



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**POSGRADO EN CIENCIAS BIOLÓGICAS**  
**FACULTAD DE ESTUDIOS SUPERIORES DE IZTACALA**  
**BIOLOGÍA EVOLUTIVA**

**VARIACIÓN GENÓMICA POBLACIONAL Y CARACTERIZACIÓN  
DE LA DISTRIBUCIÓN ACTUAL Y FUTURA DE *Vanilla planifolia*  
JACKS. EX ANDREWS (ORCHIDACEAE)**

# **TESIS**

**QUE PARA OPTAR POR EL GRADO DE:  
MAESTRO EN CIENCIAS BIOLÓGICAS**

**PRESENTA:**

**BIOL. ALEJANDRO RODRÍGUEZ AGUILAR**

**TUTOR PRINCIPAL DE TESIS:**

**DR. JORGE EDUARDO CAMPOS CONTRERAS**  
FES IZTACALA, UNAM

**COMITÉ TUTOR:**

**DRA. ESTELA SANDOVAL ZAPOTITLA**  
INSTITUTO DE BIOLOGÍA, UNAM

**DR. JORGE CIROS PÉREZ**  
FES IZTACALA, UNAM

**Los Reyes Iztacala, Tlalnepantla, Estado de México, Junio, 2021**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.





**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**POSGRADO EN CIENCIAS BIOLÓGICAS**  
**FACULTAD DE ESTUDIOS SUPERIORES DE IZTACALA**  
**BIOLOGÍA EVOLUTIVA**

**VARIACIÓN GENÓMICA POBLACIONAL Y CARACTERIZACIÓN  
DE LA DISTRIBUCIÓN ACTUAL Y FUTURA DE *Vanilla planifolia*  
JACKS. EX ANDREWS (ORCHIDACEAE)**

# **TESIS**

**QUE PARA OPTAR POR EL GRADO DE:  
MAESTRO EN CIENCIAS BIOLÓGICAS**

**PRESENTA:**

**BIOL. ALEJANDRO RODRÍGUEZ AGUILAR**

**TUTOR PRINCIPAL DE TESIS:**

**DR. JORGE EDUARDO CAMPOS CONTRERAS**  
FES IZTACALA, UNAM

**COMITÉ TUTOR:**

**DRA. ESTELA SANDOVAL ZAPOTITLA**  
INSTITUTO DE BIOLOGÍA, UNAM

**DR. JORGE CIROS PÉREZ**  
FES IZTACALA, UNAM

**Los Reyes Iztacala, Tlalnepantla, Estado de México, 2021**

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

OFICIO CPCB/352/2020

ASUNTO: Oficio de Jurado

**M. en C. Ivonne Ramírez Wence**  
Directora General de Administración Escolar, UNAM  
Presente

Me permito informar a usted que en la reunión ordinaria del Subcomité de Ecología y Manejo Integral de Ecosistemas y Biología Evolutiva y Sistemática, del Posgrado en Ciencias Biológicas, celebrada el día **20 de enero de 2020** se aprobó el siguiente jurado para el examen de grado de **MAESTRO EN CIENCIAS BIOLÓGICAS** en el campo de conocimiento de **Biología Evolutiva** del estudiante **RODRÍGUEZ AGUILAR ALEJANDRO**, con número de cuenta **308128954** con la tesis titulada "**Variación genómica poblacional y caracterización de la distribución actual y futura de *Vanilla planifolia* Jacks. Ex Andrews (Orchidaceae)**", realizada bajo la dirección del **DR. JORGE EDUARDO CAMPOS CONTRERAS**, quedando integrado de la siguiente manera:

Presidente: DR. BRAULIO EDGAR HERRERA CABRERA  
Vocal: DR. VICTOR MANUEL SALAZAR ROJAS  
Secretario: DR. JORGE CIROS PÉREZ  
Suplente: DRA. MARIANA HERNÁNDEZ APOLINAR  
Suplente: DRA. MARTHA MARTÍNEZ GARCÍA

Sin otro particular, me es grato enviarle un cordial saludo.

**ATENTAMENTE**  
"POR MI RAZA HABLARA EL ESPÍRITU"  
Ciudad Universitaria, Cd. Mx., a 18 de marzo de 2020

**COORDINADOR DEL PROGRAMA**



**DR. ADOLFO GERARDO NAVARRO SIGÜENZA**



c. c. p. Expediente del alumno

**COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS**  
**UNIDAD DE POSGRADO**

Edificio D, 1° Piso. Circuito de Posgrados, Ciudad Universitaria  
Alcaldía Coyoacán. C. P. 04510 CDMX

Tel. (+5255)5623 7002 <http://pcbiol.posgrado.unam.mx/>

## **Agradecimientos Institucionales**

Al Posgrado en Ciencias Biológicas de la UNAM, por permitirme realizar mis estudios de Maestría y brindarme herramientas para realizarme profesional y académicamente.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo otorgado mediante la beca de posgrado (CVU: 745004) durante los semestres desde el 2017-2 al 2019-1. La cual fue fundamental para realizar mis estudios de Maestría.

A mi director de tesis, el Dr. Jorge Eduardo Campos Contreras, por confiarme un proyecto tan interesante y de gran desafío, por aceptarme como su estudiante, brindarme su tiempo, conocimiento, orientación y su infinita paciencia.

A los miembros de mi comité tutorial, por su compromiso, sugerencias, comentarios y correcciones que enriquecieron en gran medida este trabajo.

- Dra. Estela Sandoval Zapotitla, por brindarme su apoyo y asesoría, así como por su amistad, confianza, comprensión y consejos.
- Dr. Jorge Ciros Pérez, por su asesoría, y apoyo durante la realización de esta tesis.

## Agradecimientos

A mis padres, por su amor y apoyo incondicional, por darme estabilidad y enseñarme a ser paciente, a confiar en mis capacidades, por alentarme a seguir creciendo académicamente, profesionalmente y como persona. Les agradezco infinitamente sus desvelos y esfuerzos que realizaron para que nunca me faltara nada. A ustedes les debo todo lo que soy. Los amo.

A Liz, que siempre estuvo al pendiente de mí, con tanto amor, consejos y entrega incondicional. No hay jefa más chida que tú, mamá.

A Alejandro, porque ahora te comprendo y eso me ha hecho valorar aún más los sacrificios que haces por nosotros, gracias por enseñarme el sentido de la responsabilidad. Gracias por ser un buen padre.

A Oscar, mi hermano, por ser un compañero de vida y travesuras, por las vivencias que tuvimos desde niños, por cuidarme y apoyarme, por tus consejos profesionales y personales, por estar conmigo en las buenas y en las malas. Te quiero, *Charly*.

A mi familia, gracias por estar y apoyarme siempre, Armando, Loth, Ili, Elena, Arturo. Gracias por ayudarme a ser quien soy. Muchísimas gracias por sus enseñanzas.

A Francisco, mi mejor amigo, por todo el tiempo y los momentos que hemos vivido estos años. Gracias por ser mi amigo y por estar ahí cuando te necesito. Te quiero, joven.

A Sayra, por apoyarme en todo sin titubear, por la motivación que tienes hacia mí, por los ánimos, por tu tiempo y por tu cariño.

A mis compañeros y amigos Jake, Julio, Mario, Beto, por las experiencias, las risas y aventuras, por aferrarse a nuestra amistad, aunque nunca podamos vernos.

A Alexa, mi mejor amiga, por estar siempre para escucharme, salir y acompañarme en las divertidas aventuras que hemos vivido. Te quiero, maldita.

A Brenda, por todo el apoyo y palabras de aliento que siempre me brindaste.

A mis grandes amigos, Alan, Gus y Jair, por todos los momentos divertidos y amenos que hemos pasado juntos, por brindarme su amistad durante tantos años, por sus consejos y por las aventuras tan chidas que nos han llenado de anécdotas memorables.

A mis compañeros en el laboratorio de bioquímica molecular, por hacer más ameno el trabajo con sus pláticas y comidas: Ari, Clau, Karen, Ana, Roberto, Andrea, Chava. En especial a Nelly, por tu amistad, apoyo, consejos y ayuda que me brindaste desde que nos iniciamos en el mundo de la bioinformática.

A la Dra. Martha, al Maestro Alex y al Dr. Víctor por sus consejos y pláticas, por su apoyo con mis dudas a lo largo de mi estadía en el laboratorio.

A los que ya no están, mis abuelos: *Nina, Má y Pá*, que siempre estuvieron al pendiente de mí. A Lalo, mi mejor amigo durante mi infancia. Y a Bethoveen, el mejor perrito del mundo. Los extraño.

A mis amigos, por estar en el momento y lugar indicado.

A la UNAM, por todo lo que me ha brindado, por la formación y por todo lo que he aprendido y conocido desde que ingresé a la preparatoria y por permitirme ser parte de tan prestigiada institución.



Ser científico es ser ingenuo. Estamos tan centrados en conocer la verdad que no consideramos la poca gente que quiere que la descubramos. Pero la verdad siempre está ahí, la veamos o no, elijamos verla o no. Le da igual nuestro gobierno, nuestra ideología, nuestra religión.  
Esperará eternamente.

-Chernóbil (miniserie HBO) 2019

¿La ciencia ha prometido la felicidad? No lo creo. Ha prometido la verdad y la cuestión es saber si con la verdad se conseguirá algún día la felicidad.

-Émile Zola

# INDICE

INDICE .....	i
INDICE DE FIGURAS.....	iii
INDICE DE TABLAS .....	iv
Resumen.....	1
Abstract.....	2
1. Introducción .....	3
2. Marco Teórico .....	5
2.1 Variación biológica.....	5
2.2 Variación genética .....	5
2.3 Genética y genómica de poblaciones .....	7
2.4 Secuenciación de nueva generación .....	9
2.5 Domesticación y recursos fitogenéticos.....	10
2.6 <i>Vanilla planifolia</i> .....	11
2.6.1 Diversidad en <i>V. planifolia</i> (Antecedentes).....	16
2.7 Variación geográfica.....	20
2.8 Unidades Evolutivas Significativas .....	21
3. Justificación .....	22
4. Objetivos.....	23
4.1 Objetivo general.....	23
4.2 Objetivos particulares .....	23
5. Material y método .....	24
5.1 Colecta de Material Biológico .....	24
5.2 Análisis Molecular.....	24
5.3 Análisis bioinformático (genoma) .....	25
5.3.1 Filtrado e identificación de SNPs.....	25
5.3.2 Diversidad y diferenciación genética.....	27
5.3.3 Análisis de Componentes Principales.....	28
5.3.4 Análisis Discriminante de Componentes Principales .....	28
5.3.5 Estructura genética .....	29
5.4 Modelos de distribución de especies .....	29
5.4.1 Modelos de distribución actual.....	29

5.4.2 Modelo de distribución futura .....	31
5.5 Relación genómica-geográfica-ambiental .....	32
6. Resultados .....	34
6.1 Variación genómica .....	34
6.1.1 Estructura genética poblacional.....	37
6.1.2 SNPs asociados a los sitios de colecta .....	40
6.2 Modelos de distribución de <i>V. planifolia</i> .....	44
6.2.1 Modelos de distribución actual.....	44
6.2.2 Modelos de distribución futura .....	51
6.3 Relación genómica-geográfica-ambiental .....	56
7. Discusión.....	57
7.1 Ensamble de novo.....	57
7.2 Variación genómica .....	58
7.2.1 Estructura genética .....	59
7.3 Modelos de distribución.....	60
7.3.1 Modelos de distribución actual de <i>V. planifolia</i> .....	60
7.3.2 Distribución futura de <i>V. planifolia</i> .....	61
7.4 Relación genómica-geográfica-ambiental .....	62
7.5 Unidades Evolutivamente Significativas (UES) y el estado de conservación de <i>V. planifolia</i> .....	63
8. Conclusiones .....	65
9. Perspectivas .....	66
10. Referencias bibliográficas .....	67
Anexo 1 .....	77
Anexo II.....	78
Anexo III.....	78
Anexo IV .....	79

## INDICE DE FIGURAS

<b>Figura 1.</b> Tipos de selección. ....	7
<b>Figura 2.</b> Esquema de <i>Vanilla planifolia</i> .....	13
<b>Figura 3.</b> Distribución conocida de <i>Vanilla planifolia</i> en México .....	15
<b>Figura 4.</b> A) Dendrograma Quimiotipos B) Dendrograma Genotipos .....	17
<b>Figura 5.</b> Neighbor-Joining de 88 individuos silvestres de la península de Yucatán .....	18
<b>Figura 6.</b> Dispersión de especímenes con base variables morfológicas.....	19
<b>Figura 7.</b> Mapa de las ecorregiones terrestres de México .....	30
<b>Figura 8.</b> Cambios en las temperaturas globales para el escenario RCP2.6 y RCP8.5.....	32
<b>Figura 9.</b> Análisis bicluster.....	35
<b>Figura 10.</b> Dendrograma por el método Neighbor-joining de los SNPs encontrados en el mapeo al genoma de <i>V. planifolia</i> .....	38
<b>Figura 11.</b> Análisis de componentes principales. ....	39
<b>Figura 12.</b> Análisis de agrupamiento.....	40
<b>Figura 13.</b> DAPC por sitio de colecta .....	42
<b>Figura 14.</b> Análisis de Componentes principales de A) variables climáticas Worldclim y B) variables climáticas de CONABIO .....	45
<b>Figura 15.</b> AUC para los modelos de distribución de <i>V. planifolia</i> .....	46
<b>Figura 16.</b> Prueba de Jackknife .....	48
<b>Figura 17.</b> Mapa de la distribución actual de <i>V. planifolia</i> utilizando Worldclim.....	49
<b>Figura 18.</b> Mapa de la distribución actual de <i>V. planifolia</i> utilizando CONABIO.....	50
<b>Figura 19.</b> Áreas de distribución en Km <sup>2</sup> .....	51
<b>Figura 20.</b> Mapa de distribución en RCP 2.6 para el año 2050.....	52
<b>Figura 21.</b> Mapa de distribución en RCP 8.5 para el año 2050.....	53
<b>Figura 22.</b> Mapa de distribución en RCP 2.6 para el año 2070.....	54
<b>Figura 23.</b> Mapa de distribución en RCP 8.5 para el año 2070.....	55

## INDICE DE TABLAS

<b>Tabla 1.</b> Descripción de los principales marcadores moleculares. ....	8
<b>Tabla 2.</b> Parámetros principales que controlan STACKS .....	26
<b>Tabla 3.</b> Descripción de las 19 variables Worldclim.....	33
<b>Tabla 4.</b> Lista de los grupo genético asignados por <i>bicluster</i> y sitios de colecta .....	36
<b>Tabla 5.</b> Descripción de la diversidad genética de los grupos genéticos.....	37
<b>Tabla 6.</b> Estadísticos de divergencia genética ( $F_{ST}$ ) entre los grupos genéticos .....	37
<b>Tabla 7.</b> Descripción de la diversidad genética de los sitios de colecta.....	43
<b>Tabla 8.</b> Estadísticos de divergencia genética ( $F_{ST}$ ) entre los sitios de muestreo .....	43
<b>Tabla 9.</b> VIF de las variables climáticas y porcentaje de contribución cada variable.....	44
<b>Tabla 10.</b> Coeficientes de correlación entre las matrices. ....	56

## Resumen

*Vanilla planifolia* es una orquídea considerada un recurso fitogenético cuyo centro de origen es México; donde sus poblaciones silvestres han disminuido o desaparecido en gran parte de su área de distribución, debido a la fragmentación o desaparición de su hábitat. Aunado a esto, diversos problemas meteorológicos, así como bajos rendimientos, irregularidad de la siembra y altos costos de producción, han provocado que su producción se detuviera casi por completo. En este estudio, se analizó la variación genómica y la distribución actual y futura de la especie considerando una colecta de los ejemplares en los estados de San Luis Potosí, Hidalgo, Oaxaca, Veracruz y Quintana Roo. Para el análisis de la variación genómica, se identificaron 252 SNPs mediante el mapeo al borrador del genoma de *V. planifolia*. Para la distribución actual se consideraron dos conjuntos de variables climáticas (WorldClim y CONABIO), seleccionando las variables que más aportaban al modelo. Para la distribución futura, se consideraron dos escenarios contrastantes de cambio climático, RCP2.6 y RCP8.5, ambos para el 2050 y 2070, utilizando proyecciones de WorldClim. Se encontró una diversidad genética media ( $H_o$  y  $H_e \sim 0.2$ ), bajos niveles de diferenciación genética ( $F_{ST}=0.03652$ ) y niveles cercanos a cero de endogamia ( $F_{IS}= -0.0326$ ). Además, el análisis de agrupamiento Bayesiano (BIC) así como uno de validación cruzada (Admixture) muestran un  $K=1$ , lo que significa que la especie no está estructurada genéticamente. Los resultados de la distribución actual muestran que la especie presenta una distribución restringida en parches de hábitat remanentes, con alta especificidad hacia las variables ambientales: Temperatura media del trimestre más seco, Precipitación del mes más húmedo, Régimen de humedad, Vegetación y Precipitación. Asimismo, las proyecciones futuras muestran una modificación en el hábitat que probablemente pueda deberse a las modificaciones en temperatura y precipitación originadas por el aumento de  $CO_2$ .

Palabras clave: *Vanilla planifolia*; genómica de poblaciones; SNPs; distribución actual y futura; cambio climático.

## Abstract

*Vanilla planifolia* is an orchid considered a fitogenetic resource whose center of origin is Mexico; where their wild populations have decreased or disappeared for a large part of their distribution area, due to the fragmentation or disappearance of their habitat. In addition to this, various weather problems, as well as low yields, irregular planting and high production costs, have caused its production to stop almost completely. In this study, the genomic variation and the current and future distribution of the species, were analyzed considering a sampling collection of the specimens in the states of San Luis Potosí, Hidalgo, Oaxaca, Veracruz and Quintana Roo. For the analysis of genomic variation, 252 SNPs were identified by mapping them on the *V. planifolia* draft genome. For the current distribution, two sets of climatic variables (WorldClim and CONABIO) were considered, selecting the variables that contributed the most to the model. For future distribution, two contrasting climate change situations, RCP2.6 and RCP8.5, both for 2050 and 2070, were considered using WorldClim projections. An average genetic diversity ( $H_o$  and  $H_e \sim 0.2$ ), low levels of genetic differentiation ( $F_{ST} = 0.03652$ ) and near zero levels of inbreeding ( $F_{IS} = -0.0326$ ) were found. In addition, the Bayesian grouping analysis (BIC) as well as a cross-validation analysis (Admixture) shows a  $K = 1$ , meaning that the species is has not a genetic structure. The results of the current distribution suggest that the species presents a restricted distribution in patches of remaining habitat, with high specificity towards the following environmental variables: Average temperature of the driest quarter, Precipitation of the wettest month, Humidity regime, Vegetation and Precipitation. Likewise, future projections show a modification in the habitat that could probably be due to the changes in temperature and precipitation caused by the increase in  $CO_2$ .

Keywords: *Vanilla planifolia*; population genomics; SNPs; Current and future distribution; climate change.

## 1. Introducción

Se considera a la vainilla (*Vanilla*) un género que presenta una distribución pantropical entre latitudes aproximadas a 27 °N y 27 °S (Hu *et al.*, 2019). Algunas especies son cultivadas debido al prestigio que ha adquirido su aromático fruto, del cual se genera un extracto utilizado en múltiples industrias, como: la farmacéutica, la cosmética, la refresquera, entre otras; utilizado principalmente en la elaboración de alimentos (Zamora *et al.*, 2016). Actualmente, es la segunda especia más valiosa en el mercado debido a la creciente demanda mundial (Hu *et al.*, 2019).

La vainilla tuvo gran importancia en la época prehispánica, siendo un condimento de diversos productos alimenticios, incluido su uso para mejorar el sabor del chocolate de los aztecas y mayas, así como un tributo en la época de la conquista europea, además de formar parte de una larga tradición cultural de México (Salazar-Rojas, *et al.*, 2011; Rosales, *et al.*, 2013). Su propagación a nivel mundial comenzó desde Mesoamérica a finales de los años 1500 y su introducción en tierras europeas en 1739 (Fouche y Jouve, 1998).

El cultivo de vainilla presenta una problemática compleja, derivada principalmente de factores relacionados con su cultivo, uso y conservación: 1) Su cultivo formal es reciente (alrededor de 200 años), lo que se traduce en un material con pocas generaciones de selección que no presenta expresión de caracteres asociados a domesticación (Soto-Arenas, 2006; Rocha *et al.*, 2018). 2) La sobreexplotación de poblaciones naturales y subutilización de poblaciones cultivadas (Soto-Arenas, 2006; Lubinsky *et al.*, 2008; Salazar-Rojas, 2011). Y 3) Problemas que afectan directamente en el cultivo, como caída prematura del fruto, nutrición deficiente, presencia de plagas y enfermedades, así como bajos rendimientos, irregularidad de la siembra y altos costos de producción. Esto condujo a una drástica disminución del porcentaje de contribución de México en la producción a nivel mundial, donde actualmente produce menos del 3% (Castillo y Engleman, 1993; Soto-Arenas, 2006; FAO, 2011).

*Vanilla planifolia* es la especie principal para la obtención del saborizante, y debido a su importancia cultural, agrícola, comercial y a su creciente demanda, existe un esfuerzo a nivel global para caracterizar su diversidad (Bory *et al.*, 2007). Trabajos previos en diferentes países, han tratado de evaluar y caracterizar la diversidad genética de vainilla utilizando diversos



marcadores moleculares, encontrando que la diversidad genética en *V. planifolia* es limitada (Nielsen y Siegismund, 1999; Cameron, 2004; Schlüter *et al.*, 2007; Soto-Arenas y Dressler, 2010). Esto probablemente se debe a la facilidad de propagación vegetativa y la distribución mundial de algunos clones fundadores (Hu *et al.*, 2019). Estos estudios incluyen diferentes marcadores moleculares como: isoenzimas, RAPD (*Random Amplification of Polymorphic DNA*; por sus siglas en inglés), AFLP (*Amplified fragment length polymorphism*; por sus siglas en inglés), microsatélites y secuencias de un solo gen. Las limitaciones que se han encontrado en estos enfoques incluyen la falta de reproducibilidad, alto costo relativo y baja densidad de información. Con el uso y desarrollo de plataformas orientadas al análisis genómico se podrían superar estas limitaciones ya que se podría analizar una cantidad de marcadores moleculares mucho mayor, además de conseguir una reducción del genoma de los organismos (Eguiarte *et al.*, 2013).

México es el centro de origen de *V. planifolia*, por lo que explorar y conocer los niveles de variación infraespecífica a través de la genómica de poblaciones, podría ser una gran herramienta para elaborar programas de fitomejoramiento que brinde beneficios a los cultivos y también apoye los programas de conservación de la especie (Eguiarte *et al.*, 2013; Herrera *et al.*, 2016; Hu *et al.*, 2019). Además, mejorar la diversidad genética y el rendimiento hortícola de *V. planifolia* es cada vez más importante a medida que aumenta su demanda.

## **2. Marco Teórico**

### **2.1 Variación biológica**

El concepto de variación biológica se refiere a las diferencias en las características de organismos del mismo tipo o clase, causados por el ambiente o por diferencias en la constitución genética del organismo, es decir, la diversidad presente en absolutamente todas las entidades vivas y en todas las diferencias individuales. La variación es una característica de la naturaleza, la cual es inherente a todos los seres vivos, y se identifica mediante la modificación genética, estructural, morfológica, geográfica o fisiológica de los caracteres que constituyen a los organismos de una población y especie (Mimura *et al.*, 2017; Lima-Morales, 2017; Noguera y Hernández, 2009; Gould, 2002).

Las variaciones fenotípicas y genéticas dentro y entre las poblaciones pueden tener efectos dramáticos en los procesos ecológicos y evolutivos, incluidas las respuestas al cambio ambiental, el mantenimiento de la diversidad de especies y la estabilidad y resistencia ecológicas. Al mismo tiempo, los cambios ambientales asociados con muchas actividades humanas, como el uso de la tierra y el cambio climático, tienen un impacto dramático y a menudo negativo en la variación (Mimura *et al.*, 2017; Hallgrímsson, 2005). Mimura *et al.* (2017) sugieren un monitoreo periódico de la variación biológica, que debe incluir al menos dos estrategias principales: 1) monitoreo intensivo de variación genética y 2) modelado de especies para predecir cambios en la variación en función de los cambios en el tamaño de la población y extensión del rango.

### **2.2 Variación genética**

La diversidad genética es la variación del material genético heredable que ocurre dentro y/o entre organismos, poblaciones o especies. Esta variabilidad se origina por mutaciones, recombinaciones y alteraciones en el cariotipo (el número, forma, tamaño y ordenación interna de los cromosomas). Siendo la selección natural y la deriva genética los procesos que dirigen o eliminan esta variabilidad genética (Piñero *et al.*, 2008).

