo se pueden distinguir dos subgrupos, uno integrado principalmente por los tipos Mexicano (*Agave rhodacantha*) y otro por los Espadín (*Agave angustifolia*) a una distancia cercana a 0.5. Espadin_3 y Espadin_4 son las especies más cercanas fenotípicamente, a una distancia más alejada del resto de aproximadamente 0.2. la distancia se representa en escala Horizontal.

Dendrograma de relaciones entre variables (modo R).

En el dendrograma de la Figura 4 se aprecian las relaciones entre las variables evaluadas, se determinaron *a posteriori* 9 grupos, estos se concretan a una distancia de 0.75. Considerando los grupos de forma descendente: el primer grupo se encuentra constituido exclusivamente por la variable CIF: ciclo de inicio de la floración, por lo que esta variable bajo las condiciones de distancia seleccionada *a posteriori* no se encuentra relacionada con alguna otra variable.

El segundo grupo se constituyó por un clado separado a una distancia de 0.7, donde la variable Tel; tamaño de la espina lateral se encuentra relacionado con Fh: forma de la hoja, por lo cual en este grupo se puede inferir que la forma de la hoja es muy cercana con el patrón del tamaño de la espina lateral.

El tercer grupo se constituyó por dos clados separados a una distancia de 0.61, en el primero se tiene a Txh; textura de la hoja, y en el segundo se encuentran las variables Ah e Ich; que corresponden al patrón ancho de la hoja e intensidad del color, por lo que se puede inferir que la textura de la hoja es muy cercana con el patrón del ancho y color de la hoja.

El cuarto grupo se encuentra constituido por: Nel y PEL por lo que en este grupo queda remarcada la relación existente entre las variables del patrón de número de espinas laterales y el perfil de la espina lateral.

El quinto grupo se constituyó por dos clados separados a una distancia de 0.6, en el primero se encuentra let; longitud de la espina lateral, y en el segundo se encuentran las variables FET y UTE; que corresponden al patrón forma de la espina lateral y uniformidad en el tamaño de las espinas laterales, por lo cual en este grupo se puede inferir que la longitud de las espinas laterales está muy relacionada con el patrón de la forma y uniformidad de las espinas laterales.

El sexto grupo se constituyó por dos clados separado por una distancia de 0.76, en el primero se tiene a Nh: número de hojas, en el segundo se encuentran las variables VT y HC; que corresponden al patrón visibilidad del tallo y hábito de crecimiento, esto se puede explicar en base a que la visibilidad del tallo está relacionada con el hábito de crecimiento, si es caulescente o acaulescente.

El séptimo grupo se encuentra constituido exclusivamente por la variable FEL: forma de la espina lateral, a una distancia de 1.25, por lo que esta variable bajo las condiciones de distancia seleccionada *a posteriori* no se encuentra relacionada con alguna otra variable.

El octavo grupo se encuentra constituido exclusivamente por la variable Del: distancia entre las espinas laterales, por lo que esta variable bajo las condiciones de distancia seleccionada *a posteriori* no se encuentra relacionada con alguna otra variable.

El noveno grupo se constituyó por dos clados separados a una distancia de 0.38, en el primero H; altura de la planta, en el segundo se encuentran las variables DIAM y lh; que corresponden al patrón diámetro de la planta y longitud de la hoja, por lo cual en este grupo se puede inferir que tanto la altura, diámetro y longitud de la hoja están relacionadas y al presentar una de las menores distancias se asumen como las variables más cercanas.

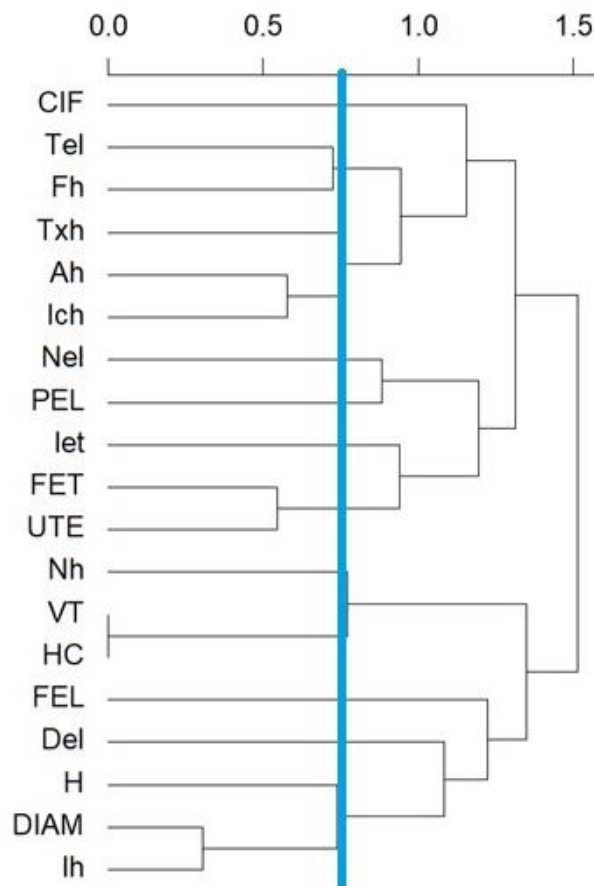


Figura 4. Dendrograma de relaciones entre variables evaluadas correspondiente al agrupamiento jerárquico (modo R) basado en el ligamiento promedio generado a partir de la MBD estandarizada de las especies del género *Agave*. La distancia se representa en la escala horizontal.

Mapa de calor

En la Figura 5 se muestra un mapa de calor donde se observan las relaciones existentes entre las especies pertenecientes al género *Agave* y las variables cualitativas (QL) y cuantitativas (QN) evaluadas durante la caracterización morfológica.

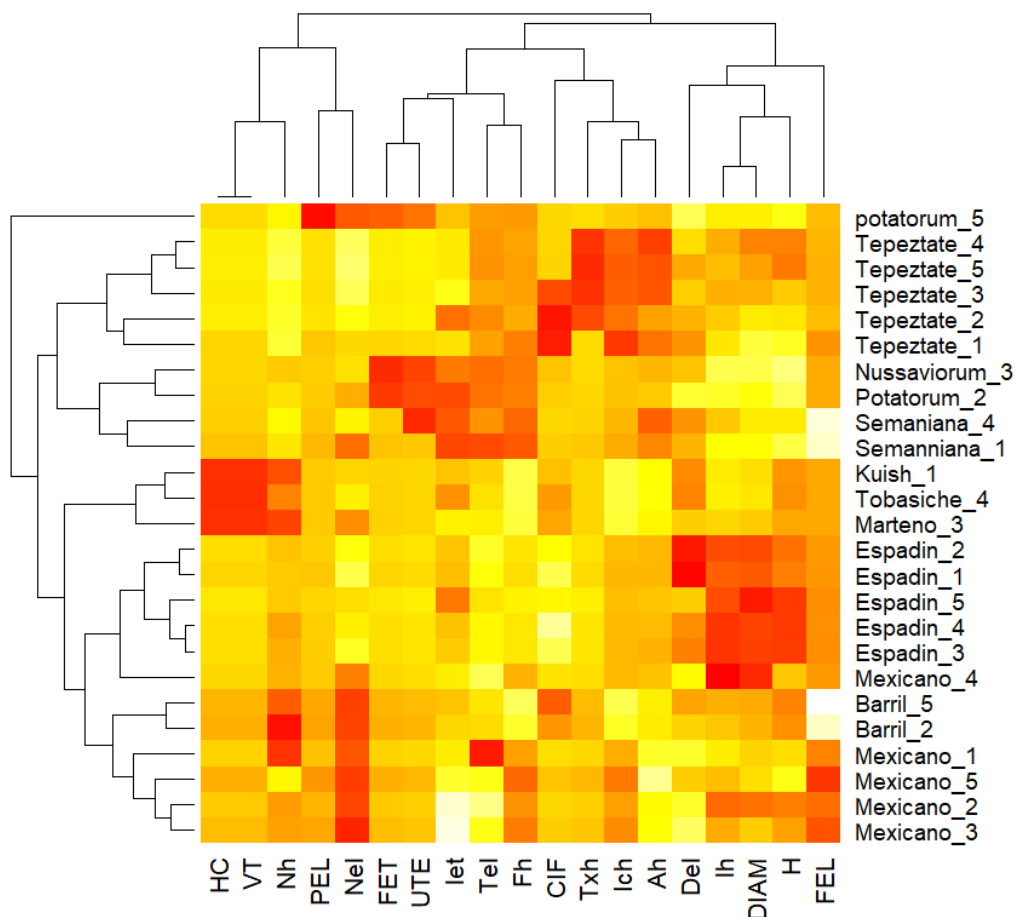


Figura 5. Mapa de calor que muestra las relaciones existentes entre las especies pertenecientes al género *Agave* sp. Con las variables QL y QN evaluadas de la caracterización morfológica en campo.