La selección natural es el proceso evolutivo que Darwin propone en su libro el origen de las especies (Futuyma 2013), donde esta fuerza genera una reproducción y supervivencia diferencial en los individuos de una población. La formulación clásica de la selección natural establece que las condiciones de un ambiente favorecen o dificultan la tasa reproductora o la eficacia biológica en relación con características fenotípicas heredables.

Se han descrito diferentes tipos de selección, que se detectan por los cambios en la frecuencia de los caracteres heredables entre generaciones (Figura 1). Esencialmente se trata de tres tipos cuya diferencia depende de la influencia sobre la media y la varianza del carácter bajo selección (Hedrick, 2011), los cuales se describen a continuación: 1) Selección estabilizadora, en la que se produce cuando individuos con un cierto rasgo han tenido éxito con respecto a la gran mayoría de los demás individuos que conforman la población donde se mantiene a la población en un óptimo estable, 2) Selección direccional, que transforma el valor de un rasgo mediante el incremento de la frecuencia de los individuos cercanos a una distancia óptima. 3) Selección disruptiva que incrementa la frecuencia de valores mayores y menores a costa de valores intermedios.

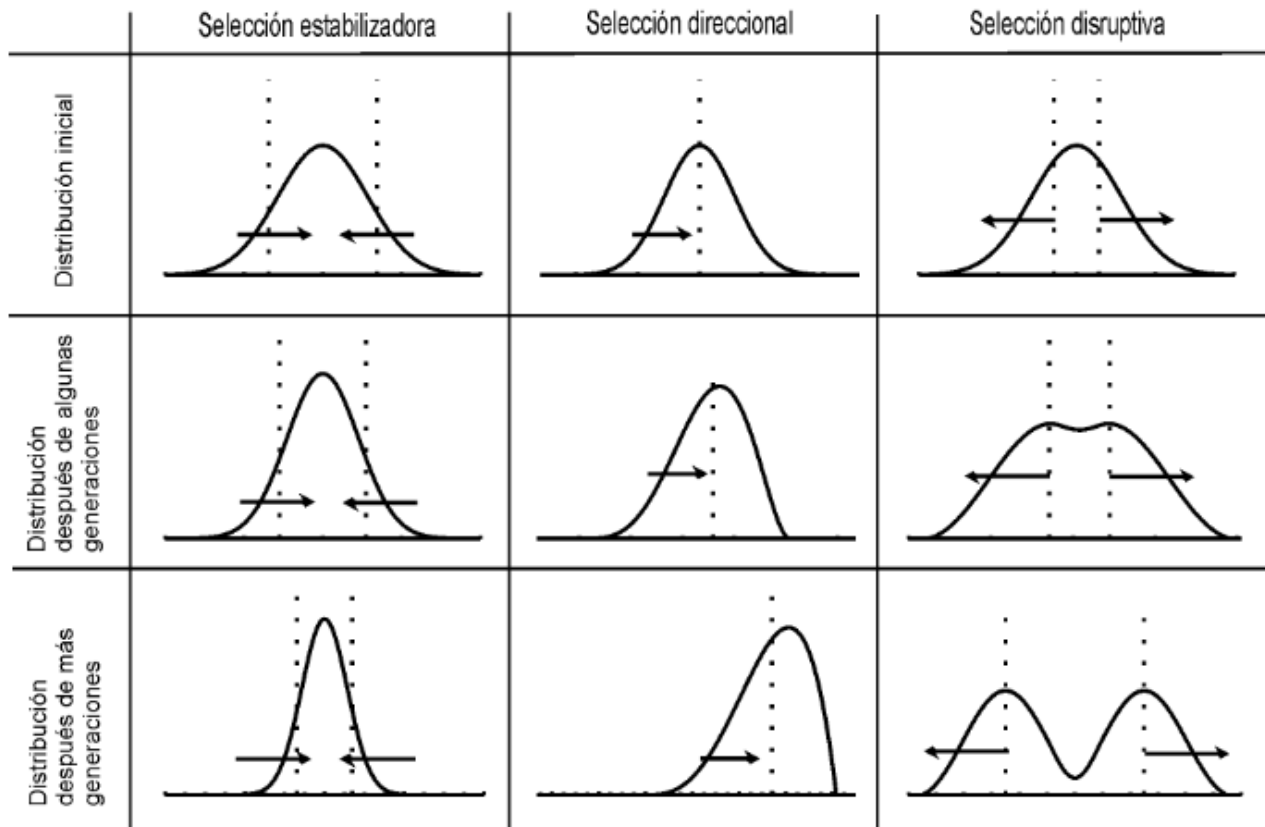


Figura 1. Tipos de selección donde la diferencia depende de la influencia sobre la media. Tomado de Soler, 2002.

Por otro lado, la deriva genética (descrita por Wrigth en 1943) es un efecto estocástico y se refiere al cambio aleatorio de la frecuencia de los genes. La deriva genética tiene efectos importantes cuando una población reduce dramáticamente su tamaño debido a un desastre natural (efecto de cuello de botella) o cuando un grupo pequeño se separa de la población principal para formar una colonia (efecto fundador; Hedrick, 2011).

### 2.3 Genética y genómica de poblaciones

La genética de poblaciones tiene como finalidad explorar los niveles de la variación genética dentro y entre las poblaciones a través del tiempo y espacio, asimismo, permite entender los diferentes procesos y fuerzas evolutivas que conducen a esta variación (Templeton, 2006); siendo una herramienta útil para el estudio de la evolución.

Durante el siglo XX, el descubrimiento de la estructura del ácido desoxirribonucleico (ADN) en 1953 trajo consigo el desarrollo posterior de diversas técnicas moleculares (Tabla 1) para el estudio de la variación genética en temas referentes a evolución y domesticación, (Hedrick, 2011; Eguiarte *et al.*, 2013).

Inicialmente la genética de poblaciones contemplaba a un gen para su estudio, explorando el comportamiento de dos alelos que segregan de acuerdo con las leyes de Mendel, utilizando el equilibrio de Hardy-Weinberg como modelo nulo de lo que sucede si no opera ningún proceso evolutivo. La técnica inicial para medir la variabilidad genética fue la electroforesis de isoenzimas que permitía evaluar loci (genes) polimórficos mediante la separación eléctrica por carga y tamaño. Posteriormente se analizaron algunas decenas de genes, y a pesar de ser un modelo predictor seguían sin ser comparables con un posible análisis brindado por el genoma completo (Eguiarte *et al.*, 2013).

Los enfoques que han sido utilizados para conocer la variación genética incluyen isoenzimas, polimorfismos de fragmentos de restricción (RFLP), polimorfismos de ADN amplificados aleatoriamente (RAPD), repetidos de secuencias internas (ISSR), microsatélites y secuencias de ADN (Tabla 1). La elección del marcador más adecuado varía dependiendo el tipo de pregunta y algunas de las características del marcador como tipo de herencia y, además, de la facilidad en su uso y costo (Piñero *et al.*, 2008).

Con los avances recientes en los métodos moleculares y estadísticos es posible reconstruir la historia evolutiva de las poblaciones con un mayor número de marcadores encontrados a lo largo del genoma, potenciando investigaciones ecológicas y de conservación (Eguiarte *et al.*, 2013; Piñero *et al.*, 2008).

Tabla 1. Descripción de los principales marcadores moleculares utilizados en genética de poblaciones. Tomado y modificado de Eguiarte, et al. 2013.

Marcador	Características	Ventajas	Desventajas
Variación Fenotípica	Variables observables y cuantificables	-Se puede observar en las poblaciones -En muchas ocasiones son adaptativos -Comparable con fósiles	-No representan la variación genética -Pueden reflejar plasticidad -Muchas veces son poligénicos
Aloenzimas	Variación en la migración de las enzimas durante la electroforesis, por cambios de aminoácidos.	-Muchos loci son polimórficos -Codominantes (identifica homocigos de heterocigos)	-Poca variación -No detecta mutaciones silenciosas

			-Pocos marcadores a lo largo del genoma -Difíciles de montar
RFLP's	Variación en los fragmentos generados por enzimas de restricción, debido a cambios mutacionales en los sitios de reconocimiento		-Información limitada -Complicados de montar -Dominantes
RAPD's	Variación en la ampliación de fragmentos aleatorios, que depende de la presencia diferencial de sitios de unión (mutaciones y cambios estructurales)	-Muchos sitios polimórficos a lo largo del genoma -Alta variación -Oligonucleótidos universales	-Dominantes -Cada sitio poco informativo -Se desconocen sitios amplificados -Muchos artefactos -Difíciles de leer
ISSR's	Similar a los RAPD's pero los oligonucleótidos son secuencias más complejas y largas en tandem (microsatélites)	-Muchos sitios polimórficos a lo largo del genoma -Alta variación -Oligonucleótidos universales -Más reproducibles que los RAPD's	-Dominantes -Cada sitio poco informativo -Se desconocen sitios amplificados -Difíciles de montar -Difíciles de leer
AFLP's	Variación en la ampliación de fragmentos cortados con enzimas de restricción y ligados con adaptadores	-Genera muchos fragmentos -Alta variación -Muy reproducibles -Oligonucleótidos universales -Alta representación del genoma	-Muy complicados de montar -Dominantes -Cada sitio es poco informativo -Difíciles de leer
Microsatélites	Variación en el número de secuencias repetidas en tandem de ADN por inserción o pérdida de motivos	-Muy polimórficos -Codominantes -Se conoce la región donde se encuentran -En muchos casos se conoce la tasa de mutación	-Difíciles de montar -Modelo de mutación complicado -Error en la estimación de tasas de Mutación conduce a resultados erróneos -Alta homoplastia -Alelos nulos
SNP's	Polimorfismos de un solo nucleótido. Mutaciones puntuales en secuencias de ADN. Se obtienen por medio de secuenciación, RAD-seq, GBS o micro-arreglos	-Existen miles en el genoma -Permite entender la evolución a nivel genómico -Nuevas herramientas de secuenciación -Permite la obtención de miles de SNP's -En sitios neutrales y codificantes	-Caros -Cada sitio por si solo es poco informativo -Difícil de analizar

## 2.4 Secuenciación de nueva generación

El desarrollo de nuevas tecnologías ha aumentado de manera exponencial. Un claro ejemplo son las denominadas tecnologías de Secuenciación de Nueva Generación (SNG; también llamada secuenciación masiva) que actualmente permite obtener millones de secuencias de ADN a una velocidad y a un costo cada vez más reducido. Estas tecnologías permiten logros

científicos trascendentales, como la identificación de nuevos genes y la resolución de las bases genéticas de diferentes enfermedades (Rodríguez y Armengol, 2012).

En las últimas dos décadas, se ha incrementado considerablemente la disponibilidad de recursos genómicos de secuenciación de nueva generación en temas como, análisis de la expresión de genes, número de marcadores moleculares, análisis de transcriptomas, estudios de asociación genómica, entre otros. Específicamente, en los últimos cinco años, se han desarrollado métodos y técnicas, que permiten el análisis de una cantidad de marcadores moleculares mucho mayor. La secuenciación masiva puede ser aplicada para conseguir una reducción del genoma de los organismos y obtener un número elevado de variantes genómicas en múltiples individuos. La reducción del genoma se consigue mediante el empleo de enzimas de restricción, que reconocen un patrón específico de bases en las secuencias de ADN y fragmentan en sitios específicos. Tal es el caso de la Genotipificación por Secuenciación (GBS por sus siglas en inglés) o RAD-Seq (derivado de *Restriction site Associated DNA*), técnicas con base en la SNG, que permite la identificación de cientos a miles de marcadores moleculares del tipo “polimorfismos de un solo nucleótido” (SNP, por sus siglas en inglés). Estas permiten rastrear el genoma completo en la búsqueda de polimorfismos, es decir, tanto regiones neutrales como regiones que se expresan, lo cual permite un estudio integral de las variaciones propias de una especie, así como la capacidad de adaptación a diferentes condiciones bióticas y abióticas (Elshire *et al.*, 2011; Glaubitz *et al.*, 2014).

El método GBS (Genotipificación por Secuenciación) o SBG (Genotipificación Basada en Secuenciación) genera pequeños fragmentos de ADN adyacentes a cada sitio de restricción, lo cual es útil y más costeable para el estudio de organismos eucariotas con tamaños de genoma de muchas megabases o gigabases (Patel *et al.*, 2015).

## **2.5 Domesticación y recursos fitogenéticos**

Uno de los casos más evidentes de variación genética son las especies domesticadas en donde se utiliza la variabilidad para crear razas y variedades a partir de un manejo agrícola que incluye mecanismos de manipulación de los genotipos de las plantas por el hombre, siendo dicha manipulación la actividad cultural que permite al hombre adecuar la diversidad biológica

a las necesidades de la sociedad humana, y, por lo tanto, a las condiciones artificiales de manejo (Díaz, 2010).

México es un importante centro de diversidad y de domesticación de numerosos cultivos, algunos de gran importancia global. Se han reportado cerca de 120 especies de plantas útiles en la agricultura, textiles o como colorantes (Mapes y Basurto, 2016). Esta diversidad está determinada no solamente como resultado de eventos biológicos, si no por la interacción entre poblaciones humanas y estos organismos a través de cientos de generaciones. Estas poblaciones humanas, a raíz de sus necesidades, intereses, prácticas y conocimientos han ido conformando y manteniendo esta diversidad. Estos organismos vegetales que contienen unidades funcionales de la herencia, y que tiene valor real o potencial para la alimentación y la agricultura son llamados recursos fitogenéticos (FAO, 2011).

La vainilla es un recurso fitogenético que presenta una problemática compleja derivada principalmente de factores relacionados con el uso y conservación de sus poblaciones naturales y cultivadas (Lubinsky, 2008; Soto-Arenas, 2006; Salazar-Rojas, 2011). Por lo que se requiere un análisis integral que resulte en una mejor comprensión de la variación intra e interpoblacional y así poder evaluar la amenaza genética para la especie (Díaz, 2010; Eguiarte *et al.*, 2013).

## **2.6 *Vanilla planifolia***

El género *Vanilla* abarca un grupo de orquídeas hemi-epífitas con origen en el continente americano, el cual se conforma aproximadamente de 110 spp. Las condiciones para el crecimiento de estos organismos son bastante restringidas, ya que requiere temperaturas entre 20 y 30°C, 50% de sombra, alta humedad, una distribución altitudinal no mayor de 600-800 msnm y un tutor para su crecimiento (Zotz y Schmidt, 2006; Rosales *et al.*, 2013).

Las raíces son delgadas y alcanzan poca profundidad en el material orgánico que se encuentra en el suelo, además de contar con raíces aéreas que les permiten adherirse y absorber nutrientes desde la corteza del tutor donde se apoya. El tallo principal es delgado y flexible el cual puede alcanzar diez metros de longitud. Las hojas, generalmente, son de color verde oscuro, planas y en algunas especies presentan cierta succulencia. Las inflorescencias crecen a manera de ramillete de forma sucesiva, estas son efímeras (solo abren un par de



horas), grandes y atractivas con coloraciones blancas, verdes, amarillas o color crema y de corola infundibuliforme. Los frutos consisten en una cápsula alargada (12-30 cm de longitud; comúnmente llamada “vaina”) carnosas, las cuales contienen miles de pequeñas semillas. A medida que va madurando, el fruto se vuelve color negro y desprende un fuerte y atractivo aroma (Rosales *et al.*, 2013).

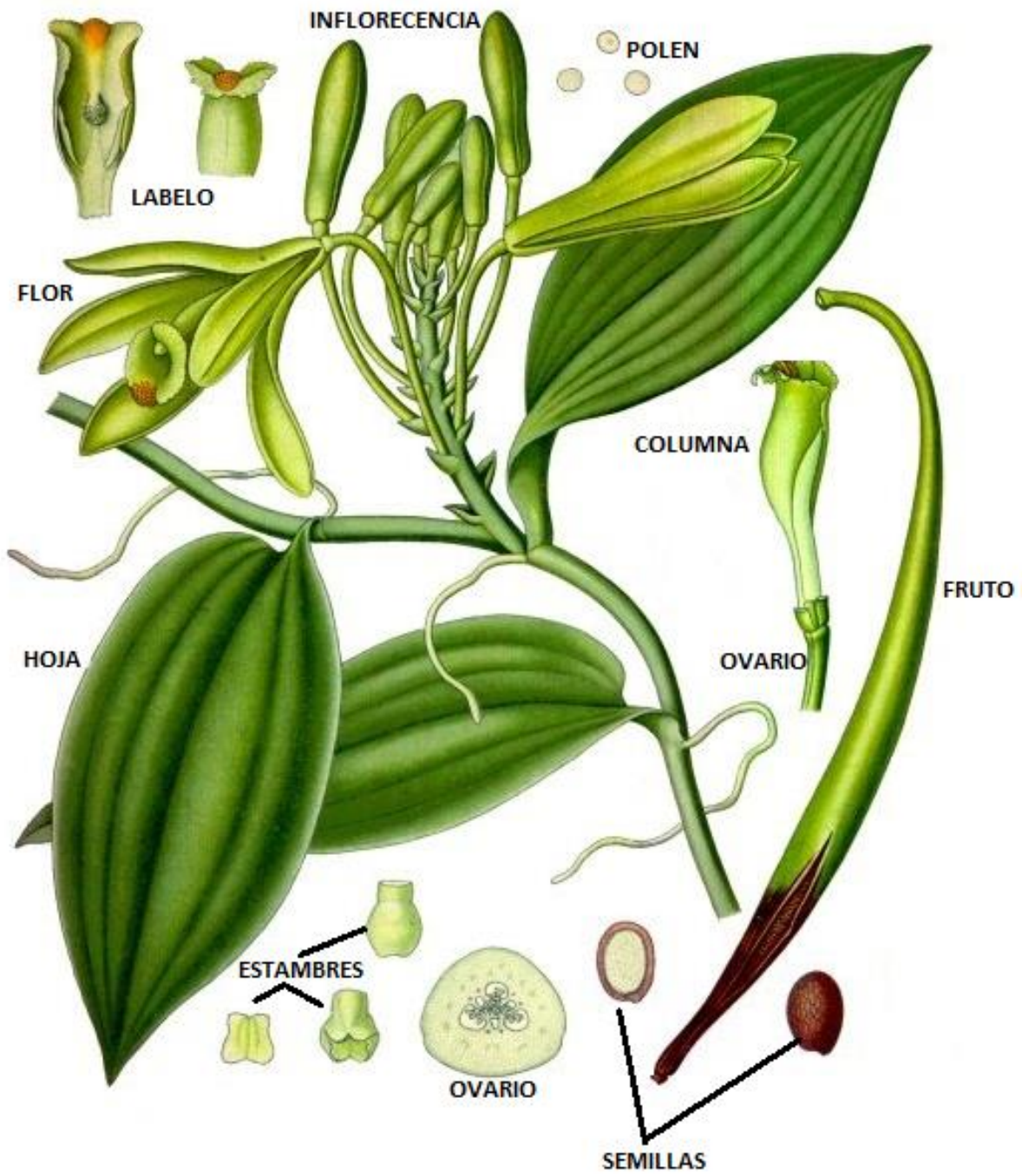


Figura 2. Esquema de las estructuras de *Vanilla planifolia*. Modificada de Kohler, 1890

Dentro de este género podemos encontrar algunas especies representativas como *V. pompona*, *V. insignis*, *V. odorata* y *V. planifolia*. De estas, *V. pompona* y *V. planifolia* producen un alto contenido de vainillina, compuesto utilizado en diferentes industrias debido a su aroma agradable.

*V. planifolia* (Fig. 2) es la especie principal para la obtención del saborizante. Es una orquídea herbácea, perenne, hemiepífita y enredadera que trepa sobre árboles hasta una altura aproximada de 10 m (Soto-Arenas y Dressler, 2010). A pesar de su importancia, se considera que es una especie no domesticada y que se encuentra en este proceso debido a que los organismos cultivados son similares a sus parientes silvestres, donde solo existe una selección empírica por parte de los productores (Flores *et al.*, 2017)

En México, se han localizado individuos silvestres de *V. planifolia* en Oaxaca, Yucatán, Quintana Roo, Tabasco y zonas limítrofes de Veracruz, Chiapas y San Luis Potosí (Fig. 3; Soto y Dressler, 2010). De la producción nacional de vainilla, el 90% se obtiene en la región de Totonacapan, el resto se produce en Hidalgo, Oaxaca, Tabasco y Quintana Roo (Morales *et al.*, 2001).

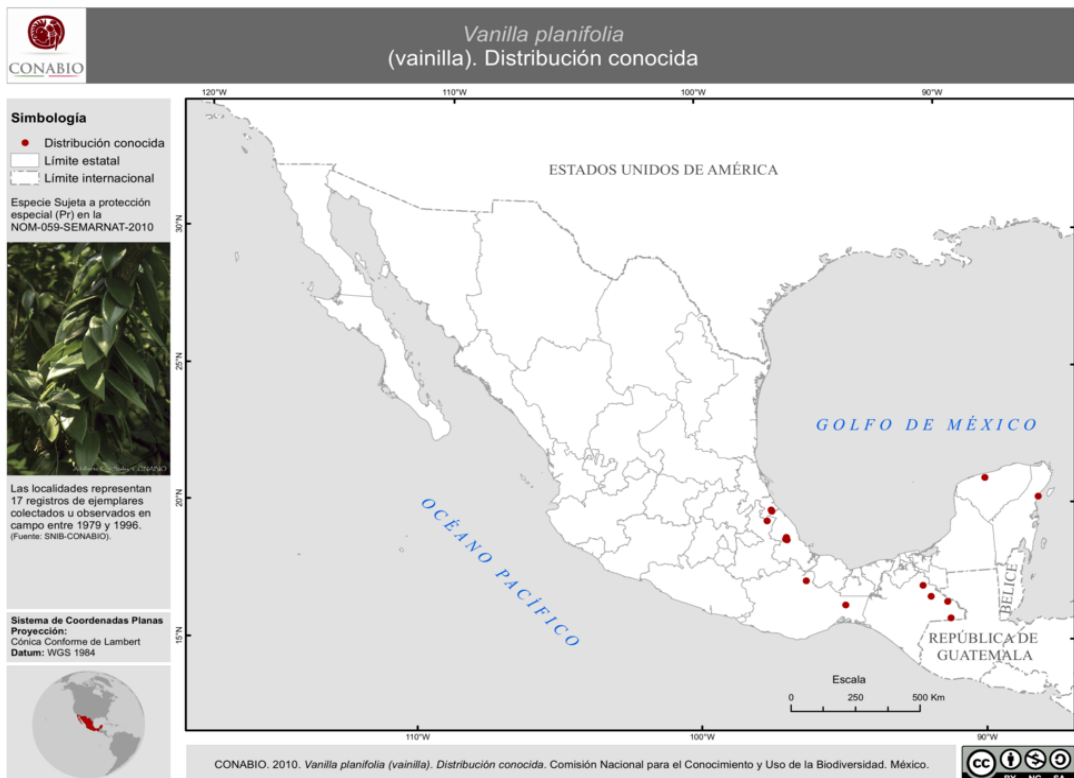


Figura 3. Distribución conocida de *Vanilla planifolia* en México ([www.conabio.gob.mx](http://www.conabio.gob.mx))

La vainilla tiene una importancia económica, social-cultural y ecológica. Dentro de los ámbitos económicos, la vainilla es la segunda especie más cara del mundo después del azafrán y es considerada la sustancia aromatizante y saborizante más importante (Osorio, 2012; Hu *et al.*, 2019). Ecológicamente, esta especie tiene la posibilidad de ser establecida en bosques y zonas de reserva forestal, contribuyendo así a la conservación y valorización de éstos y otros recursos naturales (Osorio, 2012). Culturalmente, la vainilla tuvo gran importancia en la época prehispánica, siendo un condimento de diversos productos alimenticios, así como un tributo en la época de la conquista europea, además de formar parte de una larga tradición cultural de México. Actualmente, su cultivo es una alternativa de agricultura sustentable que beneficia a pequeños productores ya que las plantaciones pueden ser establecidas en sistemas de traspatio (Azofeifa *et al.*, 2014).

Aunado a la problemática del cultivo de *V. planifolia*, sus poblaciones silvestres han disminuido o incluso desaparecido en gran parte de su área de distribución debido a la

fragmentación o desaparición de su hábitat, lo que provoca que el acervo genético se vea gravemente amenazado o incluso extinto para algunos autores (Soto-Arenas, 1999; Lubinsky *et al.*, 2008). Los especímenes silvestres y los cultivares de *V. planifolia* conforman este acervo genético (Soto-Arenas, 1999). Por lo que, al ser las plantaciones comerciales de vainilla establecidas mediante una propagación vegetativa (Ecott, 2004; Bory *et al.*, 2007), la variación genética de la especie ha sido sometida a factores de erosión (Mino *et al.*, 2006), lo que ha provocado que *V. planifolia* sea catalogada en la NOM-059-SEMARNAT-2010 como una especie bajo protección especial, en peligro por la IUCN (International Union for Conservation of Nature) y que se encuentre en el Apéndice II (especies potencialmente amenazadas) de la Convención Internacional de Especies Amenazadas de Fauna y Flora Silvestres (CITES).

#### 2.6.1 Diversidad en *V. planifolia* (Antecedentes)

Cibran (2000), analizó la diversidad mediante el uso de isoenzimas en poblaciones cultivadas y silvestres. Encontró dos grupos, las plantas del sur y las del norte de México, así como diferencias entre los organismos silvestres y los cultivados. Posteriormente, Schluter *et al.* (2007), analizaron la variación genética de las poblaciones de *V. planifolia* mediante el uso de marcadores tipo RAPDs, donde un análisis multivariado reveló tres grupos geográficos: a) un grupo costarricense b) un grupo mexicano al norte del eje volcánico Transmexicano y c) grupo mexicano al sur del eje volcánico Transmexicano, apoyando la separación mencionada por Cibran.

Un caso particular es el descrito por Salazar-Rojas *et al.* (2011), quienes revelaron la existencia de variación quimiotípica entre los especímenes cultivados de *V. planifolia*. Encontraron que los organismos analizados se agruparon en seis quimiotipos (Fig. 4A), que indirectamente sugieren la existencia de polimorfismo genético en la región de Totonacapan. Posterior a esto, Herrera-Cabrera *et al.*, (2012) mediante el uso de microsatélites identificaron variación a nivel infraespecífico en *V. planifolia*, donde 13 loci resultaron polimórficos para los quimiotipos de *V. planifolia* definiendo cinco genotipos cultivados en la región Totonacapan (Fig. 4B).

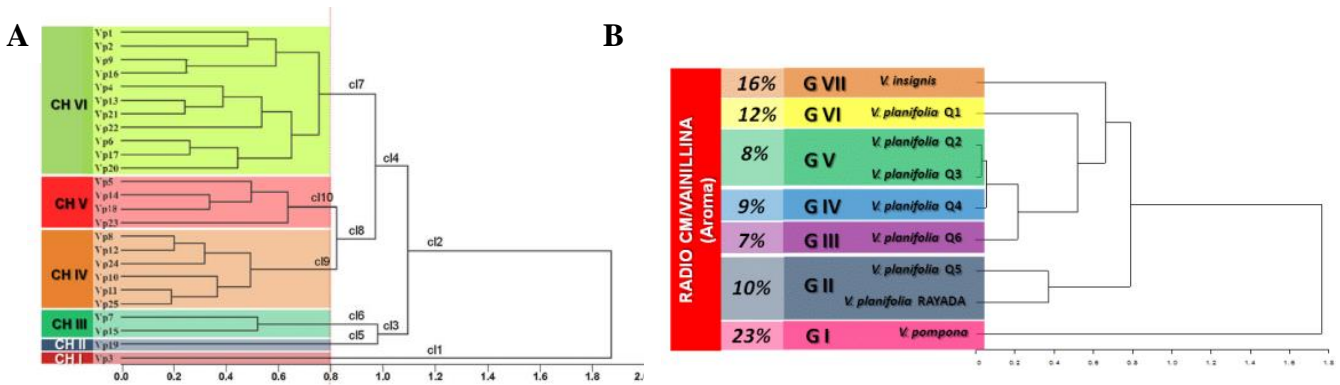


Figura 4. A) Dendrograma de 25 accesiones de *Vanilla planifolia* en la región de Totonacapan Puebla-Veracruz, México, basado en promedios de 10 variables y agrupación por distancias de similitud. Quimiotipos CH I-CH VI B) Dendrograma de colectas de *V. planifolia* en la región Totonacapan Puebla-Veracruz, con base en el promedio de 13 loci microsatélites y agrupamiento por distancias de similitud donde G (I-VI) son los genotipos. Tomado de Salazar-Rojas 2011 y Herrera-Cabrera et al. 2012.

En un trabajo reciente, Villanueva-Viramontes *et al.* (2017), se plantearon la posibilidad de discriminar a los individuos silvestres de *V. planifolia* de otras especies simpátricas del género. Analizaron 88 individuos silvestres de varias especies de vainilla mediante 84 loci ISSR utilizando tres métodos de agrupamiento. Encontraron que los marcadores ISSR discriminan claramente la *V. planifolia* silvestre, encontrando una estructura genética definida dentro de la especie. Se encontraron cinco grupos que correspondieron con las especies estudiadas (Fig. 5).

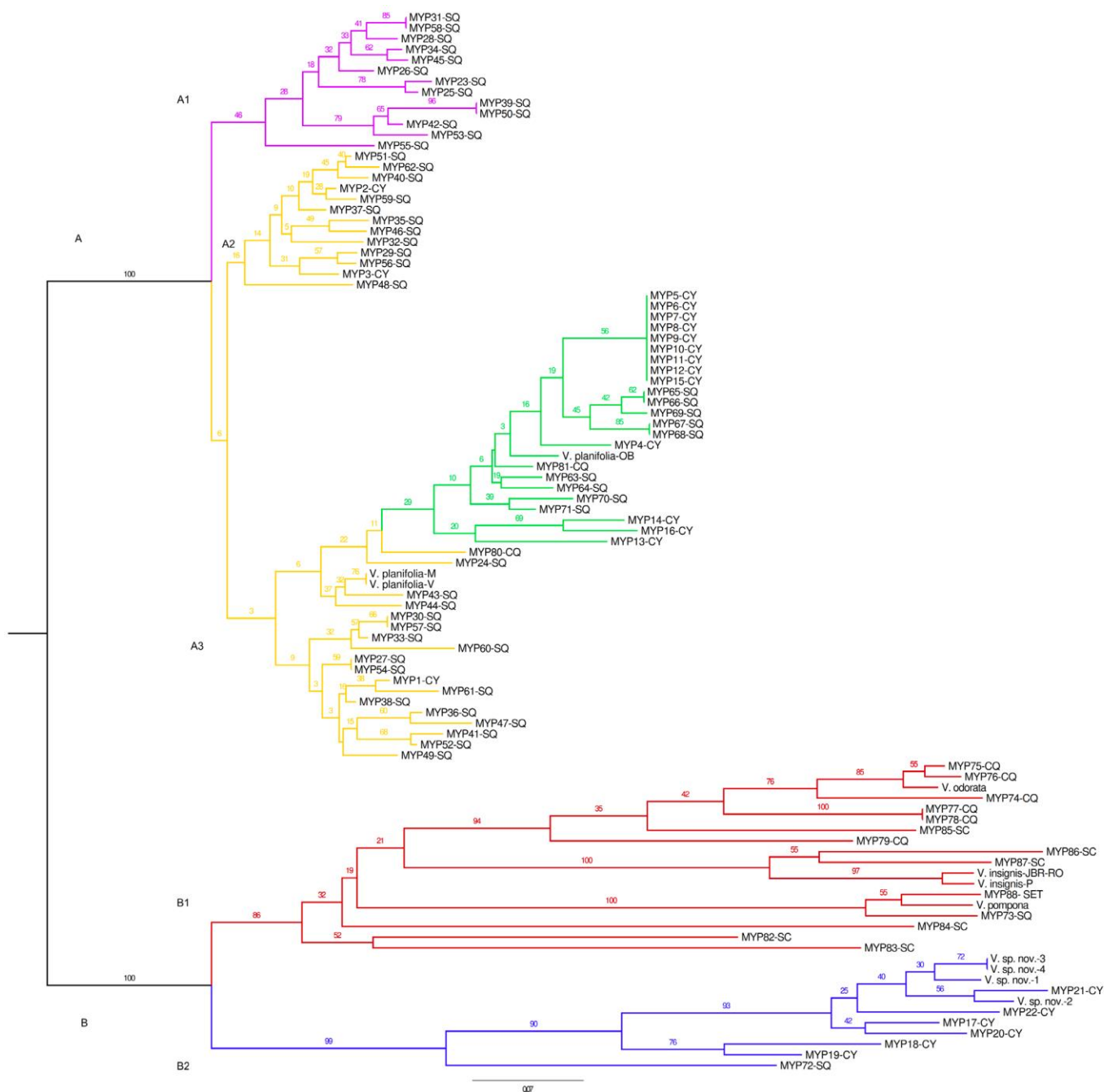


Figura 5. Dendrograma Neighbor-Joining basado en el índice de similitud de Jaccard de 88 individuos silvestres de la península de Yucatán en México y 11 muestras estándar de vainilla, utilizando marcadores ISSR. Tomado de Villanueva-Viramontes (2017)

Más recientemente, Hu *et al.* (2019), analizaron la diversidad entre diferentes individuos del género *Vanilla*. Con la finalidad de proporcionar un recurso para analizar otras colecciones de vainilla, ensamblaron un borrador del genoma de *V. planifolia* que fue utilizado para



identificar marcadores tipo SNPs usando SBG, el análisis de componentes principales de sus resultados, agrupó los individuos dentro de distintos grupos que representan las especies *V. pompona*, *V. planifolia*, *V. imperialis*, *V. odorata*, *V. palmarum*, *V. barbellata*, *V. dilloniana*, *V. phaeantha* y *V. mexicana*, validando la eficiencia de las herramientas basadas en la genómica para caracterizar e identificar la diversidad genética en el género *Vanilla*.

En estudios basados en caracteres morfológicos de la flor (labelo), Castillo y Engleman (1993), analizaron dos variedades de *V. planifolia*, encontrando características morfológicas y fenológicas que permiten separar claramente en dos tipos diferentes. Asimismo, Lima-Morales (2017) analizó el tamaño de las estructuras vegetativas y reproductivas de *V. planifolia*. Esta especie presentó cinco morfotipos, con formas redondeadas, globulares y alargadas, que se diferenciaron por los lóbulos basales, apicales y ángulos internos del labelo. Se encontró variación morfológica, y tanto los gradientes ambientales como la intensidad de manejo no influyeron en esta, por lo que es posible que su variación sea genética.

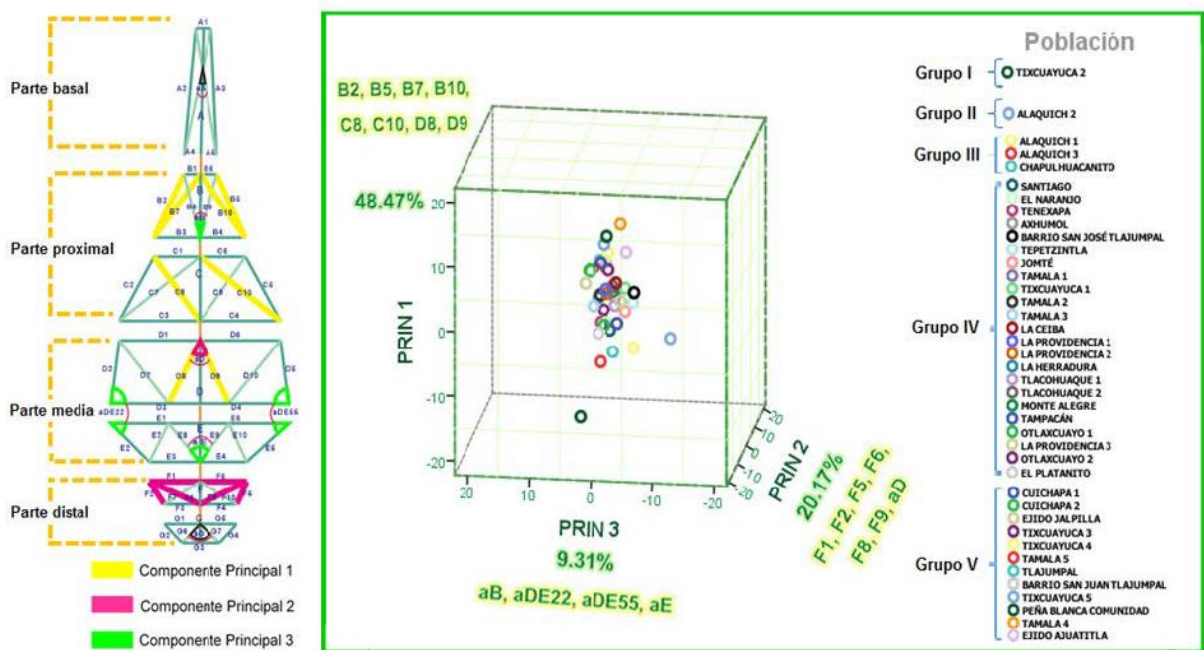


Figura 6. Dispersión de los 679 especímenes pertenecientes a 40 poblaciones de *Vanilla planifolia* en la Huasteca Potosina, con base en los tres primeros componentes principales del análisis de 76 variables basadas en la morfología del labelo. Tomado de Lima-Morales 2017.

En un trabajo reciente (Martinez *et al.*, 2016), se determinó que existe variación anatómica en la epidermis foliar y caulinar de dos genotipos de *V. planifolia* y qué caracteres



apoyan al reconocimiento de estos o cuales están genéticamente establecidos y son propios de la especie. A través de preparaciones de lámina y tallo, se determinó que existen diferencias significativas entre genotipos y que existe una correlación positiva entre el tamaño de ciertos parámetros epidérmicos y la altitud del origen geográfico de los genotipos.

## **2.7 Variación geográfica**

Además de los estudios genéticos, otro acercamiento para la conservación ha sido mediante el uso de modelos de distribución de especies (SDM). Éstos, cuantifican la correlación entre los factores ambientales y la distribución de especies. El perfil ambiental puede ser usado para describir y medir la importancia de factores específicos y predecir la distribución de especies a través de áreas no muestreadas (Miller, 2010). Existen diferentes algoritmos de modelación de la distribución geográfica y potencial, uno de ellos es máxima entropía (*maximum entropy*; *MaxEnt*) el cual ha resultado efectivo en la proyección espacial de la distribución, especialmente cuando se cuenta con pocos datos presenciales (Steven *et al.*, [http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/); Miller, 2010; Pliscoff y Fuentes-Castillo, 2011). Un análisis indirecto de la diversidad en diferentes organismos ha sido mediante las características geográficas y ambientales que determinan la distribución de las especies (Miller, 2010).

La producción de vainilla presentó un aumento considerable en el mundo, durante la década de 1990 (Lubinsky *et al.*, 2008), lo cual generó una extensa recolección de plantas para lograr el establecimiento de cultivos. Además, la destrucción de su hábitat afectó fuertemente a la especie en su entorno natural (Herrera *et al.*, 2016). Por lo que, un conocimiento detallado de la distribución actual y del impacto potencial del cambio climático sobre los nichos de distribución geográfica de la especie es un requisito previo para utilizar y conservar el recurso genético dentro de un ecosistema (Wan *et al.*, 2014).

Ejemplo de esto es el trabajo realizado por Hernández *et al.* (2016), quienes describen al régimen de humedad y precipitación como las variables que presentan mayor contribución al modelo de distribución potencial de *V. planifolia*. Además, Flores *et al.* (2017) analizan el perfil bioclimático de las especies del género *Vanilla* en México.

## **2.8 Unidades Evolutivas Significativas**

Una de las grandes problemáticas para el desarrollo de trabajos en el ámbito de la conservación es el definir los límites de una especie o de una población (Fazekas, *et al.*, 2009). De hecho, las clasificaciones erróneas pueden llevar a posibles resultados negativos para las estrategias de conservación, así como, repercusiones financieras, legales y biológicas (Hey *et al.*, 2003). Esto se complica aún más cuando la morfología está influenciada por la plasticidad fenotípica, la hibridación o la variación infraespecífica (Fazekas *et al.*, 2009). Una propuesta para la solución a estos problemas es el establecimiento de unidades de conservación, como lo son las unidades evolutivamente significativas (UES), lo que puede ayudarnos a tomar mejores decisiones para la conservación, incorporando la diversidad genética, además de la información biológica y morfológica (Ahrens *et al.*, 2017).

El uso de UES proporciona una clasificación alternativa de entidades biológicas (p.e., ecotipos, morfotipos, subespecies o poblaciones) y puede ser adecuado para ayudar a la conservación de los procesos ecológicos y evolutivos de una especie (Moritz, 1994; Ahrens *et al.*, 2017).

### **3. Justificación**

Dado que *V. planifolia* es un recurso fitogenético que tiene a México como centro de origen, resulta importante estudiar los niveles de variación biológica para contribuir a una mejor toma de decisiones para la conservación.

Esto puede lograrse mediante el estudio de la variación genómica, basándose en un mayor número de loci que permita aportar más información acerca del nivel de variación y tipo de estructura genética de sus poblaciones. Además, el estudio de la distribución actual y futura de la especie y el conocimiento de las variables ambientales que influyen en su distribución resultarían en una gran herramienta para la posterior elaboración de un programa de mejoramiento que brinde beneficios a los cultivos y también poder definir regiones prioritarias que apoyen los programas de conservación de la especie y de la diversidad del acervo genético.

## **4. Objetivos**

### **4.1 Objetivo general**

- Analizar la variación genómica y caracterizar la distribución actual y futura de *Vanilla planifolia*.

### **4.2 Objetivos particulares**

- Estimar la variación genómica de *V. planifolia*, por medio de la identificación de polimorfismos de un solo nucleótido (SNPs) obtenidos por SBG.
- Analizar los niveles de variación y estructura dentro y entre los sitios de muestreo.
- Modelar la distribución actual y la proyección futura de *V. planifolia*, bajo dos escenarios contrastantes de cambio climático para los años 2050 y 2070.

## **5. Material y método**

### **5.1 Colecta de Material Biológico**

El material biológico utilizado en este trabajo fue colectado en los estados de San Luis Potosí, Hidalgo, Oaxaca, Puebla, Veracruz y Quintana Roo, cuyas características se han descrito en trabajos previos dirigidos por el Dr. Herrera-Cabrera B. Edgar con la participación de sus alumnos Hernández-Silva, 2014; Hernández-Ruiz, 2015; Maceda-Rodríguez, 2015; Lima-Morales, 2017. La colecta se realizó tanto en cultivares, como en poblaciones silvestres. El número de ejemplares colectados dependió del número de plantas disponibles. Se recabó información geográfica del origen de las plantas. El tejido colectado fue etiquetado y almacenado en nitrógeno líquido y transportados al Laboratorio de Bioquímica Molecular (LBM) que pertenece a la Unidad de Biotecnología y Prototipos (UBIPRO) en la Facultad de Estudios Superiores Iztacala, U.N.A.M.

### **5.2 Análisis Molecular**

Las extracciones de ADN de las hojas colectadas se llevaron a cabo siguiendo el protocolo del kit Plant DNeasy (Qiagen™). Se estimó la concentración de ADN en las extracciones por fluorimetría mediante el uso de un analizador Qubit® 3.0 Fluorometer (Life Technologies™), la calidad mediante espectrofotometría Nanodrop™ y la integridad se visualizó mediante electroforesis en gel de agarosa al 0.8% el cual se corrió a 80 volts durante 40 minutos, se le añadió 1µl de Midori green como fluoróforo intercalante ADN. Con el fin de aumentar la concentración de ADN de las muestras, estas fueron liofilizadas y posteriormente resuspendidas en un volumen de 35µl. Las muestras con mayor concentración a 20ng/µl se seleccionaron y se enviaron al laboratorio especializado *Genomics Center* en la Universidad de Minnesota (EUA) para realizar pruebas de digestión con la enzima de restricción que permita la generación de una representación reducida del genoma. La preparación de las bibliotecas para su secuenciación fue realizada por medio de la plataforma Illumina NexSeq550® (Elshire *et al.*, 2011).

Ocho muestras diferentes con una concentración de al menos 20ng/μl en 100μl fueron enviadas al laboratorio especializado para la selección de las enzimas de restricción óptimas que permitieron obtener mejores lecturas de ADN y por lo tanto una mayor cobertura del genoma. Se determinó que las enzimas de restricción BamHI-Nsil, fueron la mejor opción para usar en la construcción de librerías.

La selección de los individuos para el genotipado se basó en la calidad, concentración e integridad del ADN. Se enviaron un total de 58 muestras de la siguiente manera: 5 individuos del estado de Veracruz (VER), 21 de Hidalgo (HID), 21 de San Luis Potosí (SLP), 9 de Oaxaca (OAX) y 2 individuos pertenecientes al estado de Quintana Roo (INS).

### **5.3 Análisis bioinformático (genoma)**

#### **5.3.1 Filtrado e identificación de SNPs**

Los datos sin procesar de la secuenciación (formato fastq), fueron filtrados mediante el software *Fastx-tools*, con base en su calidad utilizando solo aquellos con un valor phred superior a 30 (99.9% de confianza) y con un filtrado para que las secuencias tengan la misma longitud (120 pb). Posterior a esto, se realizó un ensamble *de novo* con las secuencias mediante el software *Stacks 2.0b*, (Rochette y Catchen, 2017; Catchen *et al.*, 2013). Se optimizaron los parámetros (Tabla 2) del ensamblaje en *Stacks* para minimizar el error y maximizar la recuperación de los SNPs (Mastretta-Yanes, *et al.*, 2014; Paris, *et al.*, 2017). Se realizaron ensambles con diferentes modificaciones de algunos parámetros con las muestras secuenciadas, posteriormente se evaluaron los efectos en el modelo de llamado de SNPs sobre las tasas de error, el número de datos perdidos (*missing data*) y el número de SNPs recuperados. Para el ensamblaje los siguientes parámetros clave fueron corridos con los valores especificados en los paréntesis: -m (3, 4, 5, 6), -M (1, 2, 3, 4, 5), -n (1, 2, 3, 4) y -N (1, 2, 3, 4). Modificar los diferentes parámetros nos permite maximizar el número de SNPs recuperados minimizando el error. El parámetro más importante a modificar es el mínimo de cobertura para un *cluster* (-m) ya que afecta directamente el llamado de los marcadores (Paris *et al.*, 2017). Entre más laxo sea este parámetro es más probable que errores los de secuenciación sean tomados como variación alélica, incrementando erróneamente la

estimación de heterocigosis (Reyes-Galindo, 2016). Por lo que, al aumentar este parámetro, el número de SNPs recuperados disminuirá, pero se evitarán llamadas ambiguas de SNPs.

Con los cuatro parámetros optimizados (-m5, -M2, -n1, -N2) se generó el *ensamble de novo* filtrado de los datos arrojados por la secuenciación.

Tabla 2. Parámetros principales que controlan la formación de locus y el polimorfismo en STACKS, los valores predeterminados, el programa componente STACKS que usa el parámetro y una descripción de qué parte de los procesos controla cada parámetro. Tomado y modificado de Paris, et al., 2017.

Parámetro	Valor default	STACKS	Descripción del parámetro
M	3	ustacks	Número mínimo de lecturas sin procesar requeridas para formar un agrupamiento (un alelo putativo)
M	2	ustacks	Número de desajustes ( <i>mismatches</i> ) permitidos entre agrupamientos (alelos putativos) para fusionarlos en un lugar putativo
N	1	cstacks	Número de desajustes permitidos entre los agrupamientos (loci putativo) durante la construcción del catálogo
N	M + 2	ustacks	Número de desajustes permitidos para alinear lecturas secundarias (lecturas que no formaron agrupamientos) con los loci putativos ensamblados para aumentar la profundidad del locus

Se realizó un mapeo al borrador del genoma de *V. planifolia*, disponible en la base de datos del NCBI (*National Center for Biotechnology Information*). Para el mapeo se utilizó el programa *Bowtie 2*, y *Samtools* y *Bcftools* para el llamado de los SNPs. La identificación de SNPs a lo largo del genoma, se determinó mediante el software *Tassel V 5.2.43* (Elshire et al., 2011; Glaubitz et al., 2014).

El filtrado de calidad de los SNPs se realizó principalmente mediante funciones integradas en *VCFtools*, el filtro de frecuencia de alelos menores (MAF) se estableció en 0.01 y la cobertura de locus mínima se estableció para retener los SNP que están cubiertos en al menos el 80% de los individuos. Los SNPs con un valor menor 0.01 del equilibrio H-W fueron descartados. Las inserciones y deleciones fueron removidas del análisis. La tasa de datos perdidos (*missing data*) máxima se estableció en 0%, para tener una mayor certeza en cuanto a la identificación de las variantes y así evitar el riesgo de identificar falsos positivos. Además, se utilizaron solo SNPs bialélicos. Los marcadores obtenidos para cada análisis se colocaron

en un formato VCF para continuar con los análisis de diversidad y estructura genética (D'Agostino *et al.*, 2018).

Para poder determinar el número de grupos genéticos se realizó un análisis bicluster mediante el software *JMP* (JMP®, 2019) para asignar la pertenencia de los individuos a estos grupos. El bicluster es una técnica que permite simultáneamente la agrupación de filas y columnas de una matriz (Peña, 2014), en este caso las filas representan a los individuos y las columnas a los SNPs. Esta herramienta de análisis de datos exploratorios se utiliza para analizar agrupaciones de rasgos y de muestras simultáneamente de forma sintética.