De este mapa de calor se puede derivar lo siguiente: para el caso de potatorum_5 perteneciente a la especie *Agave potatorum*, las variables más sobresalientes para definirla son: PEL, Nel, FET y UTE. Esta especie tiene el valor más bajo para las variables HC, VT, Nh, Del, lh, DIAM y H. Este clado se separa completamente de las otras poblaciones, donde principalmente PEL (perfil de la espina lateral) es la variable que identifica esta población (Figuras 5 y 7b).

Para Tepeztate_1, Tepeztate_2, Tepeztate_3, Tepeztate_4 y Tepeztate_5 (Figura 5 y 6), dado que están en un mismo clado y pertenecen a la especie *Agave marmorata*, las variables que caracterizan a estas poblaciones y presentan el mayor valor son let, Tel, y Fh, con alto valor representativo estas; y CIF, Txh, Ich y Ah, que

corresponden a ciclo de inicio de la floración, textura de la hoja, intensidad del color de la hoja y ancho de la hoja.



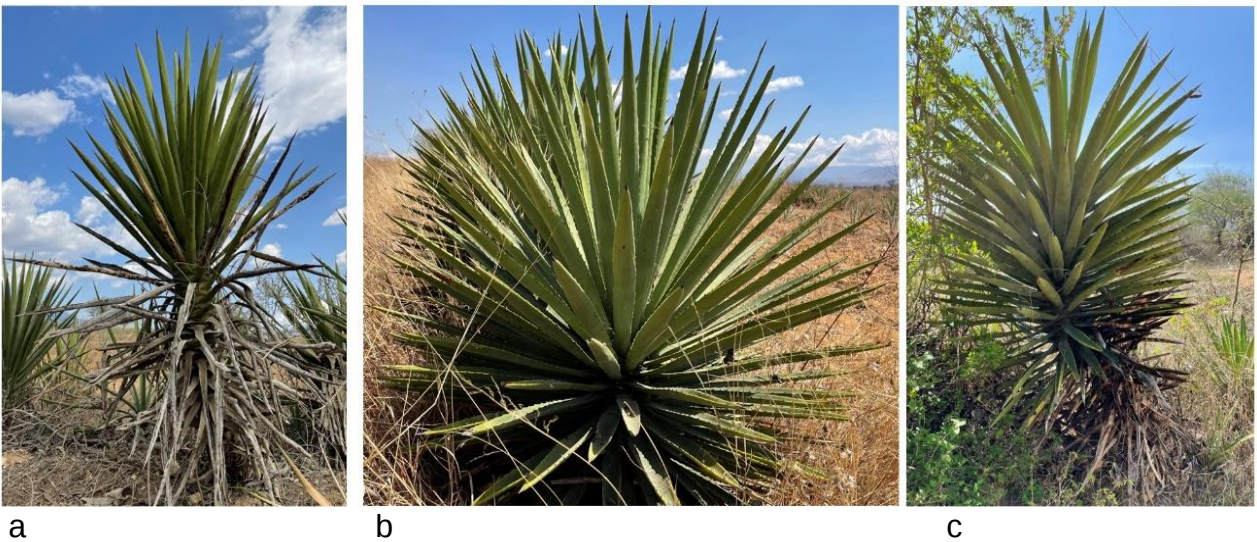
Figura 6. Fenotipo de una planta típica del grupo Tepeztate (*Agave marmorata*).

En Nussaviorum_3 que corresponde a *Agave nussaviorum* (Figuras 5 y 7a), las variables FET y UTE muestran un alto valor, y corresponden a la forma de la espina lateral y uniformidad en el tamaño de las espinas. El siguiente es potatorium_2 población de *Agave potatorium*, definida por las variables FET, UTE y let: longitud de la espina lateral. El siguiente grupo corresponde a Semanniana_4 y Semanniana_1, donde la variable más representativa es UTE: uniformidad en el tamaño de las espinas, además de las variables Let, Tel y Fh, que son longitud de la espina terminal, tamaño de la espina lateral y forma de la hoja. Estas dos últimas poblaciones son de la especie *Agave semanniana* (Figuras 5 y 7c).



a b c
Figura 7. Grupo potatorum, (a) *Agave nussaviorum*, (b) *Agave potatorum*, (c) *Agave semanniana*.

Las poblaciones Kuish_1, Tobasiche_4, Marteño_3, todas correspondientes a *Agave karwinskii* (Figuras 5 y 8), destacan en el alto valor de las variables HC, VT, Nh, que son hábito de crecimiento, visibilidad del tallo y número de hojas.



a b c
Figura 8. Grupo Karwinskii, *Agave karwinskii*, (a) kuish, (b) barril, (c) marteño.

Para el clado, Espadin_2, Espadin_1, Espadin_5, Espadin_4, Espadin_3, todas correspondientes a *Agave angustifolia* (Figuras 5 y 9), las variables más altas son Del, lh, DIAM, H y FEL, que corresponden a la distancia entre las espinas laterales, longitud de la hoja, diámetro de la planta, altura de la planta y forma de la espina lateral.



Figura 9. Fenotipo de una planta correspondiente al grupo Espadín, *Agave angustifolia*.

Mexicano_4 (*Agave rhodacantha*) mostró valores altos para lh y DIAM, longitud de la hoja y diámetro de la planta (Figuras 5 y 10).

Barril_5 y Barril_2, destacaron por su alto valor en las variables Nh y Nel, que son número de hojas y número de espinas laterales, ambas poblaciones son de la especie *Agave karwinskii* (Figuras 5 y 8b).

Para el clado, Mexicano_1, Mexicano_5, Mexicano_2, Mexicano_3, todas correspondientes a *Agave rhodacantha*, las variables más altas son: para Mexicano_1, Nh y Tel, número de hojas y tamaño de la espina lateral; para Mexicano_5, Nel y FEL número de espinas laterales y forma de la espina lateral, y para Mexicano_2 y Mexicano_3, Nel: número de espinas laterales (Figuras 5 y 10).



Figura 10. Fenotipo de una plata correspondiente al grupo Mexicano, *Agave rhodacantha*.

6.1.1.2 No jerárquica (K-medias)

El análisis de K-medias (Figura 11) muestra la integración de 5 grupos en los que el grupo más alejado se encuentra conformado exclusivamente por 5 poblaciones de Tepeztate (*Agave marmorata*), el segundo grupo se integró por 5 poblaciones de Espadin (*Agave angustifolia*) y Mexicano_4 (*Agave rhodacantha*), el tercer grupo presenta las especies (*Agave seemanniana*, *Agave potatorum*, *Agave nussaviorum*), de 5 poblaciones agrupadas en un solo grupo. El cuarto grupo se integró de 4 poblaciones de Mexicano_2, Mexicano_3, Mexicano_5 y Mexicano_1 (*Agave rhodacantha*) y de dos poblaciones de Barril_5 y Barril_2 (*Agave karwinskii*). Y el quinto grupo se formó con poblaciones Tobasiche_4, Kuish_1, y Marteño_3 (*Agave karwinskii*).

Las poblaciones más cercanas al centro del grupo denotan que están fuertemente relacionadas, el elemento más alejado del núcleo dentro del grupo es el menos relacionado.

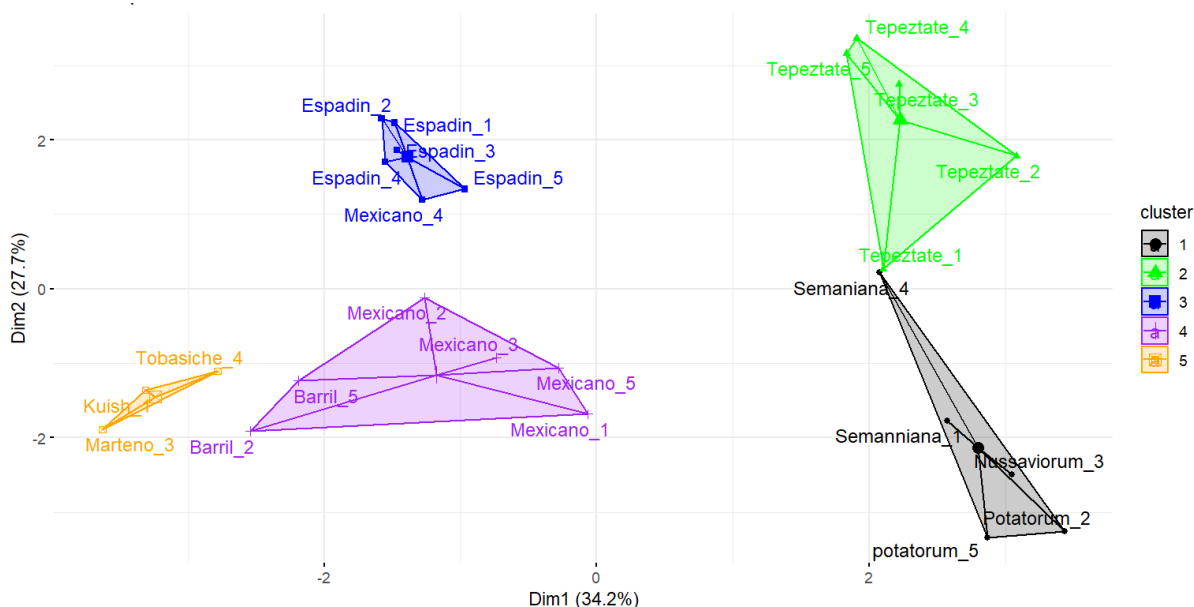


Figura 11. K-medias presentadas por medio de un análisis de componentes principales. La variación explicada por cada componente se indica por medio de los porcentajes de los ejes. Las especies se muestran en cinco grupos junto con sus centroides.