Al visualizar la calidad de las matrices de SNPs, se consideró que la matriz generada con *Stacks* no fue lo suficientemente robusta para los análisis posteriores debido a que el número de marcadores obtenidos en el ensamble *de novo* era muy bajo (11 SNPs). Por lo que se decidió solo abordar el análisis utilizando el mapeo con el borrador del genoma de *V. planifolia*.

### 5.3.2 Diversidad y diferenciación genética

Los análisis descriptivos de diversidad constaron de una prueba de neutralidad (TajimaD) entre los grupos genéticos analizados utilizando *Tassel 5.0*. Las frecuencias alélicas (heterocigocidad observadas,  $H_o$ ), la diversidad nucleotídica  $\pi$ , y la diversidad nucleotídica poblacional  $\theta$  (se relacionan directamente con las frecuencias alélicas) así como el coeficiente de consanguineidad  $F_{IS}$ . Estos análisis se estimaron usando la función *populations* de *Stacks* (Catchen *et al.*, 2013).

Para determinar el nivel de diferenciación entre los grupos se utilizó el coeficiente de coancestría  $F_{ST}$ , que compara el nivel de variación genética que existe entre poblaciones o subpoblaciones (Wright, 1943). Los pares de  $F_{st}$  se calcularon con la herramienta *populations* de *Stacks* (Catchen *et al.*, 2013). Los datos obtenidos fueron graficados mediante la paquetería *ggplot2* en R 3.4.3 (R Core Team, 2008).



### 5.3.3 Análisis de Componentes Principales

El Análisis de Componentes Principales (PCA; por sus siglas en inglés) es un método estadístico de análisis multivariado. Esta herramienta se utiliza para el tratamiento de datos compuestos por numerosas variables, con el objetivo de sintetizar la información en términos de nuevas variables no correlacionadas entre sí (que no tenga repetición o redundancia en la información), llamado conjunto de componentes principales (PCs; Serrano, 2003).

Los resultados obtenidos son utilizados para explorar las similitudes o diferencias significativas entre los individuos, lo que permite visualizar patrones de ordenamiento entre el conjunto de datos (Ringnér, 2008). El paquete *SNPRelate* (Zheng *et al.*, 2012) del programa R se utilizó para realizar un análisis de componentes principales graficando los primeros componentes principales.

### 5.3.4 Análisis Discriminante de Componentes Principales

El Análisis Discriminante de Componentes principales (DAPC; por sus siglas en inglés), es una técnica estadística de análisis multivariado cuya finalidad es analizar si existen diferencias significativas entre grupos de objetos (individuos), respecto a un conjunto de variables (PCs) sobre los mismos para explicar en qué sentido se dan estas diferencias y facilitar procedimientos de clasificación sistemática.

Para identificar si los loci presentes en este conjunto de SNPs lograba clasificar las muestras según su origen (sitio de muestreo), se llevó a cabo un DAPC (Caldu-Primo *et al.*, 2017), mediante el paquete *adegenet* del programa R (Jombart y Ahmed, 2011; Hu *et al.*, 2019). Se utilizó el procedimiento de optimización propuesto por el paquete *adegenet* para evaluar el número óptimo de PCs a retener, con 1,000 repeticiones utilizando un 90% de los datos como un conjunto de datos de entrenamiento.

### **5.3.5 Estructura genética**

Para la estructura genética se utilizaron dos enfoques evolutivos de agrupamiento. Para el primer enfoque se realizó un análisis bayesiano de agrupamiento para identificar el número óptimo de grupos según el Criterio de Información Bayesiano (BIC; por sus siglas en inglés) mediante el paquete *adegenet* de R. El segundo enfoque se realizó infiriendo la ancestría mediante el software *Admixture* 1.3.0 por medio de una validación cruzada (Pritchard *et al.*, 2000; Alexander *et al.*, 2009; Zhou *et al.*, 2009).

En ambos modelos se analizó la matriz de SNPs, para cada valor de K, de 1 a 15. Para visualizar y comparar los resultados de agrupación, se utilizaron los paquetes *ggplot2*, en el programa R V. 3.5.3.

## **5.4 Modelos de distribución de especies**

Se realizó una base de datos de las coordenadas geográficas (solo con los organismos silvestres), para realizar los modelos de distribución con los puntos muestreados, basado en la colecta de trabajos previos (ver colecta de material). Para evitar sobrerrepresentación en el modelo, se eliminaron las coordenadas repetidas e individuos que estuvieran a menos de 1km de distancia, quedando 32 individuos en la base de datos.

### **5.4.1 Modelos de distribución actual**

Primero se determinó un área de accesibilidad de la especie utilizando la capa de ecorregiones terrestres de México (INEGI *et al.*, 2008), el cual contempla como base la vegetación primaria del INEGI, complementándose con los mapas de climas, edafología, geología y topografía. Esta capa nos proporciona una historia geológica y evolutiva de los sitios donde se ha encontrado la especie.

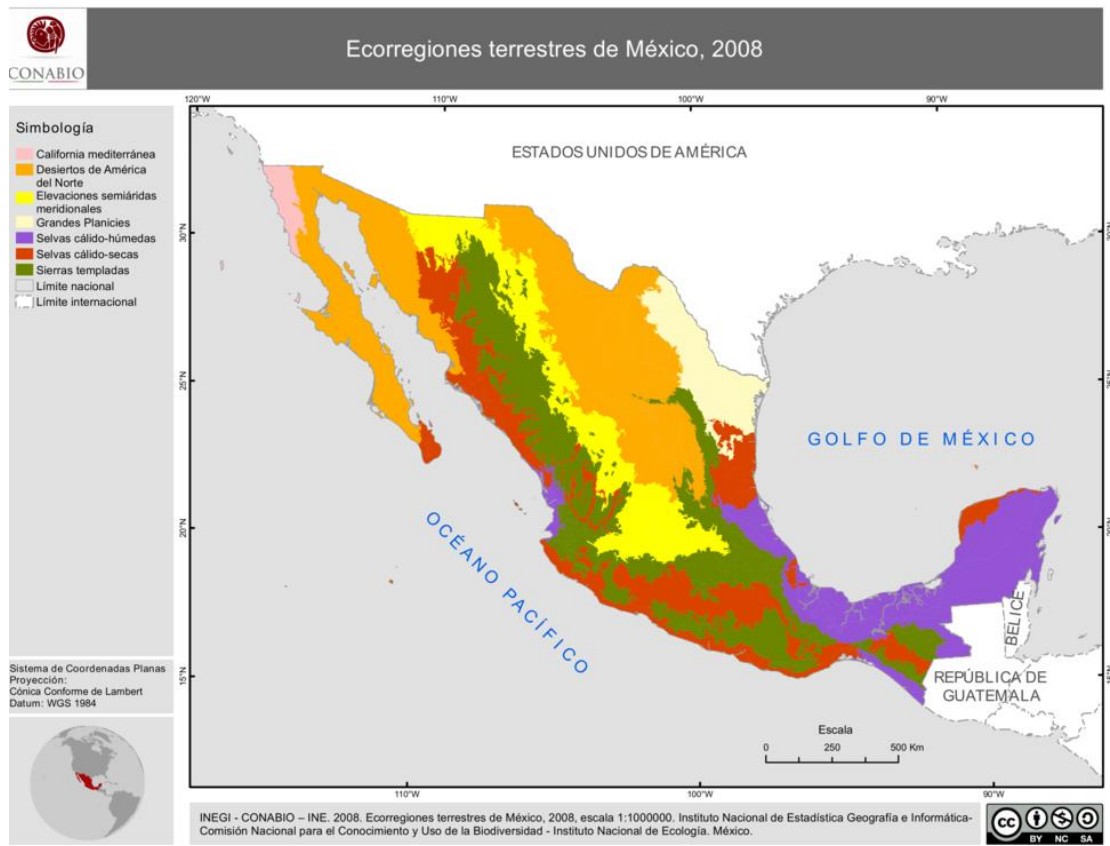


Figura 7. Mapa de las ecorregiones terrestres de México ([www.conabio.gob.mx](http://www.conabio.gob.mx))

Se descargaron las 19 capas de Worldclim con resolución espacial de 0.5 min (1 km<sup>2</sup> aproximadamente) en formato Raster (Busby, 1991; [www.worldclim.com](http://www.worldclim.com); Tabla. 3), las cuales son utilizadas en la mayor parte de los trabajos biogeográficos, para modelado de nicho y de distribución potencial (Hernández *et al.*, 2016; Hijmans *et al.*, 2005; Varela *et al.*, 2015). Posteriormente, estas capas se transformaron en formato *Ascii* mediante el software *Qgis* para estimar la distribución actual y el grado de influencia de estas variables sobre la distribución de *V. planifolia* con ayuda del algoritmo *MaxEnt*.

Se realizó un análisis de componentes principales para ver el grado de ordenación de las variables y poder discriminar las que no aportaban algún porcentaje de contribución. Además, se llevó a cabo una correlación de las variables analizando el factor de inflación de la varianza (VIF) mediante el software R, esto para conocer la multicolinealidad de las variables y así descartar las que tuvieran una mayor correlación, evitando se enmascarara la información

(Jiménez-Valverde, *et al.*, 2008). Las variables con mayor contribución (Tabla 3) se tomaron para realizar la modelación que constó de 50 réplicas teniendo 1000 iteraciones cada una, con una secuencia de arranque (*Bootstrap*) como método de remuestreo. El 20% de los registros se utilizó para la validación y el 80% restante para generar el modelo (Plasencia-Vázquez *et al.*, 2014).

Los modelos generados fueron sometidos a las pruebas de la curva de respuesta con el análisis de omisión o comisión, de sensibilidad ROC (*Receiver Operating Characteristic*), el área bajo la curva (AUC por sus siglas en inglés) y la prueba de Jackknife para verificar la confiabilidad y el efecto de cada variable dentro del modelo (Zuur *et al.*, 2009; Scheldeman y van Zonneveld, 2010).

Con fines comparativos, mediante sistemas de información geográfica, se añadieron capas vectoriales obtenidas de CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad; <http://www.conabio.gob.mx/informacion/gis>), estas fueron seleccionadas a través de una revisión bibliográfica investigando las variables que tienen mayor influencia en la distribución de *V. planifolia*. Las capas que se anexaron fueron: 1) uso de suelo y vegetación, 2) régimen de humedad, 3) clima, 4) temperatura anual, 5) precipitación anual y 6) elevación; las cuales fueron rasterizadas y posteriormente transformadas a formato Ascii (requisito de MaxEnt) para ejecutar el algoritmo de máxima entropía.

#### 5.4.2 Modelo de distribución futura

Se descargaron las proyecciones bioclimáticas del modelo de circulación general (GCM) MIROC-ESM para el año 2050 y 2070 bajo 2 escenarios de emisión de las denominadas Trayectorias de Concentración Representativas (RCP, por sus siglas en inglés). Un escenario con un nivel bajo de emisiones (RCP2.6) y un escenario con un nivel muy alto de emisiones (RCP8.5).

El escenario RCP2.6 plantea un escenario de declinación. Asume una reducción sustancial de las emisiones de gases de efecto invernadero (GEI), a lo largo del tiempo, con un aumento en la temperatura de 0.4 a 1.6°C para el periodo del 2046 al 2065, y de 0.3 a 1.7°C para el 2081 al 2100. Por otro lado, el escenario RCP8.5 plantea un incremento de las

emisiones de GEI a lo largo del tiempo con un aumento en la temperatura de 1.4 a 2.6°C para el 2046 al 2065, y de 2.6 a 4.8°C para el 2081 al 2100.

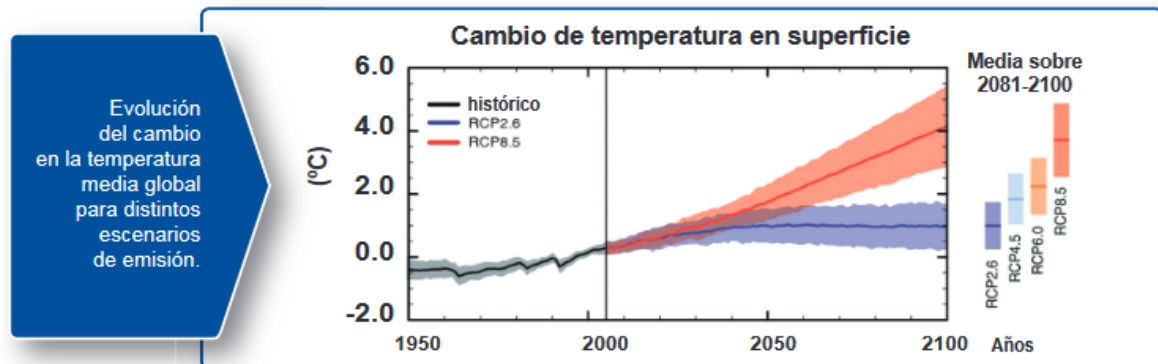


Figura 8. Cambios estimados en las temperaturas medias globales para los distintos escenarios de emisión en azul se observa el escenario RCP2.6 y en rojo escenario RCP8.5. (Tomado de Cambio Climático: Bases Físicas, [https://fundacion-biodiversidad.es/sites/default/files/informacion-institucional/ipcc5informeevaluacionresumen .pdf](https://fundacion-biodiversidad.es/sites/default/files/informacion-institucional/ipcc5informeevaluacionresumen.pdf))

Las variables con mayor contribución (Tabla 3) se tomaron para realizar la modelación que constó de 50 réplicas teniendo 1000 iteraciones cada una, con una secuencia de arranque (*Bootstrap*) como método de remuestreo.

Los resultados obtenidos de *MaxEnt* (capas GradientFile de distribución potencial), basado en ambos modelos (Worldclim y CONABIO), se exportaron al programa *Qgis* y se convirtieron a formato vectorial para estimar el tamaño de las áreas potenciales y describir sus atributos (Hernández *et al.*, 2016).

### 5.5 Relación genómica-geográfica-ambiental

Para analizar las relaciones entre la variación genómica y los factores geográficos y ambientales, se realizó una prueba parcial de Mantel, la cual, mediante un análisis de correlación entre dos matrices simétricas, se obtiene el coeficiente de correlación de Pearson.

La matriz geográfica y la matriz genética fueron generadas a partir de distancias Euclidianas. La matriz ambiental, se realizó con las coordenadas de los individuos secuenciados y las capas climáticas de Worldclim y CONABIO utilizadas para los modelos de distribución.

La prueba parcial de Mantel se realizó con el paquete ncf (<https://www.rdocumentation.org/packages/ncf>).

*Tabla 3. Descripción de las 19 variables bioclimáticas descritas por Busby (1991). Las variables sombreadas son las más predictivas y se seleccionaron para realizar el segundo modelo.*

Variable	Descripción
Bio1	Temperatura media anual
Bio2	Intervalo de temperaturas diurnas
Bio3	Isotermalidad (Bio2/Bio7) (*100)
Bio4	Temperatura estacional (Desviación Standard*100)
Bio5	Temperatura máxima del mes más cálido
Bio6	Temperatura mínima del mes más frío
Bio7	Rango anual de temperatura (P5-P6)
Bio8	Temperatura media del trimestre más lluvioso
Bio9	Temperatura media del trimestre más seco
Bio10	Temperatura media del trimestre más cálido
Bio11	Temperatura media del trimestre más frío
Bio12	Precipitación anual
Bio13	Precipitación del mes más húmedo
Bio14	Precipitación del mes más seco
Bio15	Estacionalidad de la precipitación (coeficiente de variación)
Bio16	Precipitación del trimestre más lluvioso
Bio17	Precipitación del trimestre más seco
Bio18	Precipitación del trimestre más cálido
Bio19	Precipitación del trimestre más frío

## 6. Resultados

Se secuenciaron 58 individuos de *V. planifolia* por medio de SGB para obtener marcadores tipo SNP. En general se muestra una distribución aleatoria en el número de lecturas por muestra. Al examinar el número de lecturas por muestra (min. 643,166, máx. 1,045,000, promedio 870,838), no se observa alguna variación en el número de lecturas asociada al sitio de colecta (Anexo I).

### 6.1 Variación genómica

Los resultados obtenidos después del filtrado indicaron un total de 252 SNPs para el mapeo en el borrador del genoma de *V. planifolia*.

Se identificaron 6 grupos genéticos dentro de la base de 252 SNPs mediante el análisis bicluster (Fig. 9). Los individuos pertenecientes a cada grupo, así como el número de sitios polimórficos y alelos privados se muestran en la Tabla 4.

Los análisis de diversidad genética que se realizaron con el conjunto de 252 SNPs pertenecientes a 58 individuos asignado a 6 grupos genéticos, describió una heterocigosidad observada ( $H_o$ ) = 0.2365, una heterocigosidad esperada ( $H_e$ ) = 0.2199, una diversidad nucleotídica ( $\pi$ ) = 0.2219 y poblacional ( $\theta$ ) = 0.2160, un D de Tajima = 0.0947 y un coeficiente de consanguineidad ( $F_{IS}$ ) = -0.0354.

La descripción genética de los seis grupos genéticos indicó que la heterocigosidad observada ( $H_o$ ) osciló entre los grupos teniendo una mínima de 0.2090 y una máxima de 0.2619, y fue mayor a la heterocigosidad esperada ( $H_e$ ; min. 0.1764 – máx. 0.2211), lo cual sugiere un leve exceso de heterócigos (Tabla 5). Se observó una diversidad nucleotídica parecida entre los grupos ( $\sim 0.2$ ), siendo el grupo 5 aquel con el valor más bajo (0.1925), y el grupo 6 el del valor más alto (0.2359). Los valores observados de la prueba D de Tajima para todos los grupos fueron negativos, la mayoría cercanos a -1, con valores que oscilan entre -0.9884 (grupo 2) y -1.9765 (grupo 5).



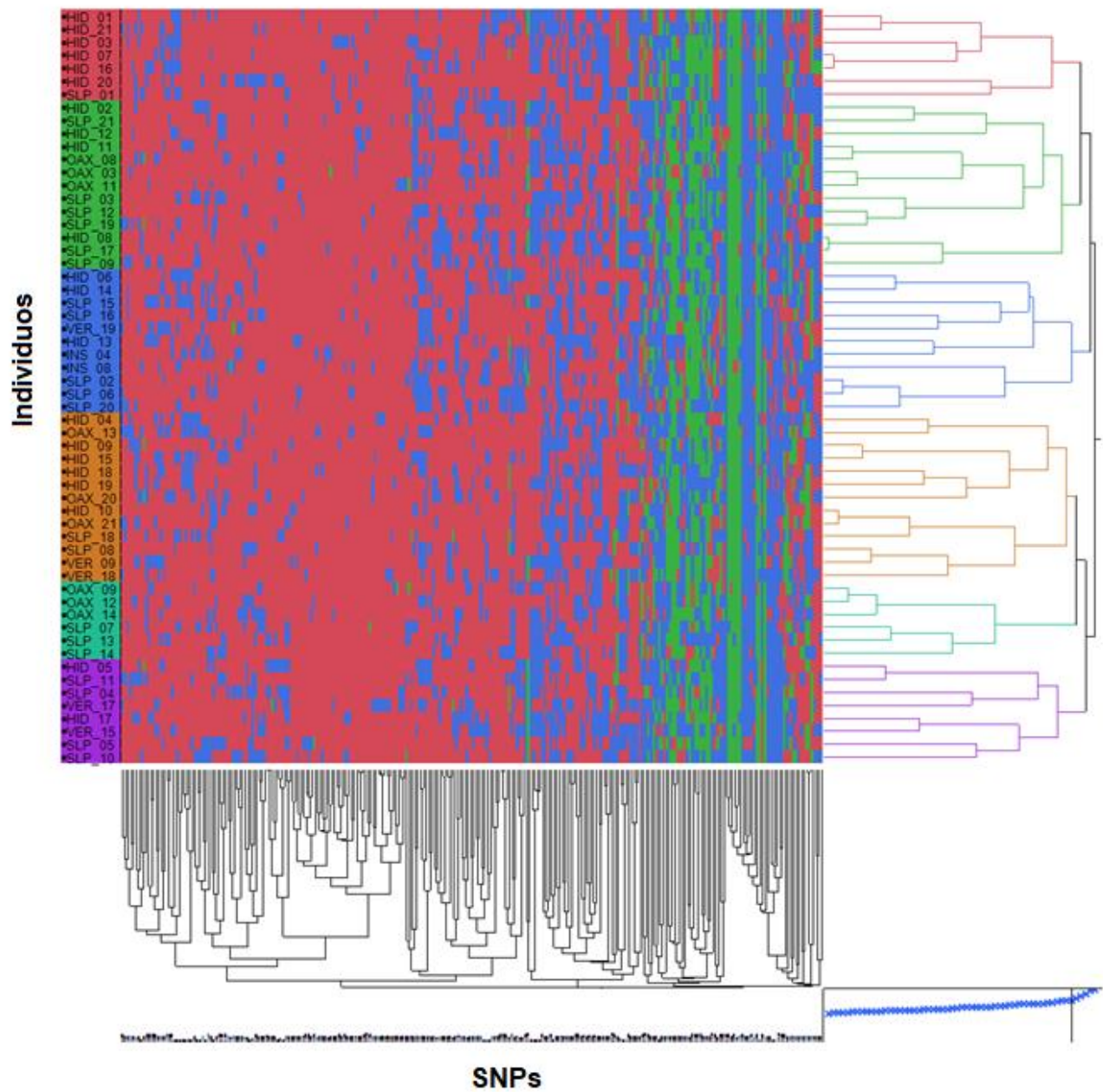


Figura 9. Análisis bicluster de los 252 SNPs resultantes del mapeo al genoma de *V. planifolia*. En la parte izquierda se encuentran los individuos, y en la inferior los SNPs. Los datos fueron normalizados mediante el método de Ward.



Tabla 4. Lista del grupo genético al cual se asignaron las muestras de *V. planifolia* incluidas en este estudio. Cada individuo recibió un identificador único dependiendo del punto de muestreo. Veracruz (VER), Hidalgo (HID), San Luis Potosí (SLP), Oaxaca (OAX) y Quintana Roo (INS). Se muestran 1) las asignaciones de grupo basadas en la prueba bicluster de los datos de GBS correspondientes al mapeo al genoma y 2) Individuos agrupados por sitio de muestreo.

1) Genoma grupos genéticos				
Grupo genético	Muestras	No. de muestras	Alelos privados	Sitios polimórficos
1	HID_01, HID_03, HID_07, HID_16, HID_20, HID_21, SLP_01	7	0	184
2	HID_02, HID_08, HID_11, HID_12, OAX_03, OAX_08, OAX_11, SLP_03, SLP_09, SLP_12, SLP_17, SLP_19, SLP_21	13	4	212
3	HID_06, HID_13, HID_14, INS_04, INS_08, SLP_02, SLP_06, SLP_15, SLP_16, SLP_20, VER_19	11	2	201
4	HID_04, HID_09, HID_10, HID_15, HID_18, HID_19, OAX_13, OAX_20, OAX_21, SLP_08, SLP_18, VER_09, VER_18	13	1	214
5	OAX_09, OAX_12, OAX_14, SLP_07, SLP_13, SLP_14	6	0	158
6	HID_05, HID_17, SLP_04, SLP_05, SLP_10, SLP_11, VER_15, VER_17	8	2	200
2) Sitio de Colecta				
Grupo genético	Muestras	No. de muestras	Alelos privados	Sitios polimórficos
HID	HID_01, HID_02, HID_03, HID_04, HID_05, HID_06, HID_07, HID_08, HID_09, HID_10, HID_11, HID_12, HID_13, HID_14, HID_15, HID_16, HID_17, HID_18, HID_19, HID_20, HID_21	21	6	235
SLP	SLP_01, SLP_02, SLP_03, SLP_04, SLP_05, SLP_06, SLP_07, SLP_08, SLP_09, SLP_10, SLP_11, SLP_12, SLP_13, SLP_14, SLP_15, SLP_16, SLP_17, SLP_18, SLP_19, SLP_20, SLP_21	21	1	232
OAX	OAX_03, OAX_08, OAX_09, OAX_11, OAX_12, OAX_13, OAX_14, OAX_20, OAX_21	9	1	194
INS	INS_04, INS_08	2	0	111
VER	VER_09, VER_15, VER_17, VER_18, VER_19	5	2	167

Tabla 5. Descripción de la diversidad genética de los seis grupos genéticos asignados por los 14 SNP's obtenidos del ensamble de novo.  $H_o$  = heterocigosidad observada,  $H_e$  = heterocigosidad esperada,  $F_{IS}$  = coeficiente de consanguineidad,  $\pi$  = diversidad nucleotídica,  $D$  = test estadístico de Tajima y  $\theta$  = diversidad nucleotídica poblacional.

# Pop ID	Genoma, grupos genéticos					
	$H_o$	$H_e$	$F_{IS}$	$\pi$	$D$	$\theta$
1	0.2619	0.2096	-0.0805	0.2258	-1.4813	0.298
2	0.2332	0.2053	-0.0379	0.2136	-0.9884	0.2711
3	0.2363	0.2071	-0.0497	0.2170	-0.9982	0.2723
4	0.2241	0.2011	-0.0431	0.2092	-1.0918	0.2736
5	0.2090	0.1764	-0.0326	0.1925	-1.9769	0.2745
6	0.2609	0.2211	-0.0608	0.2359	-1.2869	0.306

### 6.1.1 Estructura genética poblacional

Los valores de  $F_{ST}$  oscilaron entre 0.0285 y 0.0638, lo que indica valores bajos de diferenciación genética entre los grupos (Tabla 6).

Tabla 6. Estadísticos de divergencia genética ( $F_{ST}$ ) de los grupos genéticos pareados con 252 SNPs obtenidos del mapeo al genoma de *V. planifolia*.

$F_{ST}$	1	2	3	4	5	6
1		0.0317	0.0410	0.0386	0.0638	0.0442
2			0.0326	0.0285	0.0390	0.0359
3				0.0358	0.0510	0.0419
4					0.0361	0.0354
5						0.0505

Los dendrogramas construidos por NJ (Fig. 10), recuperan un agrupamiento distinto al construido por el análisis bicluster. Dentro de *V. planifolia*, el análisis NJ distingue tres clados, donde se agrupan individuos pertenecientes a diferentes grupos y a diferentes sitios de muestreo.

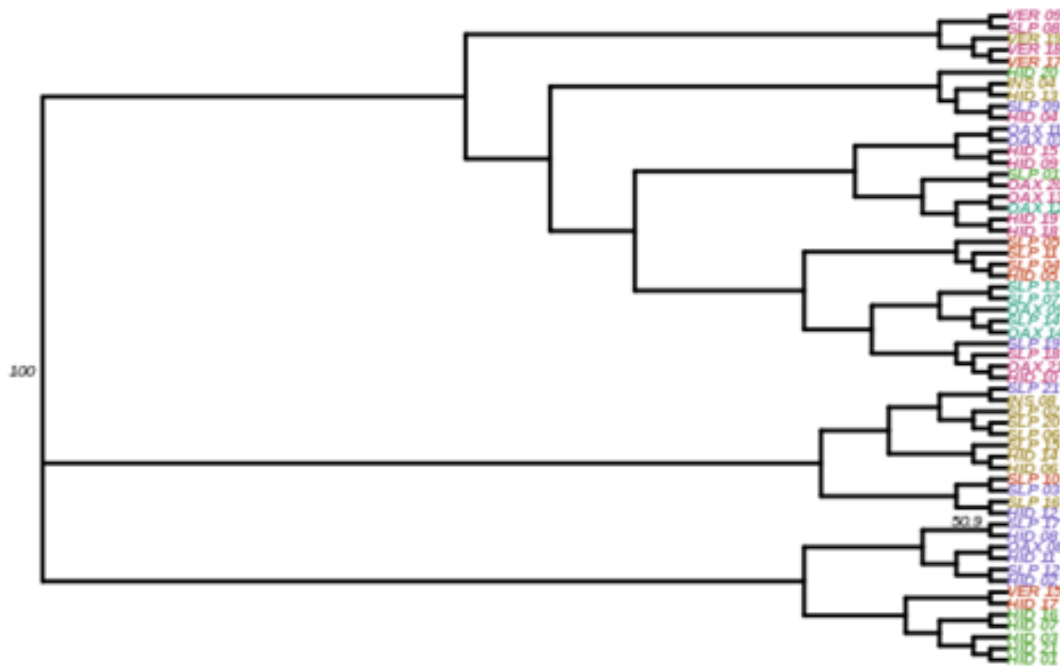


Figura 10. Dendrograma por el método Neighbor-joining de los SNPs encontrados en el mapeo al genoma de *V. planifolia* basado en distancias genéticas. El color de los individuos indica el grupo asignado mediante el análisis bicluster.

A través del análisis de componentes principales (PCA), se graficaron los dos primeros componentes que explican el 12% de la variación genética (PC1 7% y PC2 5%; Anexo II). No se observa un patrón que ordene a los individuos por grupo genético (Fig. 11).

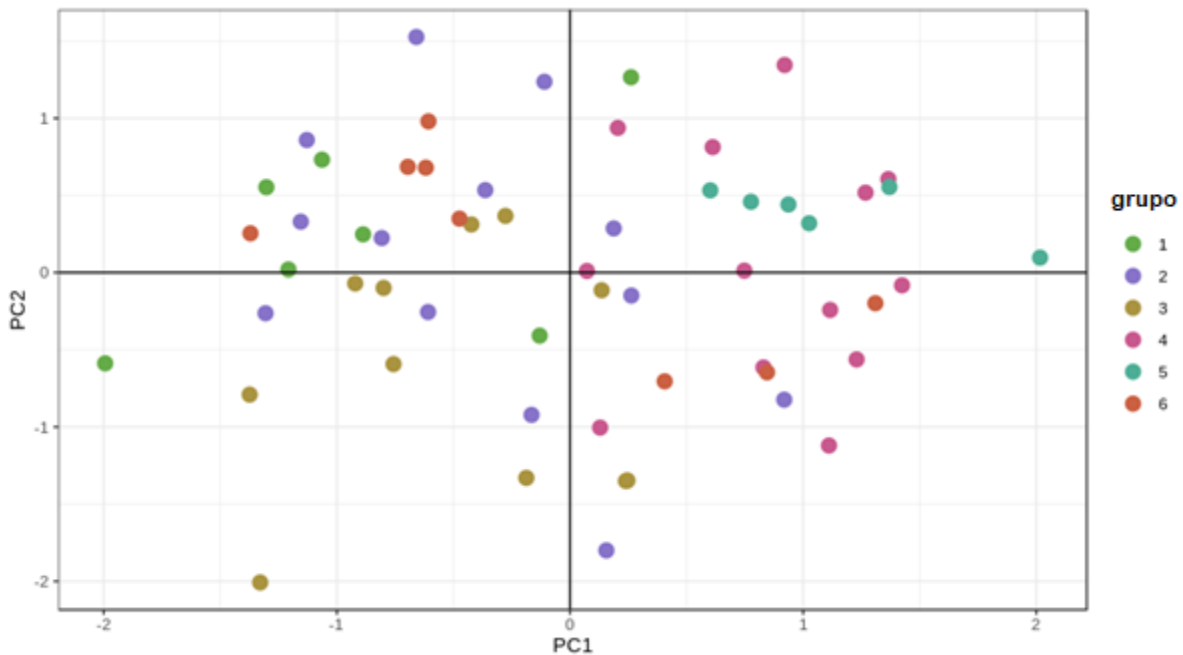


Figura 11. Resultado del análisis de componentes principales. Los 2 ejes corresponden a los dos primeros componentes que explican un 12% de la variación. Los colores corresponden a la asignación de grupos de los análisis anteriores.

La evaluación de asignación de grupos entre los 58 individuos se realizó mediante dos enfoques evolutivos de agrupamiento considerados para el análisis de la estructura poblacional, mediante la inferencia aproximada del modelo bayesiano (BIC) y las estimaciones de máxima probabilidad. El análisis BIC indicó que los individuos tienden a comportarse como un solo grupo (Fig. 12A).

Por otro lado, el software ADMIXTURE mostró un error de validación por cruzamiento para  $K=1$  como valor más bajo (Fig. 12B). Además, utilizando la función *compplot* del paquete *adegenet*, se representó la probabilidad de asignación de grupo de los individuos. Se observa que la mayoría de los individuos presentan altas probabilidades de pertenencia a un segundo o tercer agrupamiento (Fig. 12C).

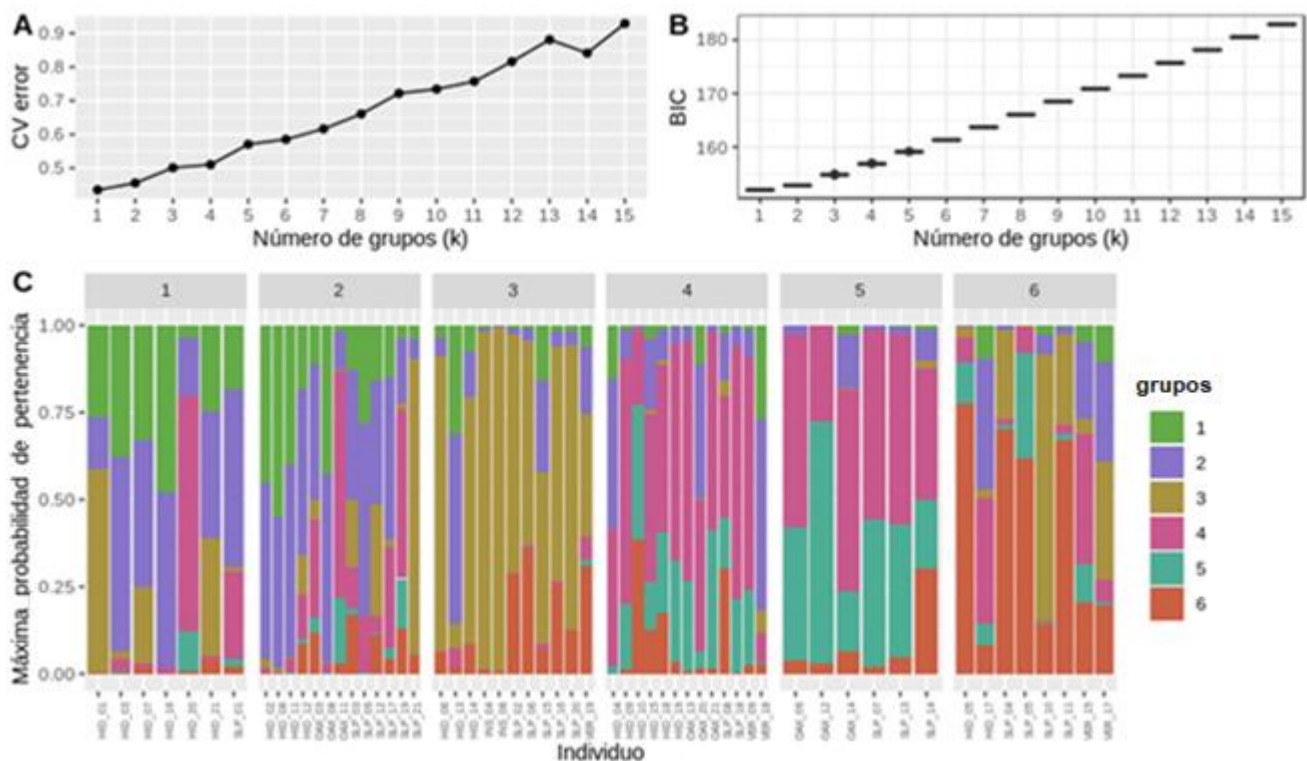


Figura 12. Análisis de agrupamiento con el conjunto de SNPs del mapeo al borrador del genoma de *V. planifolia*. A) El valor BIC más cercano a cero indica un agrupamiento óptimo ( $k=1$ ). B) Análisis de validación cruzada obtenido mediante el software *Admixture*. El valor más cercano a cero indica el número de grupos óptimo. C) Máxima probabilidad de pertenencia de los individuos de *V. planifolia* a los grupos asignados. Cada individuo analizado ( $n=58$ ), está representado por una barra vertical; cada grupo está representado por un color diferente.

### 6.1.2 SNPs asociados a los sitios de colecta

Se utilizó el conjunto de SNP previamente definidos para examinar si podrían ser útiles para distinguir los sitios de colecta en un análisis de componentes principales (PCA) y análisis discriminante de componentes principales (DPCA).

En el análisis de componentes principales (PCA), se graficaron los dos primeros componentes que explican aproximadamente el 12% de la variación genética (PC1 7% y PC2 5%). No se observa una tendencia de ordenamiento entre los individuos de acuerdo al sitio de muestreo, ni latitud como se reportó en estudios previos. Además, en el análisis de probabilidad de asignación de grupo, se observa que la mayoría de los individuos presentan altas probabilidades de pertenencia a un segundo o tercer agrupamiento (Anexo IV).

Debido a que la varianza acumulada por los eigenvalores en el PCA no mostró un comportamiento de parábola, lo cual es necesario para seleccionar el número óptimo de PCs a utilizar, se llevó a cabo un análisis de validación por cruzamiento (Anexo III), previo al Análisis Discriminante de Componentes Principales (DAPC).

El modelo de clasificación del análisis DAPC utilizó los primeros siete componentes principales del conjunto de 252 SNPs. Sin embargo, no fue posible clasificar las muestras de acuerdo al sitio de muestreo utilizando dos funciones discriminantes (Fig. 13).

En este análisis se observa un alto grado de mezcla, el cual está presente entre la mayoría de muestras (Fig. 13). Los individuos de Hidalgo, San Luis Potosí, Oaxaca y Veracruz son los que tienen más mezcla entre ellas y, en consecuencia, aquellos en los que el modelo es menos discriminante. Las muestras de Quintana Roo muestran cierto grado de mezcla, pero el modelo pudo reasignar a los individuos con cierta precisión.

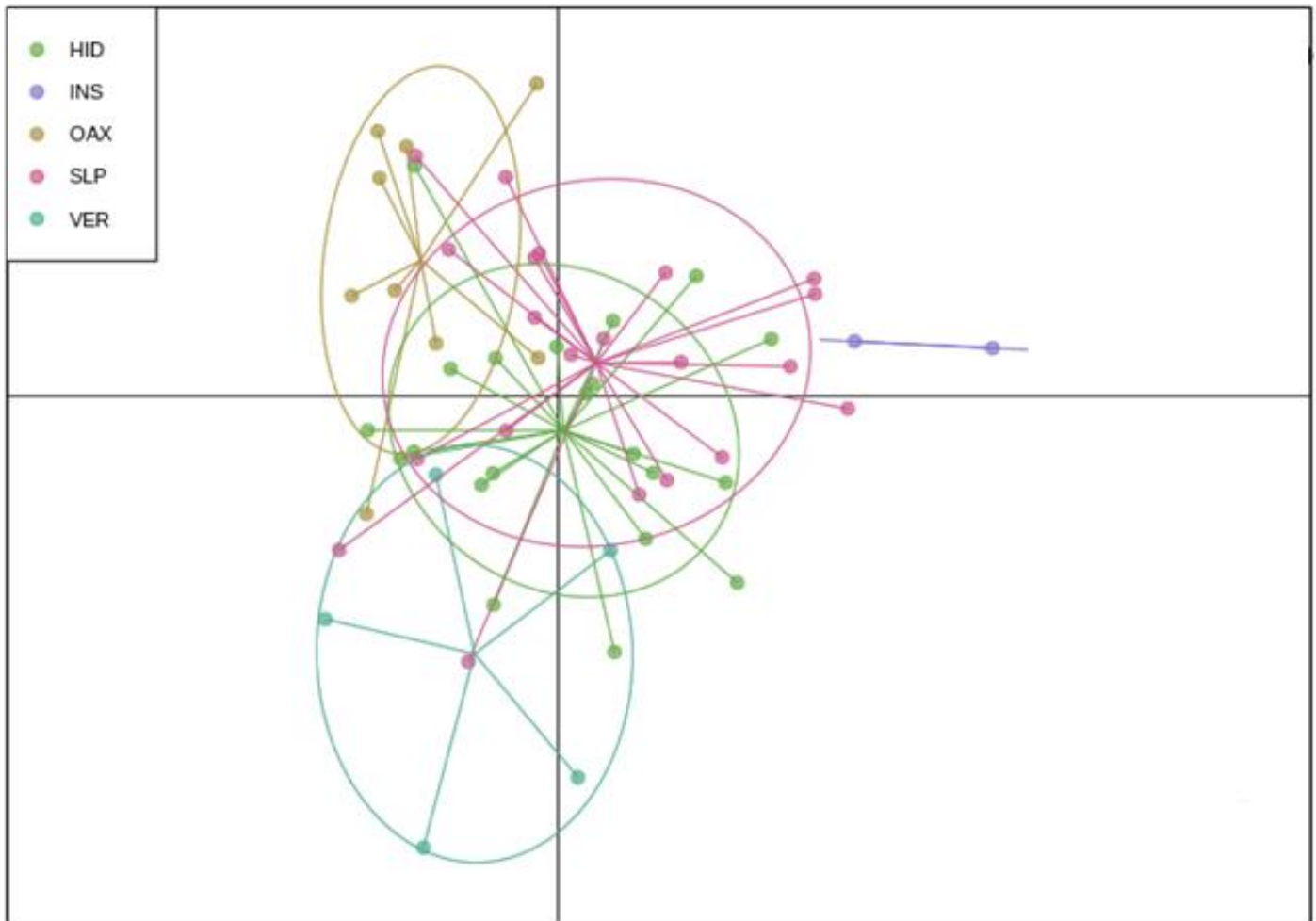


Figura13. Diagrama de clasificación: DAPC por sitio de colecta. Se representan gráficamente los primeros siete componentes principales de la función discriminante del DAPC para cada muestra.

De acuerdo con el conjunto de 252 SNPs determinados en los 58 individuos agrupados por sitio de colecta, la diversidad genética en *V. planifolia* describió una heterocigosidad observada ( $H_o$ ) que osciló entre grupos (0.2134-0.2444), lo cual fue mayor a la heterocigosidad esperada ( $H_e$ ), sugiriendo un leve exceso de heterocigos. Se observó una diversidad nucleotídica parecida entre los grupos ( $\sim 0.2$ ), siendo los individuos de Oaxaca donde se observó el valor más bajo (0.2088) y el valor más alto en los de Quintana Roo (0.2388). Los valores observados de la prueba de TajimaD para todos los grupos fueron negativos, la mayoría cercanos a -1, con valores que oscilan entre -0.5840 (SLP) y -2.4159 (VER). El valor D de Tajima para las muestras de Quintana Roo no se logró obtener debido a que solo cuenta

con dos individuos (Tabla 7). Por su parte, los valores de  $F_{ST}$  oscilaron entre 0.0139 y 0.0949, lo que indica valores bajos de diferenciación genética entre los grupos (Tabla 8).

Tabla 7. Descripción de la diversidad genética de los sitios de colecta.  $H_o$  = heterocigosidad observada,  $H_e$  = heterocigosidad esperada,  $F_{IS}$  = coeficiente de consanguineidad,  $\pi$  = diversidad nucleotídica,  $D$  = test estadístico de Tajima y  $\theta$  = diversidad nucleotídica poblacional.

Sitios de colecta						
Sitios de colecta	$H_o$	$H_e$	$F_{IS}$	$\pi$	$D$	$\theta$
HID	0.2443	0.2164	-0.0666	0.2218	-0.6038	0.2592
SLP	0.2360	0.2147	-0.0437	0.2200	-0.5840	0.2558
OAX	0.2134	0.1972	-0.0080	0.2088	-1.3700	0.2832
VER	0.2444	0.1984	-0.0500	0.2203	-2.4159	0.3181
INS	0.2441	0.1790	-0.0079	0.2388	-	0.4404

Tabla 8. Estadísticos de divergencia genética ( $F_{ST}$ ) pareados entre los sitios de muestreo, con 252 SNPs obtenidos del mapeo al genoma.

$F_{ST}$	HID	SLP	OAX	VER	INS
HID		0.0139	0.0213	0.0240	0.0330
SLP			0.0177	0.0245	0.0295
OAX				0.0464	0.0722
VER					0.0949



## 6.2 Modelos de distribución de *V. planifolia*

Se obtuvieron un total de 30 registros de presencia de *V. planifolia*. La mayor parte de los registros estuvieron localizados en los estados de Hidalgo, Querétaro y San Luis Potosí, así como algunos registros en el estado de Oaxaca.

### 6.2.1 Modelos de distribución actual

Después del análisis de contribución y de multicolinealidad, fueron las seleccionadas para realizar los modelos de distribución las siguientes variables (Tabla 12A; Fig. 14A): Intervalo de temperaturas diurnas (Bio 2), Isotermalidad (Bio 3), Temperatura media del trimestre más seco (Bio 9), Precipitación del mes más húmedo (Bio 13), Estacionalidad de la precipitación (Bio 15), Precipitación del trimestre más cálido (Bio 18). A pesar de que las variables Precipitación anual (Bio 12) y Precipitación del mes más seco (Bio 14) no eran consideradas para el modelo, después de realizar los análisis se decidió conservarlas ya que se conoce que la precipitación es una variable importante en la distribución de las orquídeas (Hernández *et al.* 2016). Para los modelos de distribución con las variables obtenidas de CONABIO, a pesar de haber realizado los análisis de contribución y multicolinealidad, se decidió conservar las 6 variables debido a su importancia en la distribución del modelo de estudio (Tabla 12B; Fig. 14B).

Tabla 9. Valores del Factor de Inflación de la Varianza (VIF), de las variables climáticas y porcentaje de contribución de la variable dentro del modelo de MaxEnt, utilizadas en el modelo de las capas de Worldclim y en el modelo con las capas de CONABIO.

Worldclim			CONABIO		
Variables	VIF	% de contribución	Variables	VIF	% de contribución
Bio_13	4.004	11.1	Vegetación	2.973	33.6
Bio_15	5.666	1.6	Reg. humedad	5.858	21.6
Bio_18	6.716	5.4	Clima	3.393	17
Bio_2	3.438	9.4	Precipitación	9.774	10.3
Bio_3	4.484	17.6	Temperatura	2.237	10.2
Bio_9	4.550	29.9	Elevación	2.642	7.3
Bio_12	-	4.6			
Bio_14	-	20.4			

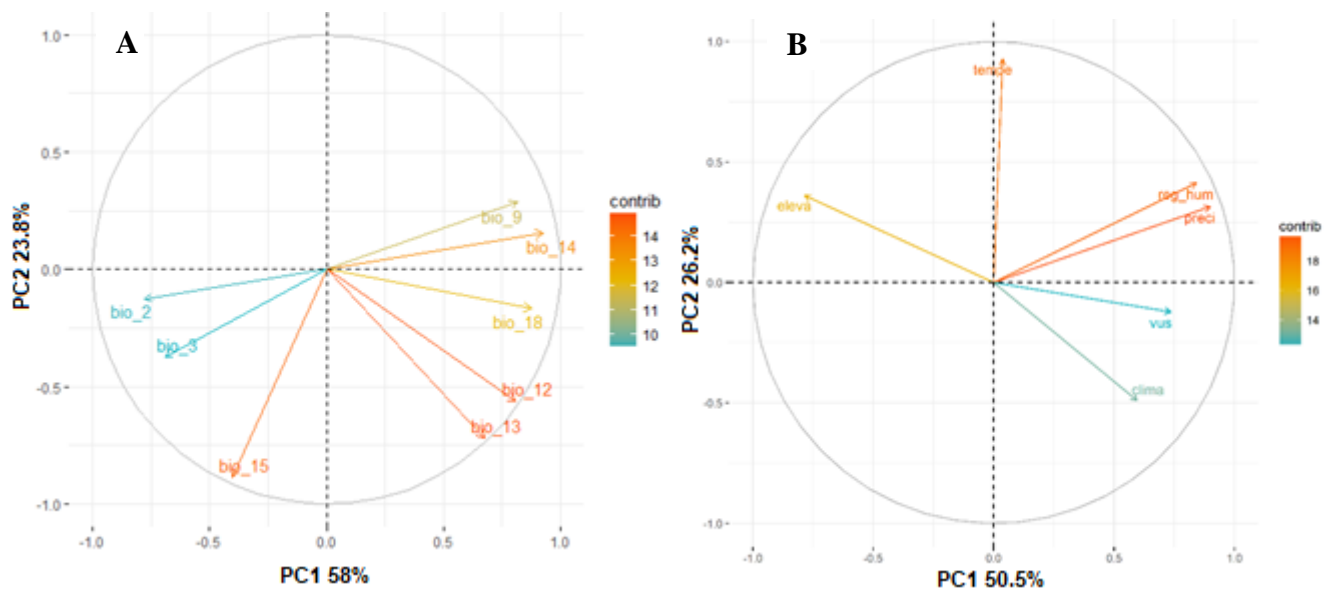


Figura 14. Análisis de Componentes principales de A) variables climáticas Worldclim y B) variables climáticas de CONABIO. Se observa el porcentaje de contribución al ordenamiento de los datos de cada variable.

El valor promedio de AUC para los modelos de distribución fue 0.968 ( $\pm 0.022$ ; Fig. 15A) para el modelo obtenido de las capas de Worldclim y un AUC de 0.986 ( $\pm 0.006$ ; Fig. 15B) para las variables de CONABIO.