6.1.2 Métodos de ordenación

6.1.2.1 Análisis de componentes principales

En el caso específico del género *Agave*, Mora-López *et al.* (2011) estudiaron la variabilidad morfológica de 62 ejemplares de la sección *Salmaniae*, con el objetivo de identificar la domesticación de estas. Con un análisis de componentes principales (ACP) encontraron las variables morfológicas que explicaban con mayor claridad las diferencias de domesticación (altura y diámetro de la roseta, también la longitud, número y distancia entre dientes). Con esto identificaron cuatro especies, una subespecie y tres variedades.

En el presente trabajo se determinó que los dos primeros componentes principales explican el 55.99% de la variación total. En los Cuadros 3 y 4 se muestran los eigenvalores y eigenvectores, respectivamente.

En la Figura 12 se muestra un gráfico que indica el porcentaje de la varianza explicada del factor 1 y cuanto explican el resto de los componentes. Para los CP1 y CP2 las variables con mayor contribución fueron las relacionadas a la forma de la hoja (Fh), tamaño de la espina lateral (Tel), número de hojas (Nh), altura (H),

uniformidad del tamaño de la espina (UTE) y la forma de la espina terminal (FET), como se muestra en la Figura 13.

Cuadro 3. Eigenvalores y varianza explicada de 10 componentes.

Componente	Eigenvalor	% varianza	% varianza acumulada
CP1	5.96	31.38	31.38
CP2	4.68	24.61	55.99
CP3	2.52	13.21	69.19
CP4	1.54	8.13	77.32
CP5	1.48	7.78	85.11
CP6	0.79	4.18	89.29
CP7	0.54	2.82	92.11
CP8	0.45	1.74	94.48
CP9	0.33	1.63	96.22
CP10	0.31	0.96	97.85

Cuadro 4. Vectores propios para las variables en el ACP

Variable	V1	V2	V3	V4	V5
H	-0.735	0.512	0.027	0.044	0.215
DIAM	-0.454	0.694	-0.379	-0.021	0.293
Nh	-0.750	-0.499	0.129	0.066	0.010
Ah	0.575	0.688	0.206	-0.007	0.045
Lh	-0.388	0.715	-0.419	-0.043	0.222
Tel	0.818	-0.018	0.368	0.040	0.012
Nel	0.146	-0.676	-0.533	0.364	-0.157
Del	-0.396	0.453	0.456	-0.454	0.160
Let	0.544	-0.125	0.451	-0.498	0.201
CIF	0.185	0.124	0.599	0.476	-0.373
FET	0.665	-0.469	-0.041	0.013	0.459
Ich	0.578	0.752	0.044	0.232	-0.019
VT	-0.561	-0.431	0.589	0.175	0.247
Fh	0.925	0.230	-0.062	-0.019	0.047
Txh	0.253	0.662	0.385	0.410	-0.116
FEL	-0.063	0.268	0.063	0.479	0.668
PEL	0.409	-0.335	-0.293	0.378	0.381
UTE	0.707	-0.417	0.016	-0.255	0.354
HC	-0.561	-0.431	0.589	0.175	0.247

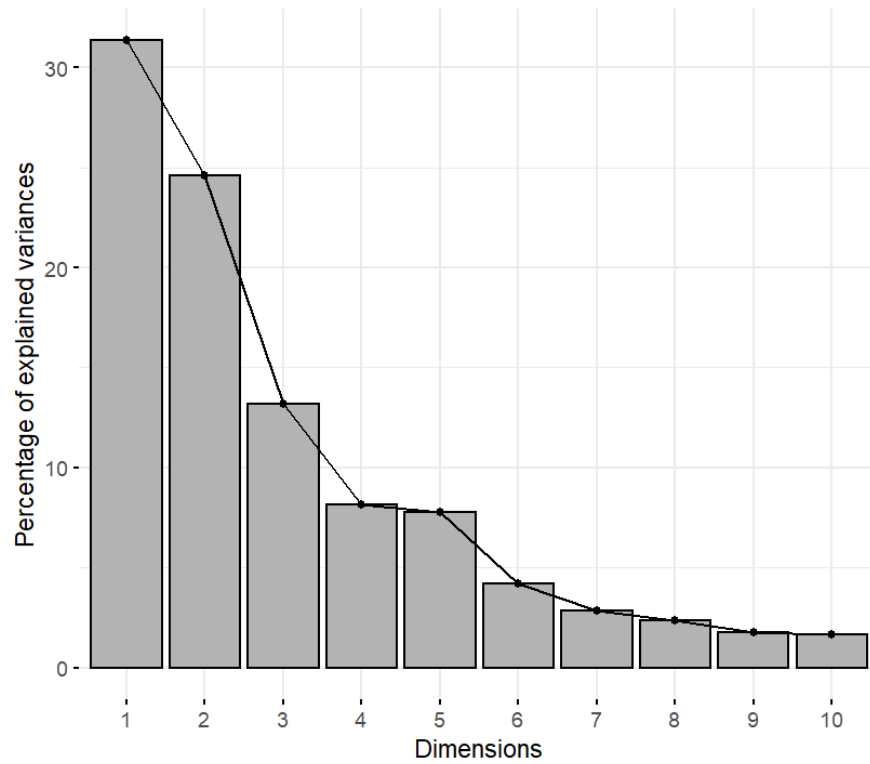


Figura 12. Porcentaje de varianza explicada de los componentes.

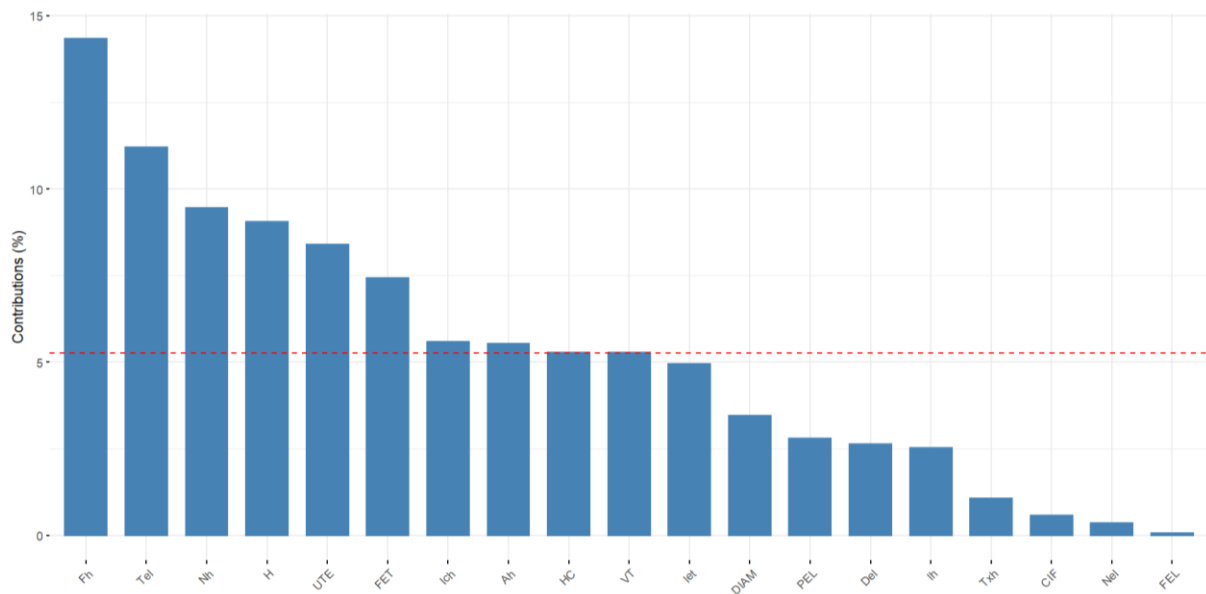


Figura 13. Contribución de cada variable a los componentes principales.