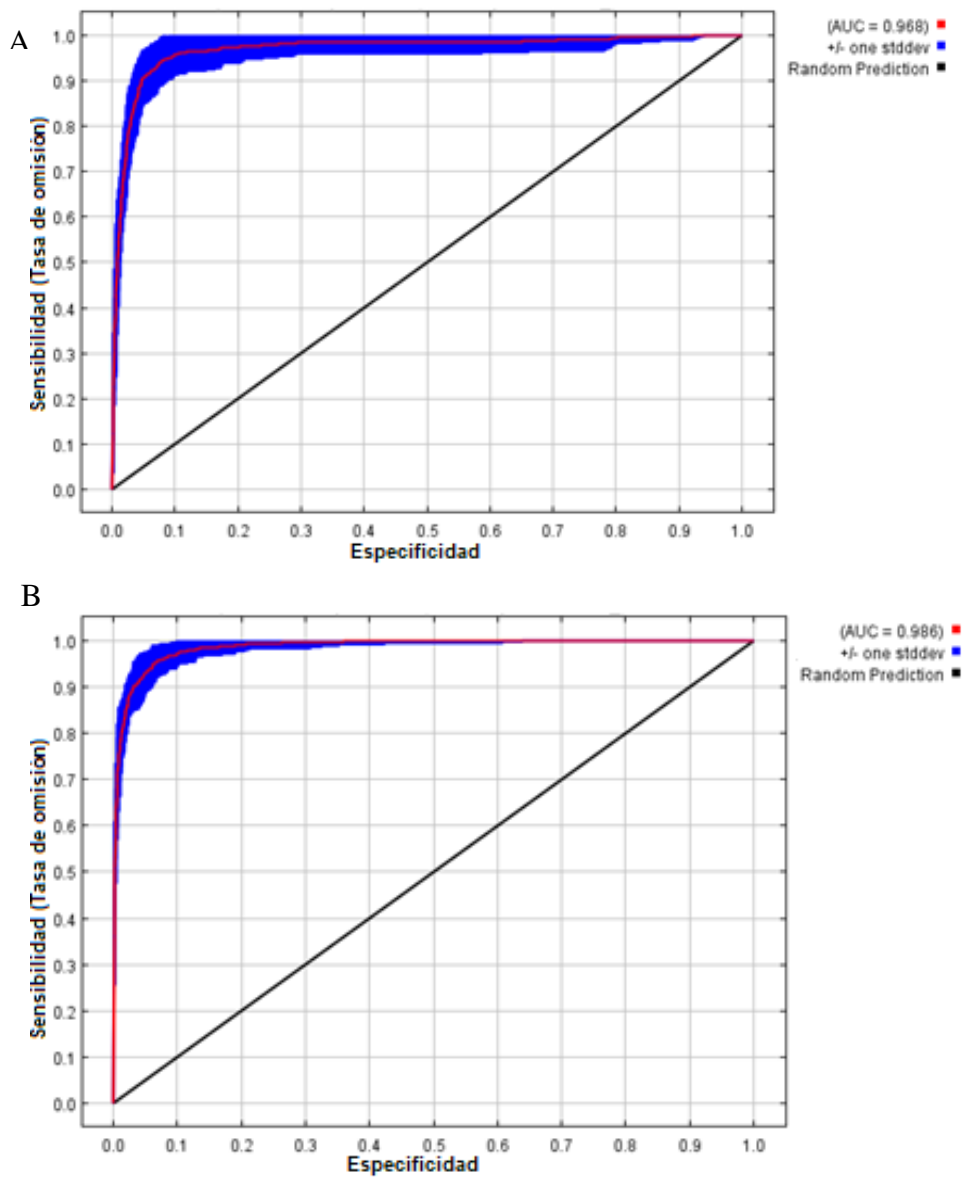


Figura 15. Sensibilidad y especificidad para los modelos de *V. planifolia*. La línea roja muestra el ajuste del modelo a los datos de información de la especie. A) Capas de Worldclim B) Capas de CONABIO.

En la Figura 16 se muestra el análisis Jackknife de las variables de Worldclim y CONABIO. Las barras color turquesa representan la entropía que se obtiene al quitar la variable, las barras color azul representan la entropía resultante solo con esa variable, teniendo como la máxima entropía la barra en color rojo.

A partir de los valores de entropía obtenidos del análisis Jackknife se determinó que, la Temperatura media del trimestre más seco (Bio 9) de Worldclim aportó la mayor ganancia al modelo cuando se considerará de forma aislada. Las otras tres variables que más contribuyeron al modelo fueron: la isothermalidad (Bio 03); la precipitación del cuatrimestre más cálido (Bio 18) y Precipitación del mes más seco (Bio 14; Fig. 16A).

Por parte del modelo del conjunto de variables de CONABIO, las variables que contribuyen como predictores para el modelo de distribución con las capas de CONABIO son: uso de suelo y vegetación, régimen de humedad y clima (Fig. 16B). El modelo de CONABIO, a pesar de incluir la elevación no la considera como una variable importante. Sin embargo, el análisis delimitó la zona de distribución en un intervalo principal de 200 a 500 msnm, lo cual concuerda con estudios previos que ubicaban a la especie en un intervalo de 250 a 750 msnm (Soto-Arenas, 1999). Aunque otro intervalo importante donde existe una alta probabilidad de presencia en 2,000 a 2,500 msnm, donde por lo general están al pie de sistemas montañosos medios de 1,300 a 2,500 msnm, haciendo su distribución más amplia (Hernández-Santana, *et al.*, 2009).

La vegetación más influyente en la distribución fue la agricultura temporal, selva alta perennifolia y en menor cantidad bosque de pino-encino, siendo selva alta perennifolia donde más se sitúan los puntos de colecta en herbarios (Soto-Arenas y Dressler, 2010). Ya que las muestras hacen referencia a poblaciones silvestres, se cree que, debido al grosor en la escala de las capas descargadas, el uso de suelo de agricultura tuvo mayor probabilidad de presencia.

El régimen de humedad resultó údico con 270 a 365 días de humedad y el clima Semicálido húmedo, con temperatura media anual mayor de 18°C, temperatura del mes más frío menor de 18°C, temperatura del mes más caliente mayor a 22°C. Precipitación anual mayor a 500 mm y precipitación del mes más seco mayor de 40 mm; lluvias de verano y porcentaje de lluvia invernal mayor al 10.2% del total anual como la mayor probabilidad de presencia.

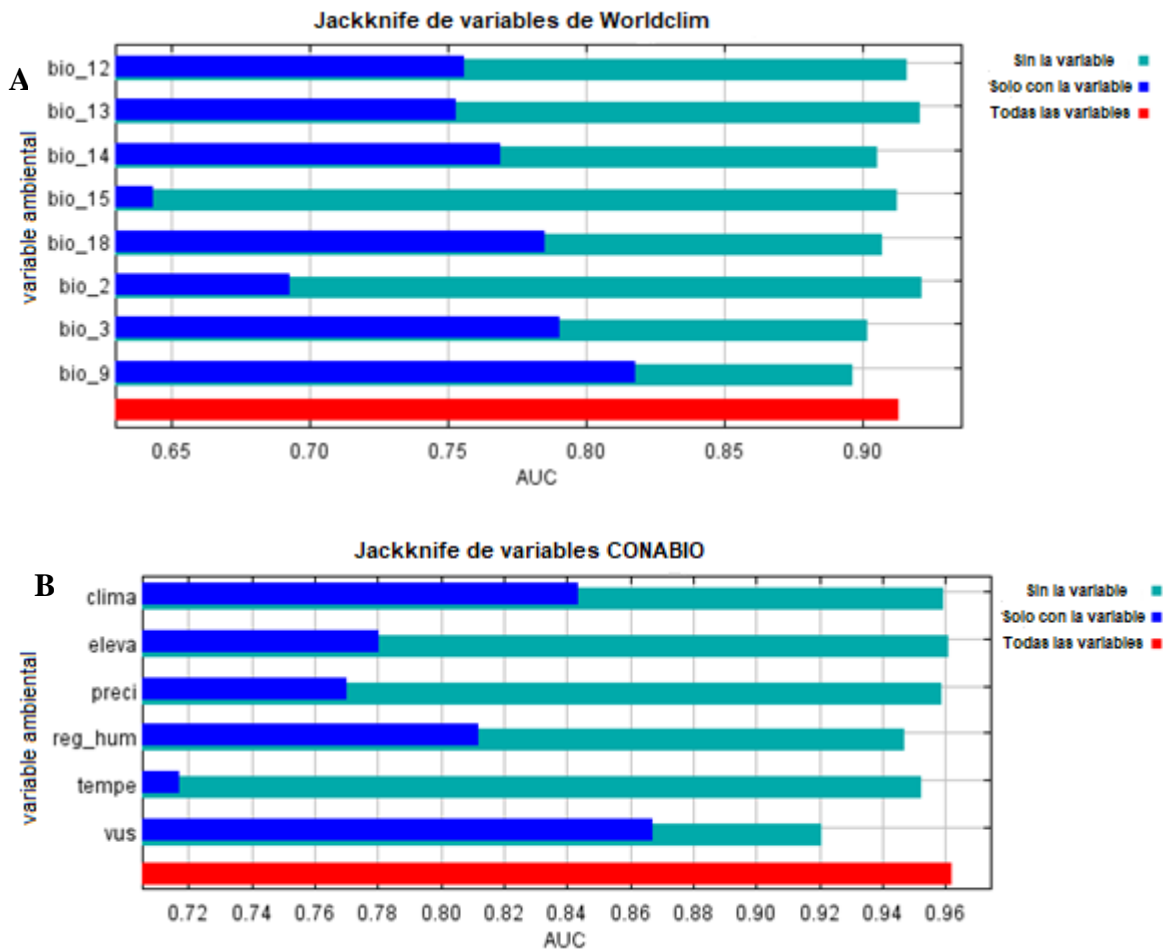
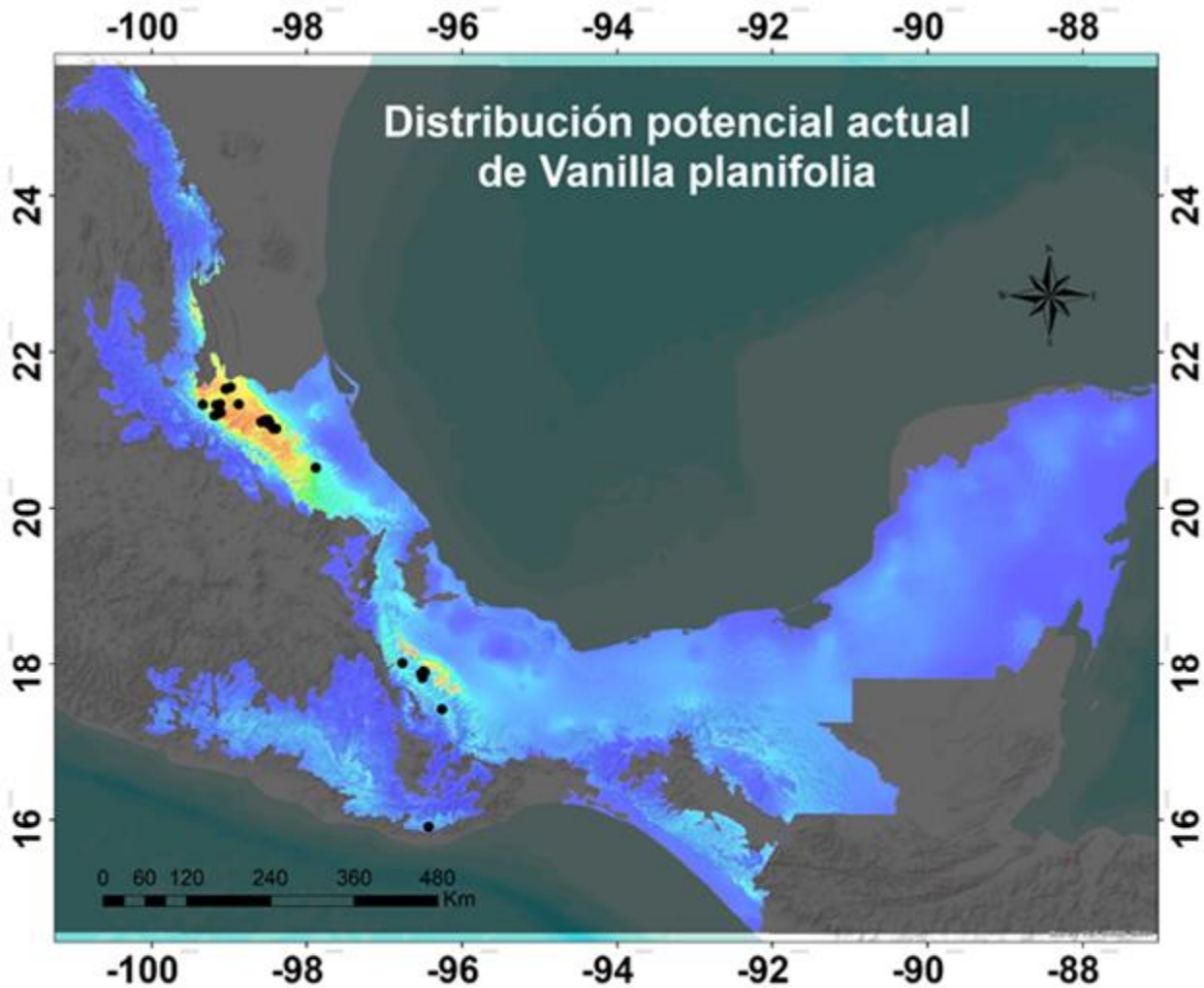


Figura 16. Prueba de Jackknife que muestra la contribución al modelo de cada variable (azul), sin la variable (turquesa) y con todas las variables (rojo): A) WorldClim. B) CONABIO. Los valores mostrados son promedio de las 100 réplicas ejecutadas.

Los mapas de la distribución actual a partir de los dos conjuntos de variables (Fig. 17 y 18), presentaron un patrón similar, indicando una distribución restringida a la zona montañosa de la sierra madre oriental, que se desplaza a través de las ecorregiones selvas cálido-húmedas y sierras templadas.



● **Puntos de presencia de *V. planifolia***

**Idoneidad ambiental**

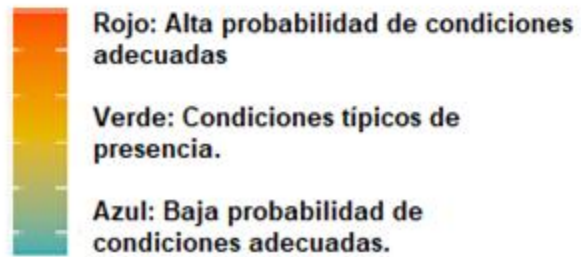
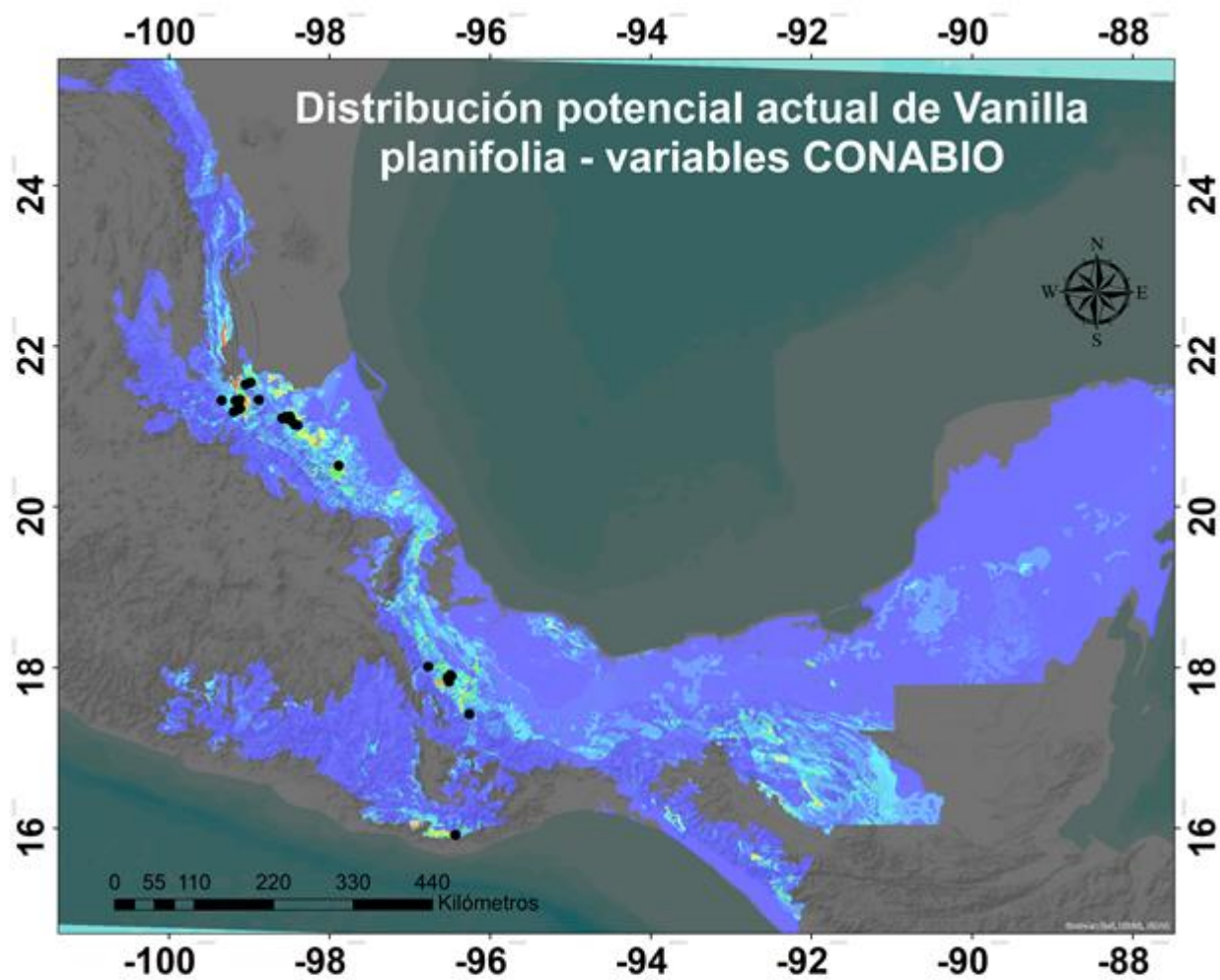


Figura 17. Mapa de la distribución actual de *V. planifolia* utilizando las variables de Worldclim.



● **Puntos de presencia de *V. planifolia***  
**Idoneidad ambiental**

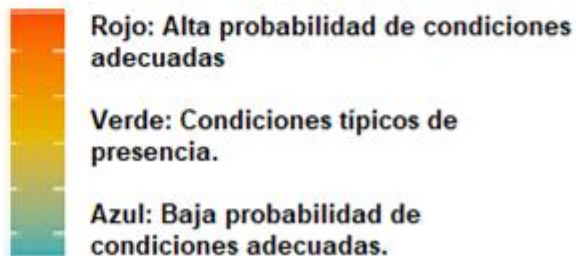


Figura 18. Mapa de la distribución actual de *V. planifolia* utilizando las variables de CONABIO.



### 6.2.2 Modelos de distribución futura

Los resultados muestran que las modificaciones en las variables climáticas propuestas por los modelos de circulación general y escenarios utilizados en este estudio, tienen efectos en la distribución de *V. planifolia*. Es decir, si se descarta la distribución por debajo de ~0.68 de probabilidad de ocurrencia, la distribución actual generada con base en las capas *lowerci* (capas generadas con un 95% de confiabilidad) de Worldclim cuenta con un área de 3,651 km<sup>2</sup>. Esta área se encuentra fraccionada en varias áreas de distribución (Fig. 17). Por parte de las proyecciones para el año 2050 la distribución potencial suma un área de 2,484 km<sup>2</sup> para un escenario con un nivel bajo de emisiones (RCP2.6) y 2,129 km<sup>2</sup> para un escenario con un nivel muy alto de emisiones (RCP8.5). Y de 1,123 km<sup>2</sup> (RCP2.6) y 2,063 km<sup>2</sup> (RCP8.5) para el 2070. Esto representa una disminución del 31.9% en RCP2.6 y 41.6% del RCP8.5 para el 2050 respecto a la distribución actual. Para el 2070, representa una disminución del 69.2% y 43.5% respectivamente (Fig. 19). Estas proyecciones se basaron en las variables seleccionadas de WorldClim (Tabla 3).

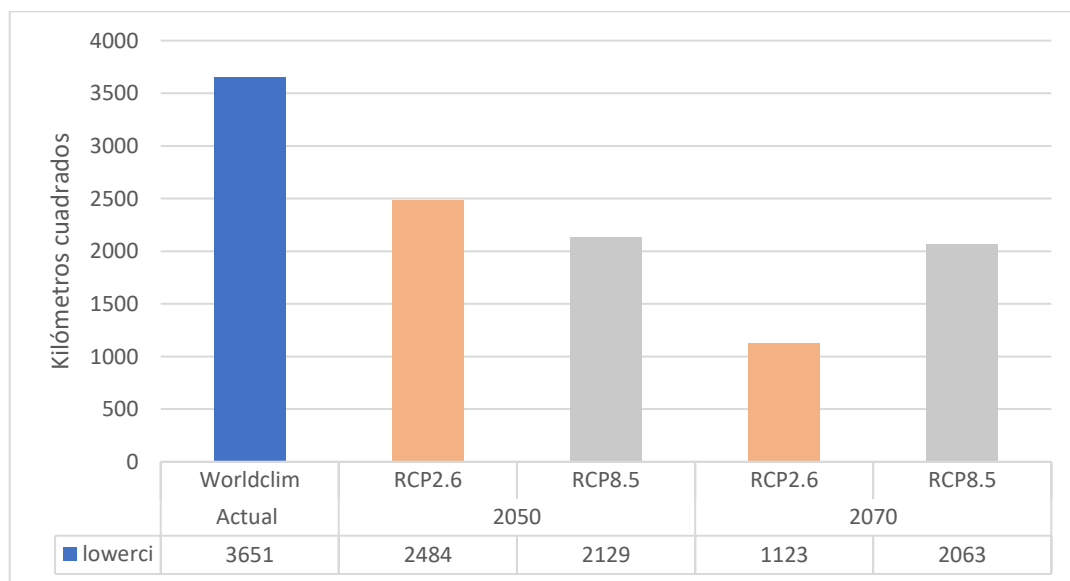
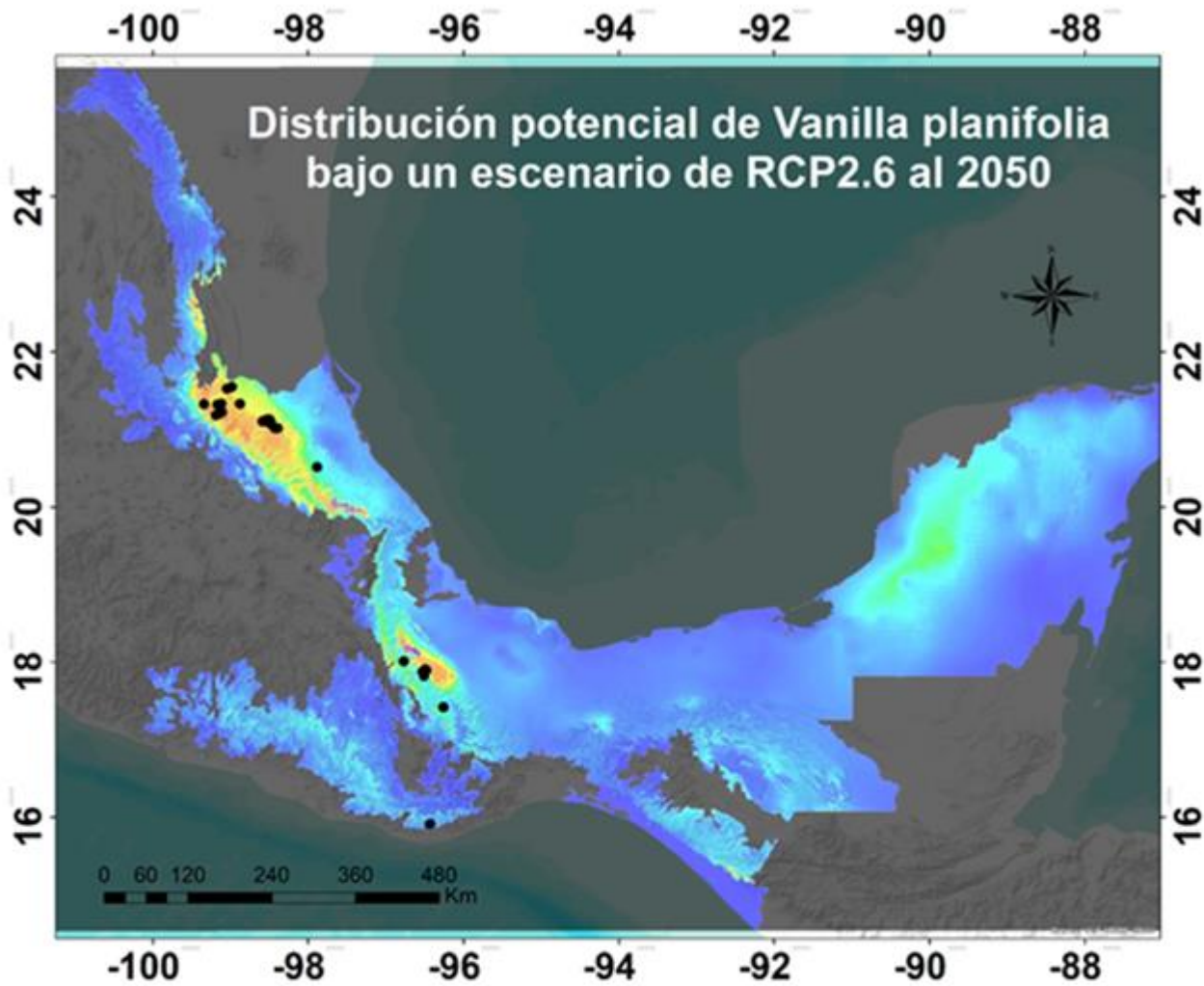


Figura 19. Áreas de distribución en km<sup>2</sup>. Actual, con el modelo de WorldClim. Proyección futura, con los modelos de WorldClim bajo los escenarios de cambio climático RCP2.6 y RCP8.5, para los años 2050 y 2070.





**Puntos de presencia de *V. planifolia***  
**● idoneidad ambiental de la proyección al año 2050**

- Rojo: Alta probabilidad de condiciones adecuadas
- Verde: Condiciones típicas de presencia.
- Azul: Baja probabilidad de condiciones adecuadas.

Figura 20. Mapa de distribución potencial proyectado al futuro en un escenario con un nivel bajo de emisiones RCP 2.6 para el año 2050 de acuerdo al modelo de circulación general MIROC-ESM.

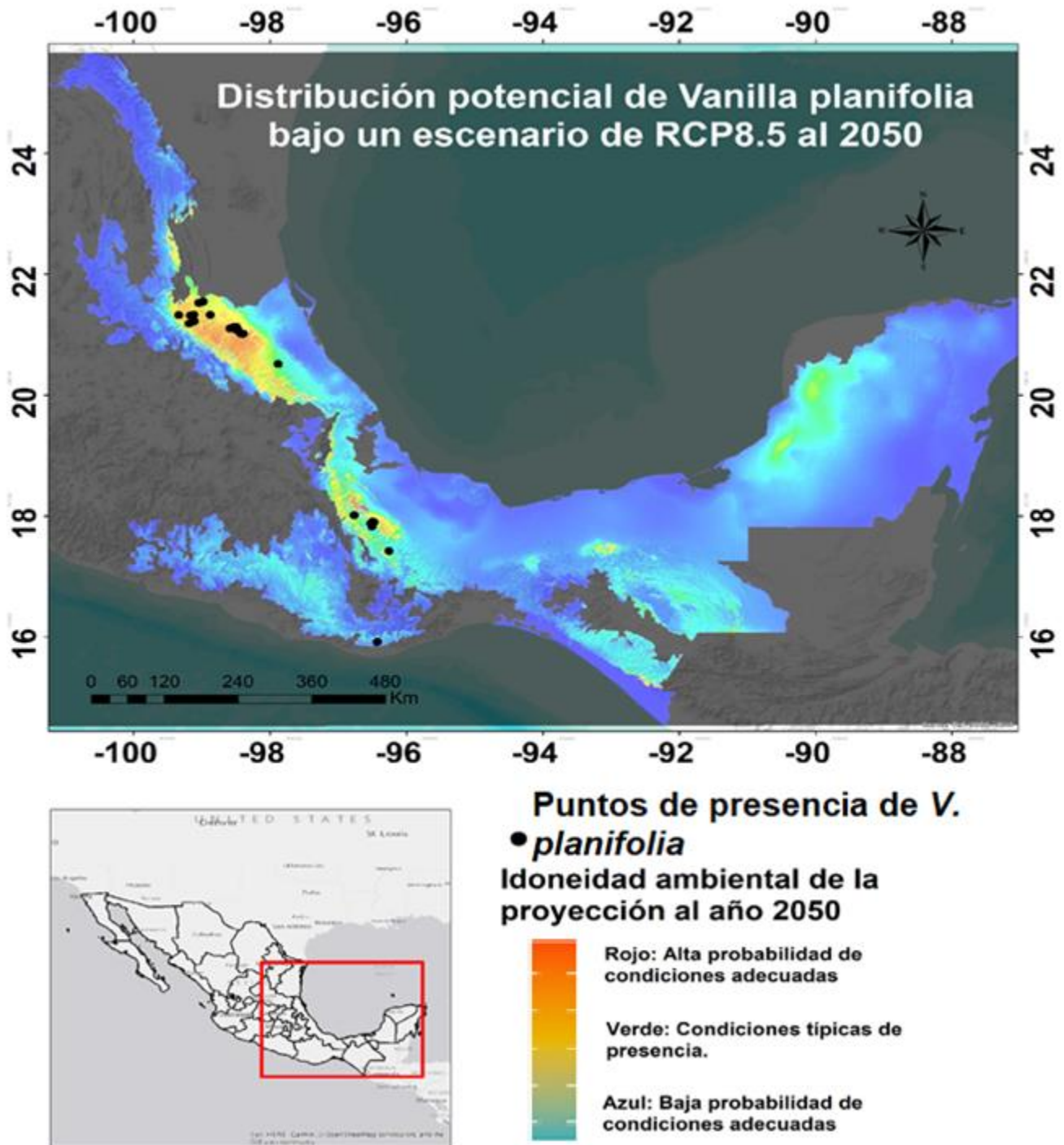
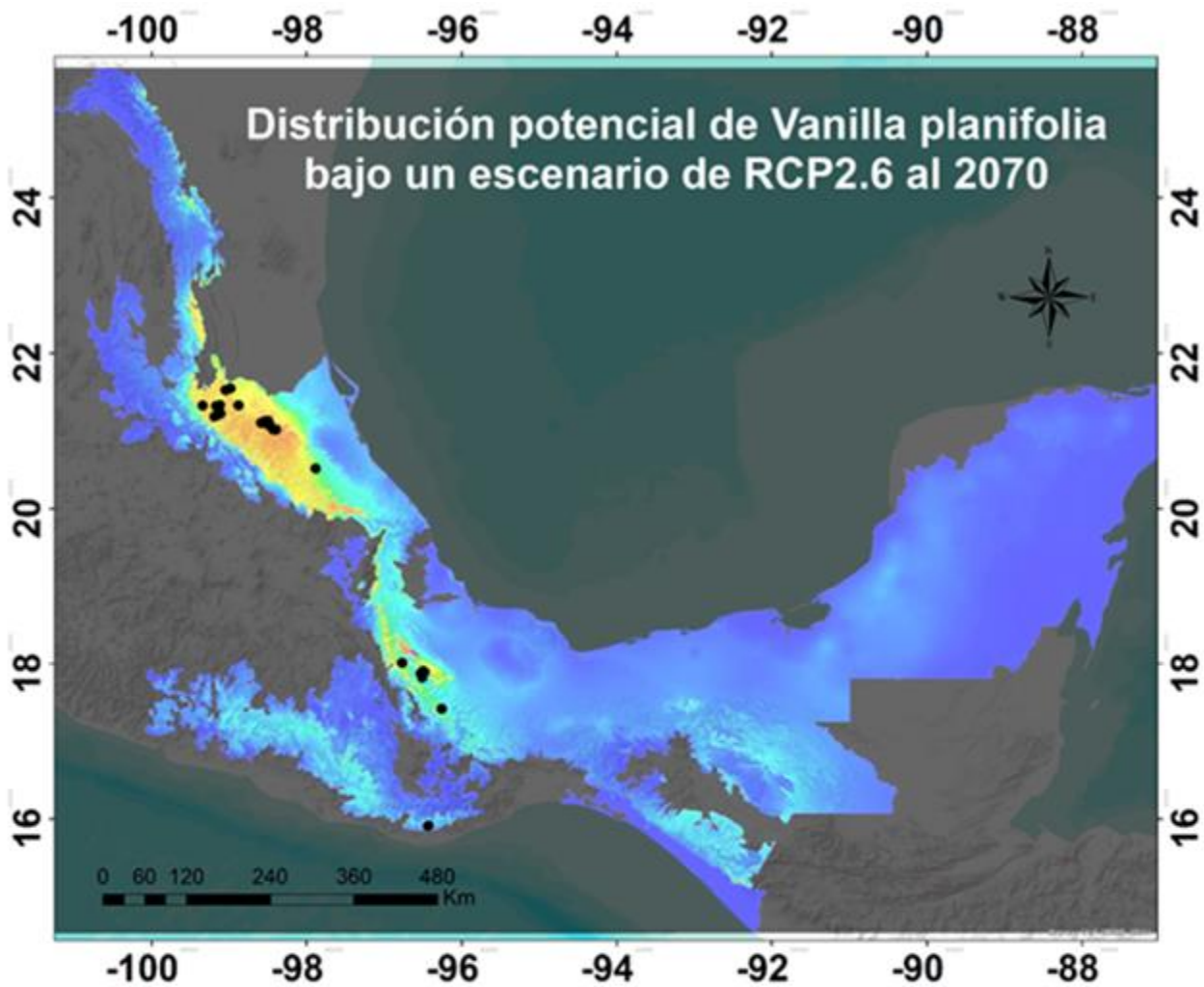


Figura 21. Mapa de distribución potencial proyectado al futuro en un escenario con un nivel muy alto de emisiones RCP 8.5 para el año 2050 de acuerdo al modelo de circulación general MIROC-ESM.



Puntos de presencia de *V. planifolia*

Idoneidad ambiental de la proyección al año 2070

- Rojo: Alta probabilidad de condiciones adecuadas
- Verde: Condiciones típicos de presencia.
- Azul: Baja probabilidad de condiciones adecuadas.

Figura 22. Mapa de distribución potencial proyectado al futuro en un escenario con un nivel bajo de emisiones RCP 2.6 para el año 2070 de acuerdo al modelo de circulación general MIROC-ESM.



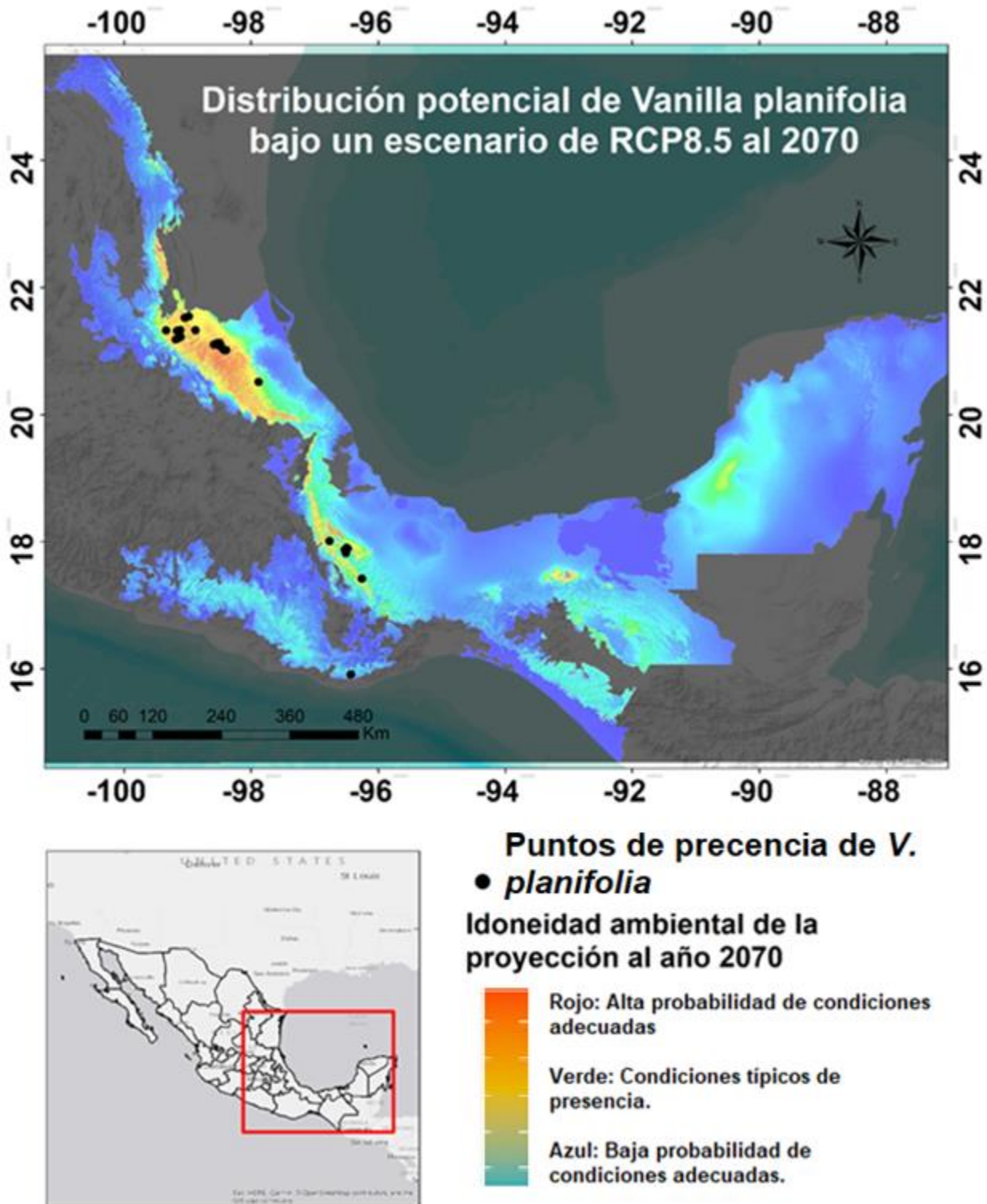


Figura 23. Mapa de distribución potencial proyectado al futuro en un escenario con un nivel muy alto de emisiones RCP 8.5 para el año 2070 de acuerdo al modelo de circulación general MIROC-ESM.

### 6.3 Relación genómica-geográfica-ambiental

Con la prueba parcial de Mantel (Tabla 10), se observa una baja correlación entre el flujo génico y las distancias geográficas (0.3069, p-valor ~0.0004). Se observa un patrón similar entre las matrices de distancias geográficas y las ambientales. Por otro lado, al analizar las matrices de variables ambientales relacionadas con la matriz de distancias genéticas, se observaron valores muy bajos de correlación para todos los casos (~0.05, p-valor >0.15).

La última fila de la tabla 10, nos indica la correlación entre las matrices genética y geográfica, controlada por el efecto de la matriz de distancias ambientales. Los valores positivos (~0.3, p-valor ~0.0003) indica que existe una correlación que puede relacionarse a la distancia geográfica.

Tabla 10. Coeficientes de correlación entre las matrices, obtenidos por medio de la prueba parcial de Mantel. Gen=matriz de flujo génico, Geo=matriz de distancias geográficas, Amb=matriz de distancias euclidianas entre las variables ambientales: CONABIO (6 variables), WorldClim (8 variables) y sumando las variables de ambos modelos.

	CONABIO		WorldClim		WorldClim + CONABIO	
	Parcial de Mantel ( <i>r</i> )	<i>p</i> -valor	Parcial de Mantel ( <i>r</i> )	<i>p</i> -valor	Parcial de Mantel ( <i>r</i> )	<i>p</i> -valor
<b><i>Gen-Geo</i></b>	0.3069	0.0002	0.3069	0.0004	0.3069	0.0004
<b><i>Gen-Amb</i></b>	0.0457	0.1549	0.0551	0.2241	0.0552	0.2283
<b><i>Geo-Amb</i></b>	0.1823	0.0039	0.3883	0.0049	0.3882	0.0044
<b><i>(Gen-Geo). Amb</i></b>	0.3040	0.0002	0.3103	0.0003	0.3102	0.0004

## 7. Discusión

*V. planifolia* es la especie principal para la obtención del prestigioso y aromático saborizante de vainilla, centrándose en esta especie, el 95% de la producción mundial (Lubisnsky, *et al.*, 2008). Se ha observado a nivel mundial que la diversidad genética al interior de las plantaciones de *V. planifolia* es limitada o prácticamente nula (Nielsen y Siegismund, 1999; Cameron, 2004; Schlüter *et al.*, 2007; Soto-Arenas y Dressler, 2010), lo que es acorde con la facilidad de propagación vegetativa y con la distribución mundial de algunos clones fundadores (Hu *et al.*, 2019).

En este estudio, se presenta un análisis de variación genética infraespecífica a nivel genómico de *V. planifolia*; en conjunto con un análisis de distribución potencial actual y futura. Se espera que este estudio, contribuya al conocimiento de este recurso y a la elaboración de programas de manejo y de conservación de la especie.

### 7.1 Ensamble de novo

Modificar los diferentes parámetros nos permite maximizar el número de SNPs recuperados minimizando el error. El parámetro más importante a modificar es el mínimo de cobertura para un cluster (-m) ya que afecta directamente el llamado de los marcadores (Paris *et al.*, 2017). Entre más laxo sea este parámetro es más probable que errores de secuenciación sean tomados como variación alélica, incrementando erróneamente la estimación de heterocigosis (Reyes-Galindo, 2016). Por lo que, al aumentar este parámetro, el número de SNPs recuperados disminuirá, pero se evitarán llamadas ambiguas de SNPs.

Si bien, se decidió descartar la matriz generada por *Stacks*, ya que no era suficientemente robusta, debido a que el número de marcadores obtenidos en el ensamble *de novo* era muy bajo (11 SNPs), esto no es sorprendente teniendo en cuenta el consenso de que la diversidad genética en *V. planifolia* es limitada (Hu *et al.*, 2019).

## 7.2 Variación genómica

Los valores de la heterocigosidad observada ( $H_o$ ), para cada grupo y para cada sitio, estuvieron cercanos a  $\sim 0.23$ , los cuales fueron siempre mayores a los de la heterocigosidad esperada ( $H_e$ ). Si bien existen algunas diferencias entre los individuos que conforman cada agrupación (*bicluster* o sitio de colecta), los valores de variación genómica ( $\pi$  y  $\theta$ ) fueron similares ( $\sim 0.2$ ). Estos valores se asemejan a los reportados por Villanueva-Viramontes en el 2017 ( $H_o = 0.295$ ), aunque son más bajos que los reportados por Salazar-Rojas en 2011 ( $H_o = 0.641$ ) y Hernández-Ruiz en 2015 ( $H_o = 0.78$ ) pero son más altos que los reportados por Cibrían en 2000 ( $H_o = 0$  a  $0.078$ ).

Cabe resaltar y dado que la variabilidad genética en esta especie se ha descrito a través de diferentes protocolos, no es posible la comparación directa de los resultados. Sin embargo, los estudios previos utilizaron los mismos marcadores por lo que haciendo (a groso modo) una comparación, se observa que la región del Totonacapan (Salazar-Rojas, 2011) y Oaxaca (Hernández-Ruiz, 2015) tiene mayor variación genética a lo observado en la península de Yucatán (Villanueva-Viramontes *et al.*, 2017).

En el presente estudio, se analizaron 262 marcadores moleculares tipo SNP con un valor de heterocigosidad promedio de 0.236. Es importante señalar que estos marcadores, al ser bialélicos, establecen un máximo valor teórico de heterocigosidad de 0.5 (Singh *et al.*, 2013; Pacheco-Cruz, 2019). Debido a que los SNPs presentan una alta cobertura a lo largo de todo el genoma tanto en regiones neutrales como en regiones codificantes, este tipo de marcador nos permite estimar una diversidad genética más real en la especie (Eguiarte *et al.*, 2013). Por lo que, los resultados sugieren que *V. planifolia* presenta una diversidad genética media.

Por otro lado, en todos los casos, las poblaciones o grupos asignados tienen valores de  $D < 0$  lo que indica, que las poblaciones han sufrido un cuello de botella. Esto podría ser el reflejo de efectos de flujo genético y/o eventos de deriva genética recientes, promovidos por condiciones ecológicas o antropogénicas. Sin embargo, el conjunto total de los 58 individuos muestra un valor  $D$  de 0.0947, un valor muy cercano a cero lo cual nos sugiere que la especie se encuentra bajo un modelo de neutralidad.

### 7.2.1 Estructura genética

Los datos genéticos evidencian niveles bajos de diferenciación genética entre los grupos genéticos y entre los sitios de muestreo ( $F_{ST}$ = min. 0.0317 - máx. 0.0638 y min. 0.0177 - máx. 0.0949 respectivamente); así como bajos niveles de coeficiente de consanguineidad ( $F_{IS}$ = máx. -0.0326 y min. -0.0805), sugiriendo la inexistencia de una estructura poblacional. Estos resultados fueron apoyados al determinarse un patrón similar en los análisis de estructura genética obtenidos con los modelos evolutivos (ADMIXTURE y BIC) que también señalan que la especie se comporta como una sola población.

El valor promedio del coeficiente de diferenciación genética  $F_{ST}$  en el presente trabajo fue de 0.03652. Este valor fue menor y difiere a lo reportado por Cibrian (2000), quien reportó altos niveles de endogamia y una fuerte diferenciación ( $F_{ST}$ = 0.826) entre las localidades, además de la existencia de dos grupos: 1) la región de Veracruz: Puntilla Aldama y Papantla; y 2) la región de Oaxaca: Oaxaca y Tabasco. Asimismo, Schlüter *et al.* (2007), sugieren la existencia de dos grupos genéticos de *V. planifolia* en México y uno en Centroamérica: 1) Sur de México (Oaxaca, Chiapas y Tabasco), 2) Sureste de México (Quintana Roo) y 3) Costa Rica. Por otro lado, Villanueva-Viramontes *et al.* (2017), reportan un valor medio de diferenciación genética  $R_{ST}$  = 0.28 y una alta estructura genética entre los niveles de organización observados en la península de Yucatán. Sin embargo, nuestros hallazgos no confirmaron los grupos descritos previamente a nivel de especie, ni correspondieron a la estructura de la población.

Por otra parte, tanto en el análisis de agrupación de PCA como en el DPCA se muestra que, a través de los SNPs analizados, no se observa una tendencia de ordenamiento entre los individuos de acuerdo al sitio de muestreo o al grupo genético asignado, lo que indica que estos *loci* no están diferenciados. En otras palabras, desde una perspectiva de genoma completo, los sitios de muestreo no son distinguibles entre ellos.

La estructura genética actual, los valores bajos de diferenciación poblacional observada y la diversidad genética encontrada, sugieren que el flujo genético no se ha visto limitado significativamente por la existencia de algún aislamiento reproductivo, geográfico o ambiental,



lo cual, de acuerdo a Ahrens y colaboradores (2017) es indicativo de una meta-población bien conectada.

Por último, si bien el objetivo del presente estudio no era evaluar el proceso de domesticación, a groso modo, al comparar las poblaciones cultivadas y silvestres, se observa que a nivel genómico no existen evidencias de este proceso.

### **7.3 Modelos de distribución**

#### **7.3.1 Modelos de distribución actual de *V. planifolia***

El hábitat potencial de la especie se distribuyó en dos ecorregiones: 1) Selvas cálido-húmedas, que incluye a la Planicie costera y lomeríos del Golfo de México, Planicie y lomeríos de la Península de Yucatán, Sierra de los Tuxtlas, Planicie y lomeríos del occidente, y Planicie costera y lomeríos de Soconusco. 2) Sierras templadas. Esta ecorregión incluye a la Sierra Madre Occidental, Sierra Madre Oriental, Sistema Neovolcánico Transversal, Sierra Madre del Sur, Sierra Madre Centroamericana y Altos de Chiapas ([www.conabio.gob.mx](http://www.conabio.gob.mx)). Siendo la Sierra Madre Occidental donde se tuvieron la mayor cantidad de registros.

Esta especie en condiciones naturales puede ubicarse no solo en selva alta perennifolia, como mayoritariamente nos mostró el modelo de CONABIO, si no que ha sido reportada en bosque mesófilo de montaña, bosque tropical perennifolio y bosque de pino-encino (Hernández-Ruiz, 2015). Esto es similar a lo reportado por Villanueva-Viramontes (2017) donde encontraron poblaciones silvestres de *V. planifolia* asociadas a vegetación primaria de selva alta perennifolia y selva mediana subperennifolia en el oeste y sur de la Península de Yucatán.