Además, León *et al.* (2013), en un estudio de diversidad determinaron la variación morfológica en cinco especies de agaves (dos variedades de *Agave Karwinskii*, *A. potatorum*, *Agave, marmorata*, *A. americana* y *A. convallis*) en la comunidad de San Miguel Tilquiápan, Ocotlán Oaxaca. En todos los casos las variables con mayor y menor variación fueron el largo de la hoja y el espacio inter-espinas, respectivamente. La Figura 14 muestra la distribución de cada población en los componentes principales de acuerdo con las 19 variables ya mencionadas; las poblaciones potatorum_5 y potatorum_2, pertenecientes a la misma especie son las que más contribuyen a los componentes, siguiéndoles Marteño_3, Nussaviorum_3, Kuisch_1, y Tobasiche_4. La línea horizontal promedio de color rojo indica que por arriba de la misma los valores son significativos.

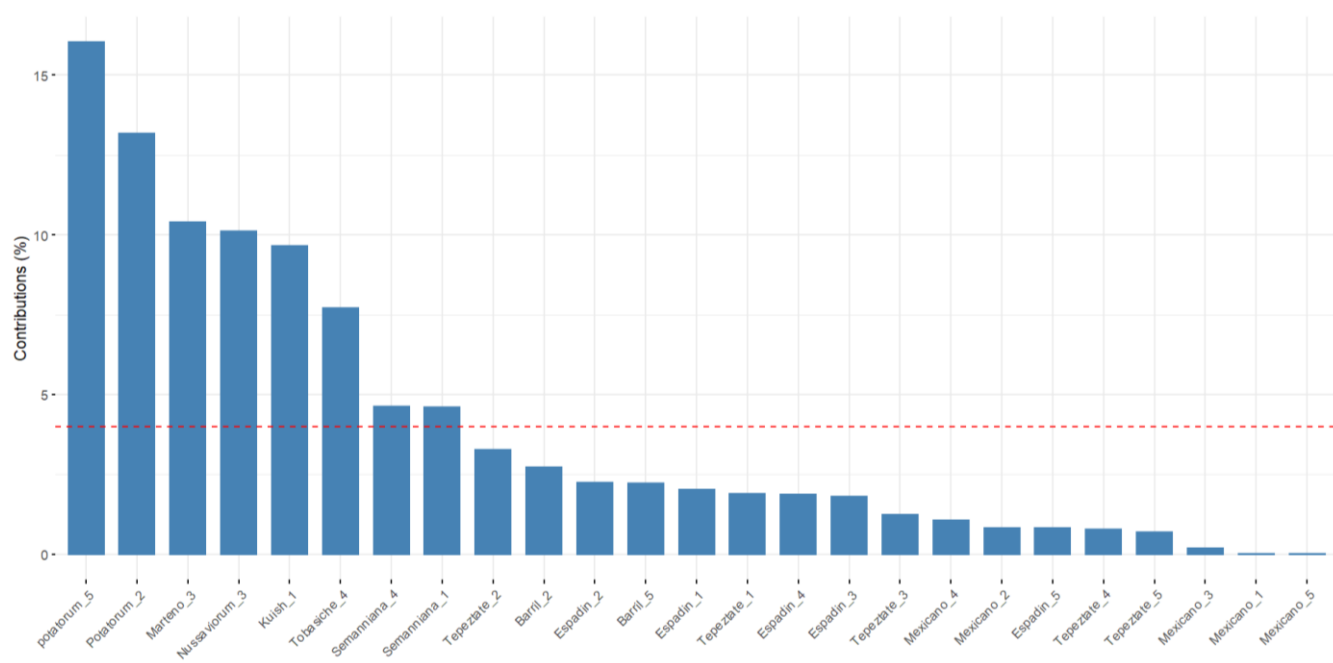


Figura 14. Contribución de cada población a los componentes principales. Los valores por arriba de la línea horizontal promedio de color rojo son altamente significativos.

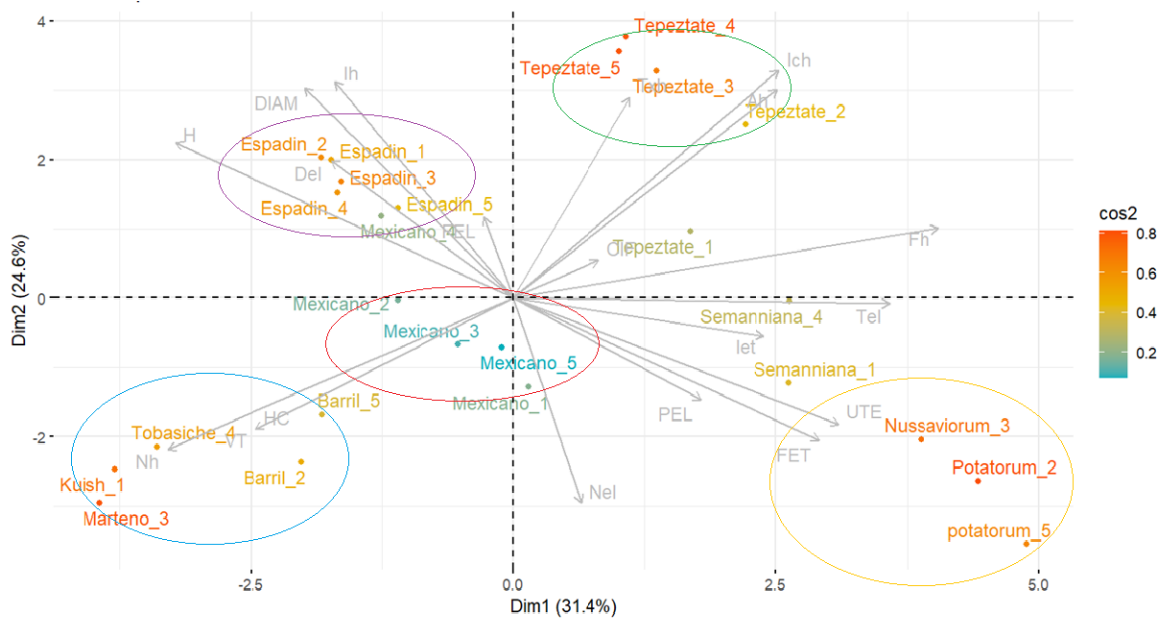


Figura 15. Representación de poblaciones y variables de componentes principales sobre el plano de PCA.

La Figura 15 muestra detalladamente los vectores de cada variable dentro del plano de PCA junto con las poblaciones. Las poblaciones se distribuyen por un gradiente de $\cos^2$ para estimar la calidad de la representación. Donde el color rojo indica mayor calidad en el factor del mapa como lo son las poblaciones Tepeztate_4, Tepeztate_5, Tepeztate_3, Potatorium_2, Potatorium_5, Nussaviorum_3, Marteno_3, Kuish_1 y Espadin_3 y los más azules menor calidad en el factor del mapa para cada población como lo son Mexicano_5, Mexicano_3, Mexicano_2 y Mexicano_4.

En la misma Figura 15 se refleja la representación de cada variable sobre los componentes principales, y los vectores que formó los componentes principales; las poblaciones agrupadas positivamente están correlacionadas como es el caso de Tepeztate que se encuentran en un mismo eje superior derecho y la variable principal que determina a esta población es la textura de la hoja ya que es la única especie (*Agave marmorata*) de las evaluadas que tiene hojas con textura rugosa. Otras variables secundarias que determinan a esta especie son también la intensidad de color (que es un verde fuerte) y el ancho de la hoja. Las poblaciones dentro del eje inferior izquierdo están correlacionadas negativamente ya que se

localizan en un eje opuesto a las demás poblaciones, tal es el caso del complejo *potatorum*, que incluye a las especies *Agave potatorum*, *Agave nussaviorum* y *Agave semanniana* y las variables con un valor alto de correlación son UTE y FET (uniformidad en el tamaño de las espigas y forma de la espiga terminal, respectivamente) que son las variables principales que permiten distinguir a estas tres especies. En el eje superior izquierdo se ubican las poblaciones correspondientes al complejo Espadín (*Agave angustifolia*) y la variable que determina a esta especie es la altura de la hoja principalmente ya que las plantas son muy altas y crecen bastante en un periodo de tiempo más corto en comparación con el resto de las especies cultivadas evaluadas. Otras variables que también determinan a *Agave angustifolia* son el diámetro y la longitud de la hoja. Y, por último, en el eje inferior izquierdo está el complejo *karwinskii* (*Agave karwinskii*) con 5 poblaciones de diferentes nombre común y mismo nombre científico. Las variables que mayormente correlacionaron son el número de hojas (con abundantes hojas a lo largo de la piña), visibilidad del tallo y el hábito de crecimiento, siendo esta última la característica principal que define a esta especie dado que las plantas son caulescentes, es decir, con tallo visible. En este mismo eje también se ubicaron poblaciones correspondientes al complejo Mexicano_3 y Mexicano_5 (*Agave rhodantha*) con poca correlación con las variables antes mencionadas, ya que en el eje las poblaciones están muy cercanas al origen. Morfológicamente estas poblaciones comparten cierto parecido con variables del complejo Espadín, lo que explica por qué no se separaron completamente las cinco poblaciones y por qué no están correlacionadas positivamente con variables específicas. Finalmente, en la Figura 16 se puede apreciar la dispersión de las variables y las observaciones dispuestas en un plano en 3D que explica lo anterior pero visualizado en un plano tridimensional con los tres primeros componentes principales.

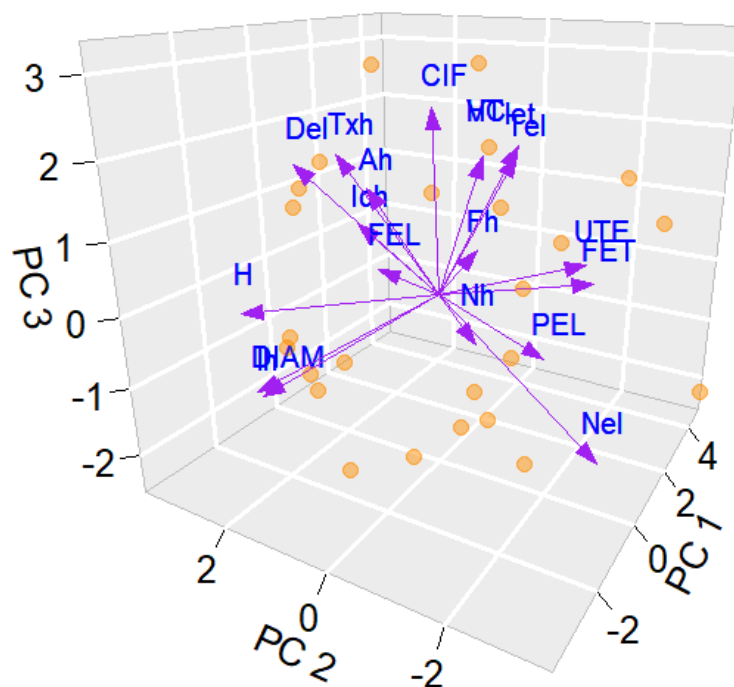


Figura 16. Grafica de Componente principales dispuestas en 3D.

7. CONCLUSIONES

Agave marmorata fue la especie menos relacionada. La variable que explica esta lejanía es Txh, ya que es la única especie que presenta textura rugosa en la hoja, mientras que las demás especies tienen las hojas lisas.

Para el complejo *potatorum*, que incluye a las especies *Agave potatorum*, *Agave nussaviorum* y *Agave semanniana* la variable que mostró un valor alto de correlación fue FET (forma de la espina terminal). Esta variable permite diferenciar muy bien a las especies de este complejo ya que *Agave potatorum* y *Agave nussaviorum* tienen espina terminal curva, mientras que *Agave semanniana* presenta espina terminal recta.

En las poblaciones del grupo *Espadín* (*Agave angustifolia*), la variable que las determina es la altura de planta principalmente, ya que incluyen plantas que crecen mucho en un periodo de tiempo corto en comparación con el resto de las especies evaluadas, y por esta característica, es la especie y poblaciones que los productores prefieren.

El complejo *karwinskii* (*Agave karwinskii*) incluye poblaciones conocidas con nombres comunes diferentes en cada región, pero comparten el mismo nombre científico pues pertenecen a la misma especie y las variables que las caracterizan son el hábito de crecimiento y el número de hojas. Es la especie que se caracteriza por presentar un tallo visible y debido a que las plantas crecen verticalmente generan un gran número de hojas.

En el caso del grupo Mexicano (*Agave rhodantha*) las variables que lo agrupan no mostraron alta correlación en el análisis del PCA, ni con otro método utilizado ya que se comportaron muy dispersas. Morfológicamente, en relación con las variables muestreadas estas poblaciones comparten cierto parecido con variables del complejo Espadín, y con las poblaciones de Marteño.

Los complejos varietales *potatorum* y *karwinskii*, fueron los que mostraron mayor variabilidad fenotípica dentro de la familia Asparagáceae de acuerdo con el análisis multivariado utilizado.

8. LITERATURA CITADA

- Barrera, G.L.A., Legaria, S.J.P., Ortega, P.R. (2020). Diversidad genética en poblaciones de razas mexicanas de maíz. *Revista Fitotecnia Mexicana* 43(1), 121-125.
- García, M. A. (2002). Distribution of Agave (Agavaceae) in Mexico. *Cactus and Succulent Journal*, 74, 177-186.
- García, M. A. (2004). *Agaváceas*. (García, M. A., Ordoñez, M. & Briones, S. M. Eds.). Biodiversidad de Oaxaca. 165p.
- García, M. A. (2011). Flora del Valle Tehuacán-Cuicatlán. Agavaceae. *Instituto de Biología, Universidad Nacional Autónoma de México*, 88, 1-95.
- León, V. N., Campos, A. G. V., Enríquez, J. R., Velasco, V. V. A., Marini, Z. F. & Rodríguez, O. G. (2013). Diversidad de agaves en San Miguel Tilquiápan, Ocotlán, Oaxaca. *Revista Mexicana de Ciencias Agrícolas*, 6, 1185-1195.
- Mora L. J. L., Reyes A., Flores, F.J. L., Peña, V. C. B & Aguirre, R. J. R. (2011). Variación morfológica y humanización de la sección Salmianae del género *Agave*. *Agrociencia* 45, 445-467.

- Moreno, G. J., Cubero, J. L. (1993). *selection strategies and choice of breeding materials. Plant Breeding: Principles and Prospects*. (Hayward, M. D., Bosemark, N. O., Romagosa, I. Eds.). Chapman & Hall. 81-313.
- Servicio Nacional de Inspección y Certificación de semillas. (2014). Guía técnica para la descripción varietal de agave (*Agave spp.*). Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación. 28p.
- Vázquez, M., & de Bernard, M. (1988). Técnicas Del Análisis Estadístico Multivariado: Objetivo y Aplicación. *Revista Cubana de Psicología* (1).
- Villaseñor, J.L. (2016). Lista de las plantas vasculares nativas de México. *Revista Mexicana de Biodiversidad* 87, 559-902.

3 CAPITULO II

ANÁLISIS FILOGENÉTICO DEL GÉNERO AGAVE (ASPARAGACEAE)

1. RESUMEN

Se desarrollo un estudio Basado en secuencias de los marcadores (genes) ITS, *matK* y *trnH-psbA* de individuos depositados en el NCBI. Se efectuó el alineamiento de las secuencias de ADN por medio del programa Clustal W anclado al programa de Bioedit, se utilizó el programa Mesquite para la concatenación de las secuencias y se determinó la filogenia mediante el método de inferencia bayesiana, la cual se ejecutó en el programa MrBayes con 10 millones de generaciones. El análisis filogenético presentó en su mayoría una alta fiabilidad por lo que se puede considerar como una representación adecuada del desarrollo evolutivo del género *Agave*, con nodos que indican los valores de probabilidad posteriores mayor a 70% para ITS, mientras que para los genes *matK* y *trnH-psbA* de cloroplasto, los valores de probabilidad posteriores encontrados en los nodos fueron menores al 70%, proporcionando información menos robusta que los de ITS.

Palabras clave: inferencia bayesiana, análisis multivariado, ITS, *matK*, *trnH-PsbA*, filogenia molecular.

Tesis de Maestría en Ciencias, Universidad Autónoma Chapingo.

Autor: Vásquez Maya María Patricia

Director: Dr. Juan Porfirio Legaria Solano

PHYLOGENETIC ANALYSIS OF THE GENUS *AGAVE* (ASPARAGACEAE)

1. 1 ABSTRACT

A study was developed based on sequences of ITS, *matK* and *trnH-psbA* markers (genes) from individuals deposited in the NCBI. The DNA sequences were aligned using the Clustal W multiple alignment in Bioedit program, the Mesquite program was used to concatenate the sequences and the phylogeny was determined by using the Bayesian Inference Method, which was executed in the MrBayes program with 10 million generations. The phylogenetic analysis was mostly high parsimonious, and can therefore be considered an suitable representation of the evolutionary development of the genus *Agave*, with nodes indicating posterior probability values greater than 70% for ITS, while for the chloroplast *matK* and *trnH-psbA* genes, the posterior probability values found in the nodes were less than 70%, providing less robust information than ITS.

Keywords: Bayesian inference, ITS, *matK*, *trnH-PsbA*, molecular phylogeny.

Thesis, Universidad Autónoma Chapingo.

Author: Vasquez Maya Maria Patricia

Advisor: Dr. Juan Porfirio Legaria Solano

1. INTRODUCCIÓN

La filogenia tiene actualmente un gran impacto sobre todas las subdisciplinas de la Biología, y muy especialmente en la Biogeografía, la Biología de la Conservación, la Ecología, la Etología y la Epidemiología (Palacio *et al.*, 2020).

Aunque el estudio de Jiménez *et al.* (2020) representa un avance respecto a estudios previos, consideramos que la resolución limitada en este y los demás estudios filogenéticos del *Agave* no es un artefacto o error de los estudios revisados, sino el resultado de diferentes procesos biológicos del *Agave*, incluyendo su larga generación. tiempo y facilidad de hibridación porque la mayoría de los agaves e incluso *Manfreda* pueden hibridar entre sí (Eguiarte *et al.*, 2021).

Los marcadores moleculares, como las secuencias de genes, son una herramienta importante en el estudio de las especies vegetales. Han sido usados en especies del género *Agave* para estudios genéticos, caracterización de poblaciones, filogenia, identificación de cultivares e híbridos. En la presente investigación se han utilizado para llevar a cabo una reconstrucción filogenética robusta del género *Agave*.

2. Objetivos

El objetivo del presente estudio consistió en hacer una reconstrucción filogenética del género *Agave*, aumentando sustancialmente el muestreo taxonómico y seleccionando los marcadores moleculares apropiados (secuencias de genes), en relación con estudios previos, con el fin de obtener un mayor nivel de resolución y valores de soporte.

3. MATERIALES Y MÉTODOS

En el Cuadro 5 se describe el subgénero y grupo al que pertenece cada especie analizada en este estudio. Para ello, se utilizaron dos tipos de marcadores: cloroplasto y nuclear ribosomal, este último sólo se había utilizado a nivel intergenérico o en el número limitado de especies dentro del *Agave* (Bogler & Simpson, 1996, Lledías *et al.*, 2020). Además, se utilizaron enfoques bayesianos para estimar el tiempo de divergencia de los principales grupos que conforman el *Agave sensu lato*.

Cuadro 5. Lista de taxones incluidos en los análisis de inferencia filogenética, siguiendo la clasificación infragénica de Agave propuesto por Gentry (1982) y Álvarez (1995).

Subgénero Littaea		
Grupo Littaea	No. De especies muestreadas	Nombre de las especies muestreadas
Amolae (8)	2	<i>attenuata, nizandensis</i>
Marginatae (27)	8	<i>convallis, ghiesbreghtii, horrida, kerchovei, lechuguilla, pelona, titanota, triangularis</i>
Filiferae (7)	1	<i>schidigera</i>
Parviflorae (4)	1	<i>polianthiflora</i>
Polycephalae (5)	2	<i>chiapensis, pendula</i>
Urceolatae (2)	0	
Striatae (5)	3	<i>dasyliroides, rzedowskiana, petrophila</i>
Choripetalae (3)	2	<i>ellemeetiana, guiengola</i>
Subgénero Agave		
Grupo Agave	No. De especies muestreadas	Nombre de las especies muestreadas
Americanae (6)	4	<i>americana, lurida, asperrima, scaposa</i>
Antillanae (14)	0	
Campaniflorae (3)	0	
Ditepalae (12)	3	<i>applanata, colorata, wocomahi</i>
Hiemiflorae (13)	4	<i>atrovirens, isthmensis, potatorum, seemanniana,</i>
Crenatae (7)	3	<i>cupreata, inaequidens, maximiliana</i>
Deserticolae (10)	0	
Marmoratae (4)	1	<i>marmorata</i>
Parryanae (8)	1	<i>parryi</i>
Rigidae (13)	3	<i>angutifolia, rhodacantha, tequilana</i>
Samianae (5)	1	<i>salmiana</i>
Sisalanae (5)	0	
Umbeliflorae ()	1	<i>shawii</i>
Agave sensu lato	No. De especies muestreadas	Nombre de las especies muestreadas
Manfreda (37)	2	<i>scabra, hauniensis,</i>
Polianthes (14)	0	
Prochanyanthes (1)	0	
Grupo externo	No. De especies muestreadas	Nombre de las especies muestreadas
Dasyliirion	2	<i>wheeleri, taxanum</i>
Chlorogalum	2	<i>parviflorum, purpureum</i>
Hesperoyucca	1	<i>whipplei</i>
Hesperaloe	1	<i>parviflora</i>
Yucca	3	<i>filifera, linearifolia, brevifolia</i>
Camassia	2	<i>quamash, leichtlinii</i>
Furcraea	1	<i>pubescens</i>
Beschorneria	1	<i>rigida</i>

* Los números del paréntesis representan el número de especies actualmente conocidas que conforman cada uno de los grupos subgenéricos analizados.

4.1 Secuencias

Las secuencias especificadas en el Cuadro 6 se descendieron de la base de datos depositadas en GenBank del National Center for Biotechnology Information (NCBI), secuencias seleccionadas de la lista de accesiones del estudio filogenético de Jiménez *et al.* (2020) y otros autores.

Cuadro 6. Lista de accesiones del GenBank para las secuencias implementadas, consultadas de la lista de Jiménez et al. (2020) y otros autores.

Especie	ITS	Referencia	matK	Referencia	trnH-psbA	Referencia
<i>Dasyllirion wheeleri</i>	U24014.1/U24035.1	Bogler, et al., 1995	HM640588	Kim et al., 2016	-	-
<i>Dasyllirion texanum</i>	U24034.1/U24013.1	Bogler, et al., 1995	-	-	-	-
<i>Hesperoyucca whipplei</i>	MF963981.1	Thornhill et al., 2017	MF963603	Thornhill et al., 2017	-	-
<i>Chlorogalum parviflorum</i>	MF964117.1	Thornhill et al., 2017	MF963609.1	Thornhill et al., 2017	-	-
<i>Chlorogalum purpureum</i>	MF963989.1	Thornhill et al., 2017	MF963610	Thornhill et al., 2017	-	-
<i>Camassia quamash</i>	KP008320.1	Archibald et al., 2015	MH748950	Zuniga y Mulcahy, 2019	-	-
<i>Camassia leichtlinii</i>	KP008304.1	Archibald et al., 2015	KX676620	Fenneman y Graham, 2019	-	-
<i>Hesperaloe parviflora</i>	MT083804	Jimenez et al., 2020	-	-	-	-
<i>Yucca filifera</i>	MT083805	Jimenez et al., 2020	MT250133	Jimenez et al., 2020	MT053314	Jimenez et al., 2020
<i>Yucca linearifolia</i>	MT083806	Jimenez et al., 2020	-	-	-	-
<i>Yucca brevifolia</i>	MF964125.1	Thornhill et al., 2017	-	-	-	-
<i>Beschorneria rigida</i>	MT083810	Jimenez et al., 2020	MT250134	Jimenez et al., 2020	-	-
<i>Furcraea longaevea</i>	-	-	-	-	MT053315	Jimenez et al., 2020
<i>Furcraea pubescens</i>	MT083813	Jimenez et al., 2020	MN893690	García et al., 2021	MT053316	Jimenez et al., 2020
<i>Manfreda hauniensis</i>	MT083814	Jimenez et al., 2020	MT250135	Jimenez et al., 2020	MT053317	Jimenez et al., 2020
<i>Manfreda scabra</i>	U23983.1/U24042.1	Bogler, D.J., 1995	-	-	-	-
<i>Agave americana</i>	MT083820	Jimenez et al., 2020	MT250137	Jimenez et al., 2020	MT053319	Jimenez et al., 2020
<i>Agave angustifolia</i>	MT083821	Jimenez et al., 2020	MT250138	Jimenez et al., 2020	MT053320	Jimenez et al., 2020
<i>Agave angustiarum</i>	AM884821	García et al., 2008	AM884156	García et al., 2016	-	-
<i>Agave applanata</i>	MT083823	Jimenez et al., 2020	MT250139	Jiménez et al., 2020	-	-
<i>Agave aspérrima</i>	MT083825	Jiménez et al., 2020	MT250140	Jiménez et al., 2020	MT053321	Jimenez et al., 2020
<i>Agave atrovirens</i>	MT083826	Jiménez et al., 2020	MT250166	Jiménez et al., 2020	-	-
<i>Agave attenuata</i>	MT083827	Jiménez et al., 2020	MT250141	Jiménez et al., 2020	MT053322	Jimenez et al., 2020
<i>Agave chiapensis</i>	MT083830	Jiménez et al., 2020	MN893674	García et al., 2021	-	-
<i>Agave colorata</i>	MT083831	Jiménez et al., 2020	MT250142	Jiménez et al., 2020	MT053323	Jimenez et al., 2020
<i>Agave convallis</i>	AM884822	García et al., 2008	AM884157	García et al., 2016	MT053324	Jimenez et al., 2020

<i>Agave cupreata</i>	MT083832	Jiménez et al., 2020	MT250143	Jiménez et al., 2020	MT053325	Jimenez et al., 2020
<i>Agave dasylirioides</i>	MT083833	Jiménez et al., 2020	MT250144	Jiménez et al., 2020	MT053326	Jimenez et al., 2020
<i>Agave ellemeetiana</i>	MT083836	Jiménez et al., 2020	MT250145	Jiménez et al., 2020	-	-
<i>Agave ghiesbreghtii</i>	-	-	-	-	MT053327	Jimenez et al., 2020
<i>Agave guiengola</i>	MT083841	Jiménez et al., 2020	MT250147	Jiménez et al., 2020	MT053328	Jimenez et al., 2020
<i>Agave horrida</i>	MT083842	Jiménez et al., 2020	MT250148	Jiménez et al., 2020	-	-
<i>Agave isthmensis</i>	MT083844	Jiménez et al., 2020	-	-	-	-
<i>Agave karwinskii</i>	MN900424	García et al., 2008	-	-	-	-
<i>Agave kavandivi</i>	MN900425	García et al., 2008	-	-	-	-
<i>Agave kerchovei</i>	MT083845	Jiménez et al., 2020	AM884161	García et al., 2016	-	-
<i>Agave inaequidens</i>	MT083843	Jiménez et al., 2020	MT250149	Jiménez et al., 2020	MT053329	Jimenez et al., 2020
<i>Agave lechuguilla</i>	MT083846	Jiménez et al., 2020	MT250150	Jiménez et al., 2020	MT053330	Jimenez et al., 2020
<i>Agave lurida</i>	MT083847	Jiménez et al., 2020	MN893678	García et al., 2021	MT053331	Jimenez et al., 2020
<i>Agave macracantha</i>	MN900428	García et al., 2008	-	-	-	-
<i>Agave mapisaga</i>	MN056968	Trejo et al., 2019	-	-	-	-
<i>Agave marmorata</i>	MT083848	Jiménez et al., 2020	-	-	MT053332	Jimenez et al., 2020
<i>Agave maximiliana</i>	MT083849	Jiménez et al., 2020	MT250151	Jiménez et al., 2020	-	-
<i>Agave nizandensis</i>	MT083851	Jiménez et al., 2020	MT250152	Jiménez et al., 2020	MT053333	Jimenez et al., 2020
<i>Agave parryi</i>	MT083852	Jiménez et al., 2020	-	-	-	-
<i>Agave pelona</i>	MT083853	Jiménez et al., 2020	MT250153	Jiménez et al., 2020	-	-
<i>Agave pendula</i>	MT083854	Jiménez et al., 2020	-	-	-	-
<i>Agave petrophila</i>	MT083855	Jiménez et al., 2020	MT250154	Jiménez et al., 2020	-	-
<i>Agave polianthiflora</i>	MT083857	Jiménez et al., 2020	MT250155	Jiménez et al., 2020	MT053334	Jimenez et al., 2020
<i>Agave potatorum</i>	MT083858	Jiménez et al., 2020	MT250156	Jiménez et al., 2020	MT053335	Jimenez et al., 2020
<i>Agave rhodacantha</i>	MT083859	Jiménez et al., 2020	MT250157	Jiménez et al., 2020	MT053336	Jimenez et al., 2020
<i>Agave rzedowskiana</i>	MT083860	Jiménez et al., 2020	MT250158	Jiménez et al., 2020	MT053337	Jimenez et al., 2020
<i>Agave salmiana</i>	MT083861	Jiménez et al., 2020	MT250159	Jiménez et al., 2020	-	-
<i>Agave scaposa</i>	MT083862	Jiménez et al., 2020	-	-	-	-
<i>Agave schidigera</i>	MT083863	Jiménez et al., 2020	MT250160	Jiménez et al., 2020	MT053338	Jimenez et al., 2020
<i>Agave seemanniana</i>	MT083865	Jiménez et al., 2020	MT250161	Jiménez et al., 2020	MT053339	Jimenez et al., 2020

<i>Agave stricta</i>	AM884825	García et al., 2008	AM884160	García et al., 2016	-	-
<i>Agave shawii</i>	MT083866	Jiménez et al., 2020	MT250162	Jiménez et al., 2020	-	-
<i>Agave tequilana</i>	MT083869	Jiménez et al., 2020	MT250163	Jiménez et al., 2020	-	-
<i>Agave titanota</i>	MT083870	Jiménez et al., 2020	MT250164	Jiménez et al., 2020	MT053340	Jimenez et al., 2020
<i>Agave triangularis</i>	MT083871	Jiménez et al., 2020	MT250165	Jiménez et al., 2020	MT053341	Jimenez et al., 2020
<i>Agave wocomahi</i>	MT083874	Jiménez et al., 2020	MT250167	Jiménez et al., 2020	-	-

*-indica accesiones no reportadas en el GenBank.

4.2 Alineamiento de las secuencias

Las secuencias se alinearon y editaron usando BioEdit versión 7.2 (Hall, 1999) para crear un alineamiento con Clustal W que está anclado a la interfase de BioEdit. seguido de ajuste manual en PhyDE para editar las secuencias.

4.3 Concatenación de las secuencias

Se realizó la concatenación con el programa Mesquite. La matriz concatenada fue utilizada para la filogenia. Se concatenaron 740 pb del espaciador transcrito interno (ITS), de MatK (1445 pb) y trnH-psbA (562pb) de 62 especies, Asparagaceae (46), *Yucca* (3), *Dasyilirion* (2), *Hesperoyucca* (1), *Chlorogalum* (2), *Camassia* (2), *Heperaloe* (1), *Manfreda* (2), *Furcraea* (2), *Bescorneria* (1), utilizando a *Yucca* como grupo externo.

4.4 Construcción del árbol filogenético

Un árbol filogenético bayesiano fue reconstruido usando el programa MrBayes 3.2.7 (Roquist & Huelsenbeck, 2003). El mejor modelo de sustitución para cada conjunto de particiones se seleccionó utilizando una estrategia de salto reversible y la heterogenia de tasas se modeló con una distribución gamma (Huelsenbeck & Rannala, 2004). Se realizaron dos corridas independientes con cuatro cadenas (tres calentadas y una fría) simultáneamente para 10.000.000 de generaciones y muestreos cada 1.000 generaciones. Cuando el valor estimado del tamaño de la muestra (SEE) excedió de 200 y el factor de reducción de la escala potencial (PSRF) fue cercano a 1.0, se consideró que se produjo la convergencia de las cadenas. Las muestras del 25% fueron desechadas como burn-in.

La matriz utilizada para la filogenia del cloroplasto incluyó secuencias de matK (1445 pb) y trnH-psbA (562pb), el mejor modelo de sustitución fue el modelo GTR+ G+ I. El mejor modelo de sustitución para cada partición se corroboró en jModelTest (Darriba *et al.*, 2012) bajo un BIC y luego se utilizó en los análisis filogenéticos. Los análisis con Mrbayes se realizaron durante 10 000000 de generaciones con 1000 repeticiones de arranque, y *Yucca filifera* se especificó como grupo externo.

5. RESULTADOS Y DISCUSIÓN

Para el conjunto de datos ITS, se analizaron 740 pb para el total de 60 especies. El análisis de inferencia Bayesiana dio lugar a topologías congruentes; en el Cuadro 7 se determina el grado de parentesco entre las secuencias y se visualiza en los árboles filogenéticos. En la Figura 17 se muestra el árbol filogenético molecular de *Agave* sensu lato utilizando a *Yucca* como un grupo externo. La filogenia se reconstruyó a partir de las secuencias del espaciador transcrito interno ITS y utilizando un análisis de inferencia bayesiana. Los números junto a los nodos indican los valores de probabilidad posteriores mayor a 70%, que son los de mejor soporte.

Cuadro 7. Resumen general de los parámetros genéticos para cada locus.

Locus	ALPB (%)	SC (%)	SV (%)	SPI (%)
ITS	740	441 (59)	297 (40)	122 (16)
matK	1445	884 (61)	561 (39)	541 (37)
trnH-psbA	562	532 (95)	30 (5)	5 (0.9)

ALPB= alineamiento longitudinal en pares de bases; SC= sitios conservados; SV= sitios variables; SPI= sitios parsimonio-informativos.

El filograma obtenido usando las secuencias del ITS (Figura 17) coincide topológicamente con el obtenido en el trabajo de Jiménez *et al.* (2020). El filograma reconstruido en el presente estudio con las secuencias ITS, contiene mayores sitios variables con un porcentaje de 40% con relación a los genes de cloroplastos, matK con un 30% y trnH-psbA con solo un 5%, esto confirma la posición filogenética de las Agavaceae dentro de las Asparagales, aunque plantea algunas dudas sobre los límites filogenéticos de la familia.

El promedio de soporte encontrados en los nodos para el conjunto de datos de MatK y trnH_psbA de cloroplastos tiene valores menores al 70% (Figura 18), lo que no permitió una buena separación de los clados, lo que es también congruente con los análisis anteriores que incluyeron marcadores de plastidios para el clado *Agave* sensu lato (Jiménez *et al.*, 2020). Los filogramas obtenidos previamente en diferentes estudios detectaron clados en los que las especies *Manfreda*, *Polianthes* y *Prochnyanthes* se sitúan dentro del grupo *Agave* sensu stricto (Good-Ávila *et al.*, 2006; Flores Abreu *et al.*, 2019). De hecho, el *Agave* sensu lato es un grupo que ha sido difícil de clasificar taxonómicamente debido a la superposición de la variación de caracteres morfológicos entre especies (Gentry, 1982; García, 2002), lo que podría deberse al reciente origen del grupo, a los recientes eventos de diversificación, así como a la permisiva hibridación entre especies y a un largo tiempo de generación (Mckain *et al.*, 2016).

El análisis filogenético mostró una alta fiabilidad por lo que se puede considerar como una representación apta del desarrollo evolutivo del género *Agave*. En el caso de las secuencias de genes de cloroplasto y nuclear ribosomal, los análisis de inferencia Bayesiana respaldaron la conformación de su clado, especialmente con los marcadores nucleares ribosomales como ITS, donde el soporte de los nodos fue superior a 0.70 de probabilidad, excepto por dos nodos con valores de 0.58 y 0.56; la razón de esto puede ser por el uso de diferentes accesiones provenientes de diferentes autores, o incluso porque se podría requerir un incremento en el número de las generaciones de corrida en Mr. Bayes, posiblemente hasta 20000000.

Las diferencias en la clasificación taxonómica de las especies de Asparagáceae son producto de la amplia variabilidad morfológica presentes entre las muestras estudiadas para cada especie. Esta variación se debe a los procesos de selección natural principalmente, ubicación geográfica de las muestras recolectadas y flujo genético entre especies. En estudios anteriores no habían recuperado a *Agave* como un género monofilético incluso en el estudio más actual aún se sitúa *Agave* sensu stricto como no monofilético, esto podría deberse a la existencia de rasgos plesiomórficos y apomórficos lo que también provoca cambios en la clasificación

taxonómica. Se requiere una revaloración taxonómica para ubicar correctamente a *Agave* y un estudio a más profundidad utilizando accesiones nuevas con previa revisión taxonómica. Los nuevos estudios para *Agave* requieren un enfoque integral de bases de datos para determinar con mayor precisión las relaciones filogenéticas dentro del género.

El arranque no paramétrico es indicativo de la cantidad de evidencia en una matriz de datos que respalda cada clado en el árbol, mientras que las probabilidades posteriores bayesianas no pretenden representar esa propiedad. Los clados con alta probabilidad posterior pueden no tener una gran cantidad de caracteres que los favorezcan, y sus frecuencias son el resultado del procedimiento de muestreo particular del método Monte Carlo de la cadena bayesiana de Markov, que tiende a muestrear topologías muy similares de acuerdo con sus probabilidades posteriores. Ambas métricas pueden relacionarse con la noción de confianza, pero representan diferentes propiedades (García, 2014).

La manera en que estos cambios acumulados son inferidos es una de las diferencias entre los métodos paramétricos (inferencia bayesiana y máxima verosimilitud) y los métodos no-paramétricos (parsimonia). Esto implica que en un filograma inferido con métodos paramétricos la longitud de las ramas es calculada de una manera diferente a como se hace con un método de parsimonia (García, 2015).

6. CONCLUSIONES

El análisis filogenético mostró una alta fiabilidad por lo que se puede considerar como una representación apta del desarrollo evolutivo del género *Agave*.

En el caso de las secuencias de genes de cloroplasto y nuclear ribosomal, el análisis de inferencia Bayesiana respaldó la conformación de los clados, especialmente con los marcadores nucleares ribosomales como ITS, donde el soporte de los nodos fue superior a 0.70 de probabilidad.

Los nuevos estudios para *Agave* requieren un enfoque integral de bases de datos para determinar con mayor precisión las relaciones filogenéticas dentro del género.

7. LITERATURA CITADA

- Alvarez, A. (1995). Los agaves de las antillas. *Botanical Sciences*, (57), 37-48.
<https://doi.org/10.17129/botsci.1475>
- Bogler, D. J., & Simpson, B. B. (1996). Phylogeny of Agavaceae Based on ITS rDNA Sequence Variation. *American Journal of Botany*, 83(9), 1225-1235.
<https://doi.org/10.2307/2446206>
- Bogler, D. J., Pires, J. C., & Francisco, O. J. (2006). Phylogeny of Agavaceae Based on ndhF, rbcL, and ITS sequences. *Aliso: A journal of Systematic and Floristic Botany*, 22, 313-328. <https://scholarship.claremont.edu/aliso/vol22/iss1/26>
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772.
<https://doi.org/10.1038/nmeth.2109>
- Eguiarte, L. E., Jiménez, B. O. A., Aguirre, P. E., Scheinvar, E., Gámez, N., Gasca-P. J., Castellanos, M. G., Moreno, L. A., & Souza, V. (2021). Evolutionary ecology of Agave: distribution patterns, phylogeny, and coevolution (an homage to Howard S. Gentry). *Revista American Journal of Botany*, 108(2), 216–235.
<https://doi.org/10.1002/ajb2.1609>
- Flores, A. I. N., Trejo, S. R. E., Sánchez, R. L. L., Good, S. V., Magallón, S., García, M. A., & Eguiarte, L. E. (2019). Tempo and mode in coevolution of Agave sensu lato (Agavoideae, Asparagaceae) and its bat pollinators, Glossophaginae (Phyllostomidae). *Molecular Phylogenetics and Evolution*, 133, 176–188.
<https://doi.org/10.1016/j.ympev.2019.01.004>
- García, S. R. (2014). Why some clades have low bootstrap frequencies and high Bayesian posterior probabilities. *Israel Journal of Ecology & Evolution*, 60(1), 41-44. <https://doi.org/10.1080/15659801.2014.937900>
- García, S. R. (2015). Modelos de sustitución de nucleótidos (y otros modelos). *Facultad de Ciencias UNAM*. 10p. 10.13140/RG.2.1.2048.8168.
- García, M. A. (2002). Distribution of Agave (Agavaceae) in Mexico. *Cactus and Succulent Journal*, 74, 177-186.
- Good, A. S. V., Souza, V., Gaut, B. S., & Eguiarte, L. E. (2006). Timing and rate of speciation in Agave Agavaceae). *Proceedings of the National Academy of*

- Sciences U.S.A*, 103(24), 9124-9129. <https://doi.org/10.1073/pnas.0603312103>
- Gentry, H. S. (1982). *Agaves of Continental North America*. Tucson: University of Arizona. 670p.
- Hall, B. K. (2007). Homoplasy and homology: dichotomy or continuum? *Journal of Human Evolution*, 52(5): 473-479.
- Huelsenbeck, J. P., & Rannala, B. (2004). Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology*, 53(6), 904-913. <https://doi.org/10.1080/10635150490522629>
- Jiménez, B. O., García, S. R., Magallón, S., García, M. A., Nieto, S. J., Aguirre, P. 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>
- Lledías, F., Gutiérrez, J., Martínez, H. A., García, M. A., Sosa, E., Hernández, B. F. (2020). Mayahuelin, a type I Ribosome Inactivating Protein: characterization, evolution, and utilization in phylogenetic analyses of *Agave*. *Frontiers in Plant Science*, 11 (573). <https://doi.org/10.3389/fpls.2020.00573>
- McKain, M. R., McNeal, J. R., Kellar, P. R., Eguiarte, L. E., Pires, J.C. y Leebens-Mack, J. (2016). Timing of rapid diversification and convergent origins of active pollination within Agavoideae (Asparagaceae). *American Journal of Botany*, 103(10), 1717-1729. <https://doi.org/10.3732/ajb.1600198>
- Palacio, F., Apocada, M., & Crisci, J. (Eds.). (2020). Análisis multivariado para datos biológicos: teoría y su aplicación utilizando el lenguaje R. Fundacion de Historia Natural Felix de Azara. <https://fundacionazara.org.ar/img/libros/analisis-multivariado-para-datos-biologicos/analisis-multivariado-para-datos-biologicos.pdf> 268p.
- Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed modes. *Bioinformatics*, 19 (12), 1572-1574. <https://doi.org/10.1093/bioinformatics/btg180>