Si bien, los modelos consideran diferentes variables, ambos muestran una distribución similar a lo largo del área de accesibilidad planteada con las ecorregiones. Sin embargo, es importante mencionar que la calidad (precisión, sesgo, incertidumbre) entre ambos modelos es contrastante. Por parte del modelo de Worldclim, sus variables son comúnmente utilizadas para el modelado de nicho por su fácil accesibilidad y sin alguna justificación biológica, y ya que estas son obtenidas a partir de mediciones meteorológicas interpoladas mundialmente,

por lo que de acuerdo a Busby (1991) y Escobar (2016) es recomendable el uso de otras variables, con mejor ajuste a los datos y al área de estudio. Por otro lado, las variables de CONABIO, provienen del portal de geoinformación de la mejor fuente de datos nacionales (datos topográficos, climáticos y satelitales), sin embargo, la resolución de sus capas es mucho menor a la presentada en las capas de Worldclim por lo que su uso no resulta del todo adecuado cuando se pretende conocer la distribución de alguna especie con distribución restringida.

Las variables con mayor porcentaje de contribución (para ambos modelos) fueron: Temperatura media del trimestre más seco (Bio 9), Precipitación del mes más húmedo (Bio 13), Régimen de humedad, Vegetación y Precipitación. Esto coincide con lo reportado por Hernández-Ruiz (2015) y Lima-Morales (2017) quienes, mencionan que las variables que contribuyeron con mayor valor porcentual para determinar la estimación del modelo de distribución potencial en vainilla, en el estado de Oaxaca y San Luis Potosí (respectivamente) son: precipitación, régimen de humedad del suelo y precipitación del cuatrimestre más lluvioso. Las variables de precipitación, humedad y temperatura son aspectos fundamentales que controlan la distribución de las plantas y son sumamente importantes en la determinación de la abundancia de poblaciones y presencia de orquídeas (Reese, *et al.*, 2005; Zotz y Schmidt, 2006).

Los mapas de distribución del hábitat ayudan a encontrar poblaciones no exploradas, por lo que visitar los lugares con la mayor idoneidad ambiental (entropía máxima) podría resultar en la identificación de nuevas poblaciones. Así mismo, lugares con una alta idoneidad ambiental podrían resultar en sitios óptimos para la conservación *in situ*, *ex situ* o el establecimiento de nuevas plantaciones.

### 7.3.2 Distribución futura de *V. planifolia*

Las modificaciones en el área de distribución de *V. planifolia*, probablemente puedan deberse a las modificaciones en temperatura y precipitación originadas por el aumento de CO<sub>2</sub>, lo cual provocará la aparición de nuevas áreas con estas características ambientales (Parmesan, 2006).

Sin embargo, estos cambios a pesar de que resulten neutros o favorables para *V. planifolia*, pueden no ser así para especies con las que tiene una estrecha relación, debido a que las tolerancias entre las especies son distintas y esto hace que algunas sean más vulnerables al cambio climático (Thuiller *et al.*, 2011). Esta situación altera la sustentabilidad de una región debido a la modificación en la composición de especies en el ecosistema (Parmesan, 2006). Si bien las especies tienen capacidad de cambiar y adaptarse, la presencia de las poblaciones en ciertas condiciones, son un indicador de sus preferencias y requerimientos, por lo que en el caso de la especie de interés se sugiere un análisis con las especies que se relaciona en su ambiente silvestre y los tutores utilizados para su cultivo.

En el futuro, para los modelos de distribución de especies y modelos dinámicos globales de vegetación, será importante incorporar diferentes variables que permitan conocer la capacidad de respuesta de las especies del ecosistema al cambio climático, como la dispersión de las semillas, dinámica poblacional, aspectos eco-fisiológicos, biología y la diversidad genética de las especies (Dawson *et al.*, 2011). De esta manera, identificar aquellas zonas que se verán favorecidas o afectadas será imprescindible para comenzar a implementar las estrategias necesarias para asegurar la conservación de la especie como recurso fitogenético.

#### **7.4 Relación genómica-geográfica-ambiental**

En las pruebas “parcial de Mantel”, se identificó una correlación significativa entre las distancias genéticas y geográficas (*P-valor* ~0.0003). Sin embargo, el coeficiente de correlación para ambos modelos (0.3) a pesar de dar indicios sobre una estructura espacial que puede relacionarse a la distancia geográfica, al ser cercano a cero, permite suponer que la estructura genética en *V. planifolia*, no se ajusta a un modelo de aislamiento por distancia. Esto concuerda con lo mostrado en el análisis de  $F_{ST}$  pareadas (Tabla 6 y 10) y en la estructura identificada  $K = 1$  (Fig. 12). Además, es interesante señalar que no se observó una correlación entre la distancia genética y ambiental. Esto indica que la estructura espacial de las variables ambientales no explica la estructura espacial de nuestros datos genéticos (Guillot *et al.*, 2009; Garrido-Garduño y Vazquez-Dominguez, 2013).

El bajo gradiente de diferenciación genética más que estar relacionado con un patrón de aislamiento por distancias geográficas o por las diferencias ambientales, podría estar más relacionado con el movimiento histórico del germoplasma en *V. planifolia*. Es decir, este movimiento se ha llevado mediante intercambio de fragmentos de plantas (p.ej. esquejes) y/o plántulas entre productores de diferentes regiones del país, por la migración del polen mediante abejas del género *Euglossa*, quienes visitan las flores transportando el polen, y/o por la dispersión de las semillas de la vaina por murciélagos, aves o algún otro animal atraído por el aroma (Soto-Arenas, 2009; Villanueva-Viramontes *et al.*, 2017).

### **7.5 Unidades Evolutivamente Significativas (UES) y el estado de conservación de *V. planifolia***

Los beneficios de conocer la diversidad dentro de *V. planifolia*, apoyarían los programas y futuras investigaciones de fitomejoramiento y la creación de cultivares de vainilla mejorados, necesarios para enfrentar sus desafíos (alta demanda, irregularidad en la siembra, enfermedades en cultivos) y preparar a la vainilla para entrar en una era moderna de desarrollo de cultivares (Hu *et al.*, 2019). Además, el mantener la diversidad alélica en el grupo *V. planifolia*, facilita los procesos evolutivos que podrían amortiguar los impactos negativos del cambio ambiental facilitando las respuestas genéticas a un futuro incierto (Ahrens *et al.*, 2017).

*V. planifolia* es una de las orquídeas mexicanas enlistadas en la NOM-059-SEMARNAT-2010, en la lista roja de IUCN y el apéndice II del CITES, catalogada como una especie en protección especial y potencialmente amenazada. Esto se debe a la transformación de áreas naturales por diversas actividades antropogénicas lo que ha alterado de gran manera los ecosistemas. Las consecuencias inmediatas de este proceso incluyen la pérdida de hábitat, la formación de parches de hábitat remanentes y la reducción del tamaño de las poblaciones. Las poblaciones silvestres de *V. planifolia* han disminuido o incluso desaparecido en gran parte de su área de distribución debido a esta fragmentación, desaparición de su hábitat y/o a la extracción ilícita de algunos ejemplares (Soto-Arenas, 1999; Lubinsky *et al.*, 2008; Villanueva-Viramontes, 2017), encontrando un escaso número de individuos a lo largo del territorio mexicano, lo que ha llevado a contemplar la necesidad de mover a *V. planifolia* a la categoría

de “en Peligro de extinción” considerando los factores reales y potenciales que producen la disminución de los tamaños poblacionales, del número de poblaciones viables y de las áreas de distribución, así como también el estado genético, los factores que causan el deterioro y la modificación del hábitat (Villanueva-Viramontes, 2017).

Por otra parte, las unidades evolutivamente significativas (UES), son tomadas en cuenta como unidades de conservación y estas pueden llegar a comprender poblaciones, morfotipos, ecotipos, subespecies, etc. y puede ser adecuado para ayudar a la conservación de los procesos ecológicos y evolutivos de una especie (Moritz, 1994).

El establecimiento de UES puede ayudar a tomar mejores decisiones para la conservación de la diversidad genética de la especie. Y de acuerdo con los resultados en el presente trabajo, los bajos valores de diferenciación y de diversidad genética encontrada, sugieren que el flujo genético no se ha visto limitado, lo cual es indicativo de una metapoblación bien conectada, y dado el rápido declive demográfico y la falta de linajes genéticos distintos en las poblaciones de *V. planifolia*, proponemos que tratar al grupo como una UES es una estrategia menos riesgosa que usar los sitios de colecta (estados) como unidades de conservación.

Sin embargo, debe tomarse en cuenta que la diversidad morfológica y aromática, que ha sido descrita en *V. planifolia* (Lima-Morales, 2017; Salazar-Rojas, 2011), es dependiente de la diversidad genética y el ambiente. Por ello, en el futuro, un análisis que integre la diversidad morfológica, aromática, genética y ambiental, será de suma importancia para identificar el surgimiento incipiente de un mayor número de unidades de conservación que tengan base en morfotipos, quimiotipos, etc. Asimismo, la protección conjunta de ambiente y diversidad genética, será la base para la conservación de la especie. Además, en las comunidades será necesaria la capacitación para el manejo y conservación de esta especie.

## 8. Conclusiones

- Se observan niveles medios de diversidad genética para *V. planifolia*.
- La especie se comporta como una sola meta-población bien conectada.
- La especie presenta una distribución restringida en parches de hábitat remanentes, con alta especificidad hacia las variables ambientales: como la temperatura media del trimestre más seco, la precipitación del mes más húmedo, el régimen de humedad, la vegetación y la precipitación.
- Las proyecciones al futuro bajo escenarios del cambio climático, sugieren una reducción en el área de idoneidad ambiental para su permanencia.
- Los resultados sugieren que los niveles de diferenciación genética observados no se ajustan a un modelo de aislamiento por distancias geográficas ni por diferencias ambientales. No obstante, es necesario realizar un estudio más detallado de las asociaciones entre el ambiente, la variación observada y la diversidad morfológica y aromática.
- Se sugiere priorizar la conservación y monitoreo de la especie en todo su rango de distribución.

## 9. Perspectivas

Los datos genómicos obtenidos en este estudio pueden dar paso a sugerir los siguientes análisis futuros:

- Análisis de la variación genómica entre individuos silvestres y cultivados de *V. planifolia* con la finalidad de encontrar evidencias de un proceso de selección.
- Análisis de genómica funcional, por ejemplo, caracterizando genes funcionales que son críticos para la sanidad y productividad de los cultivos.
- Un estudio acerca de las relaciones filogenéticas con otros individuos del género *Vanilla*, lo cual podría brindar más información acerca del pasado evolutivo del género.
- Análisis que integre la diversidad morfológica, aromática, genética y ambiental.
- Análisis de los SNPs atípicos (SNPs de alto  $F_{ST}$ ) asociados a altitud, quimiotipos, morfotipos y/o ecotipos.

Además, es necesario continuar con el estudio de la distribución de *V. planifolia* añadiendo puntos de presencia en la zona de la Península de Yucatán. Todo esto con el fin de seguir conociendo el estado de conservación de esta especie en México. Asimismo, es necesario realizar un estudio de la distribución de la especie en todo el continente americano.

## 10. Referencias bibliográficas

- Ahrens, C. W., Supple, M. A., Aitken, N. C., Cantrill, D. J., Borevitz, J. O. y James, E. A. (2017). Genomic diversity guides conservation strategies among rare terrestrial orchid species when taxonomy remains uncertain. *Annals of Botany*. 1-11.
- Alexander, D. H., Novembre, J. y Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. 19: 1655-1664.
- Azofeifa, B. J. B., Paniagua, V. y García G. J. A. (2014). Importancia y desafíos de la conservación de *Vanilla spp.* (Orquidaceae) en Costa Rica. *Agronomía Mesoamericana* 25(1): 189-202.
- Bory S., Grisoni M., Duval M. F. y Besse P. (2007). Biodiversity and preservation of vainilla: present state of knowledge. *Genetic resources and crop evolution* 55: 551-571.
- Busby, J. R. (1991). BIOCLIM a bioclimate analysis and prediction system. *Plant Protection Quarterly* 6: 8-9.
- Caldu-Primo, J. L., Mastretta-Yanes, A., Wegier, A., y Piñero, D. (2017). Finding a needle in a haystack: Distinguishing Mexican maize landraces using a small number of SNPs. *Frontiers in genetics* 8: 45
- Cameron, K. M. (2004). Utility of plastid *psaB* gene sequences for investigating intrafamilial relationships within Orchidaceae. *Molecular Phylogenetics and Evolution*. 31: 1157–1180.
- Castillo, M. R. y Engleman, E. M. (1993). Caracterización de dos tipos de *Vanilla planifolia*. *Acta Botanica Mexicana*. 25: 49 - 59.
- Catchen, J., Hohenlohe, P., Bassham, S., Amores, A. y Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*.
- Cibrán, A. (2000). Variación genética de *Vanilla planifolia* en México. Tesis de maestría. UNAM. México.



D'Agostino, N., Taranto, F., Camposeo, S. (2018). GBS-derived SNP catalogu unveiled widw genetic variability and geographical relationships of Italian olive cultivars. Scientific Reports. 8: 15877.

Dawson, T., Jackson, S. T., House, J., Colin I., y Mace, G. M. (2011). Beyond predictions: Biodiversity conservation in a changing climate. Science. 332: 53-58

Diaz, G. F. (2010). El proceso de domesticación en las plantas. Casa del tiempo. 28(1). 66-70 pp.

Ecott, T. (2004). *Vanilla* Travels in search of the ice cream orchid. Grove Press, New York. 352.

Eguiarte, L. E., Aguirre, L. J. A., Jardón, B. L, Aguirre, P. E. y Souza, V. (2013). Genómica de poblaciones: nada en evolución va a tener sentido si no es a la luz de la genómica, y nada en genómica tendrá sentido si no es a la luz de la evolución. TIP Revista Especializada en Ciencias Químico-Biológicas. 16(1):42-56.

Elshire, R. J., Glaubitz, J. C., Sun Q., Poland, J.A., Kawamoto, K., (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6(5).

Escobar, L. E. (2016). Modelos de nicho ecológico en salud pública: cinco preguntas cruciales. Universidad de Minnesota. EUA. Revista Panamericana de Salud Pública. 40(2). 98.

FAO. (2011). El Segundo Informe sobre EL ESTADO DE LOS RECURSOS FITOGENETICOS PARA LA ALIMENTACIÓN Y LA AGRICULTURA EN EL MUNDO. Roma, Italia. 372.

Fazekas, A. J., Kesanakurti, P. R., y Burgess, K. S., (2009). Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Molecular Ecology Resources. 9: 130–139.

Flores, J. A., Reyes, L. D., Jiménez, G. D., Romero, A. O., Rivera, T. J., Huerta, L. M., Pérez, S. A. (2017). Diversidad de *Vanilla* spp. (Orchidaceae) y sus perfiles bioclimáticos en México. Revista de Biología Tropical. 65(3): 975-987.

Fouche, J. G. y Jouve, L. (1998). *Vanilla planifolia*: history, botany and culture in Reunion island. *Agronomie* 19, 689–703.

Futuyma, D. J. (2013). *Evolution*. Third edition. Sunderland, Massachusetts U.S.A: Sinauer Associates, Inc. Publishers.

Garrido-Garduño, T. y Vazquez-Dominguez, E. (2013). Métodos de análisis genéticos, espaciales y de conectividad en genética del paisaje. *Revista Mexicana de Biodiversidad* 84: 1031-1054.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* 9(2).

Gould, S.J. (2002). *The structure of evolutionary theory*. Harvard. University Press.

Guillot, G., Leblois, R., Coulon, A., y Frantz, A. C. (2009). Statistical methods in spatial genetics. *Molecular Ecology* 18: 4734-4756.

Hallgrímsson, B. y Hall, B. K. (2005) *Variation: A central concept in Biology*. Elsevier.

Hedrick, P.W. (2011). *Genetics of populatios*. Subdury. Massachusetts. Jones and Bartlett publishers. 4.a Edicion. 675.

Hernández-Ruiz, J. (2015). Caracterización del pool genético silvestre de *Vanilla planifolia* G. Jack. (Orchidaceae) en Oaxaca. Tesis de doctorado. Colegio de postgraduados. Puebla.

Hernández, R.J., Herrera, C.E., Delgado, A.A., Salazar, R.V.M., Bustamante, G.A., Campos, C.J.E. y Ramírez, J.J. (2016). Distribución potencial y características geográficas de poblaciones silvestres de *Vanilla planifolia* (Orchidaceae) en Oaxaca, México. *International Journal of Tropical Biology and Conservation*. 64 (1). 251-262.

Hernández-Santana, J. R., Ortiz-Pérez, M. A., Eng, M., y Figueroa, M. (2009). Análisis morfoestructural del estado de Oaxaca, México: un enfoque de clasificación tipológica del relieve. *Investigaciones geográficas*, 68, 7-24.

Hernandez-Silva, N. (2014). Análisis de la variación específica de *Vanilla insignis* Ames bajo el enfoque de recursos genéticos. Tesis de maestría. Colegio de postgraduados. Puebla.

Herrera, C.B.E., Hernández, R.J. y Delgado, A.A. (2016). Variación de aroma en *Vanilla planifolia* Jacks ex Andrews silvestre y cultivada. *Agroproductividad*. 9 (1), 10-17.

Herrera, C.B.E., Salazar, R.V.M., Delgado, A., Campos, C.J.E. y Cervantes, J. (2012). Use and conservation of *Vanilla planifolia* J. in the Totonacapan Region, México. *European Journal of Environmental Sciences*. 2. 43-50.

Hey, J., Waples, R.S., Arnold, M.L., Butlin, R.K., Harrison, R.G. (2003). Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology and Evolution*. 18: 597-603.

Hijmans, R.J., Cameron, S. y Parra, J. (2005). WorldClim, Versión 1.3. <http://biogeو.berkeley.edu/worldclim/worldclim.htm>.

Hu, Y., Resende Jr, M.F.R., Bombare, A., Brym, M., Bassil, E. y Chambers, A.H. (2019). Genomics-based diversity análisis of *Vanilla* species using a *Vanilla planifolia* draft genome and Genotyping-By-Sequencing. *Nature*. 9: 3416. 1-16.

Instituto Nacional de Estadística, Geografía e Informática (INEGI) -Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) - Instituto Nacional de Ecología (INE). (2008). 'Ecorregiones Terrestres de México'. Escala 1:1000000. México.

Jiménez-Valverde, A., Lobo, J. M., y Hortal, J. (2008). Not as good as they seem: the importance of concept in species distribution modelling. *Diversity and Distributions* 14: 885-890.

JMP®, Version 15. SAS Institute Inc., Cary, NC. 1989-2019.

Jombart, T., y Ahmed, I. (2011). Adegnet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27(21): 3070-3071

Kohler, H. A. (1890). *Medizinal-Pflanzen in naturgetreuen Abbildungen mit kurz erläuterndem Texte* Tomo II. Atlas zur Pharmacopoea germanica, austriaca, belgica, danica, helvetica, hungarica, rossica, suecica, Neerlandica, British pharmacopoeia, zum Codex

medicamentarius, sowie zur Pharmacopoeia of the United States of America. G. Pabst (Ed.). Editorial de Franz Eugen Köhler. Gera Unterhaus. Alemania. 200.

Lima-Morales, M. (2017). Análisis de la variación morfológica del recurso genético *Vanilla planifolia* Jacks. Ex Andrews en la huasteca potosina, México. Tesis de maestría. Colegio de postgraduados. Puebla.

Lubinsky, P., Bory, S., Hernández Hernández, J., Kim, S.-C., Gómez-Pompa, A. (2008). Origins and Dispersal of Cultivated *Vanilla* (*Vanilla planifolia* Jacks. [Orchidaceae]). *Economic Botany* 62: 127-138.

Maceda-Rodríguez, Agustín. (2015). Distribución potencial, caracterización morfológica y conocimiento tradicional de *Vanilla planifolia* J. en la región de la huasteca hidalguense, México. Tesis de maestría. Colegio de postgraduados. Puebla.

Mapes, C. y Basurto, F. (2016). Biodiversity and edible plants of México. *Ethnobotany of México*. Springer New York. 83-131.

Martínez, Q. D. M., Sandoval, Z. E., Solís, D. C. J., Velázquez, V. D. E. y Herrera, C. E. B. (2016). Variación de epidermis foliar y caulinar entre dos genotipos de *Vanilla planifolia* Jacks. ex Andrews. *Agroproductividad*. 9 (1), 26-33.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D. y Emerson, B. C. (2014). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*. 15(1): 28-41.

Miller, J. (2010). Species Distribution Modeling. *Geography Compass*. 4(6): 490-509.

Mimura, M., Yahara, T., Faith, D.P., Vázquez, D.E., Colautti, R.I., Araki, H., Javadi, F., Núñez, F.J., Mori, A.S., Zhou, S., Hollingsworth, P.M., Neaves, L.E., Fukano, Y., Smith, G.F., Sato, Y.I., Tachida, H. and Hendry, A.P. (2017). Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications Impact Factor*. 10: 121-139.

Minoo, D., Nirmal B., y Peter, K. (2006). Conservation of *Vanilla* species, in vitro. *Scientia Horticulturae*. 110: 175–180.

Morales, S.S., Becerril R. A. E., Tijerina C. L. y Santizo R. J. A. (2001). Crecimiento y desarrollo de vainilla en tres sistemas de producción en Papantla, Veracruz. *Fitotecnia mexicana*. 24(1): 49-56.

Moritz C. (1994). Defining 'evolutionarily significant units' for conservation. *Trends in Ecology and Evolution* 9: 373-375.

Nielsen, L. R. y Siegismund, H. R. (1999). Interspecific differentiation and hybridization in *Vanilla* species (Orchidaceae). *Heredity* 83, 560-567.

Noguera, S.R., Hernández, M.V.R. (2009). Variación: el universo infinito de las entidades biológicas. *Revista Digital Universitaria [en línea]* Vol. 10, No. 6 [Consultada: 11 de febrero de 2014]. In: <http://www.revista.unam.mx/vol.10/num6/art35/int35.htm> ISSN: 1607-6079.

Osorio, M A. I. (2012). Efecto de materiales orgánicos, fertilizantes e inóculos microbiales sobre el crecimiento y nutrición de plántulas de vainilla (*Vanilla planifolia* Jacks). Tesis de Maestría. Universidad Nacional de Colombia. Medellín, Colombia. 80.

Pacheco-Cruz. N. J. (2019). Variación genómica y distribución potencial de *Quercus macdougalii* Martínez, especie endémica de Oaxaca. Tesis de Maestría. Universidad Nacional Autónoma de México, México. 97.

Paris, J. R., Stevens, J. R. y Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*. 8: 1360-1373.

Parmesan, C. 2006. Ecological and evolutionary response to recent climatic change. *Annual Review of Ecology and Systematics* 37: 637-669.

Patel, D. A., Zander, M., Dalton-Morgan J. y Batley J. (2015). Advances in Plant Genotyping: Where the Future Will Take Us. In: Batley J. (ed) *Plant Genotyping (Methods and protocols)*, Humana Press.

Peña, L. (2014). Algoritmos de biclustering para el análisis de datos provenientes de microarreglos de ADN: una revisión. Universidad Autónoma de Occidente Cali, Colombia. *El Hombre y la Máquina*.44: 92-102.

Piñero, D., Barahona, A., Eguiarte, L., Rocha, A. y Salas, R. (2008). La variabilidad genética de las especies: aspectos conceptuales y sus aplicaciones y perspectivas en México. *CONABIO*. 1(1): 415-435.

Plasencia-Vázquez, A. H., Escalona, S.G., y Esparza, O. L. G. (2014). Modelación de la distribución geográfica potencial de dos especies de psitácidos neotropicales utilizando variables climáticas y topográficas. *Acta zoológica mexicana*, 30(3): 471-490.

Pliscoff, P. y Fuentes-Castillo, T. (2011). Modelación de la distribución de especies y ecosistemas en el tiempo y en el espacio: una revisión de las nuevas herramientas y enfoques disponibles. *Revista de geografía Norte Grande*. (48): 61-79.

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

R Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reese, G. C., Wilson, K. R., Hoeting, J. A., Flather, C. H. (2005). Factors affecting species distribution predictions: a simulation modeling experiment. *Ecological Applications*. 15: 554-564.

Reyes-Galindo, V. (2016). Genómica de poblaciones de *Juniperus blancoi* y *J. monticola*: diversidad genética y variación altitudinal. Tesis de Licenciatura. UNAM. 76p.

Ringnér, M. (2008). What is principal component analysis? *Nature biotechnology* 26(3): 303.

Rocha, F. G., Herrera, C. B. E., Velasco, V. J., Salazar, R. V. M., Delgado, A. A. y Mendoza, C. M. C. (2018). Determinación preliminar de componentes de rendimiento para el cultivo de vainilla (*Vanilla planifolia* Jacks. ex Andrews) en la región Totonacapan, México. *Agroproductividad*. 11(3): 9-14.

Rochette, N. y Catchen, J. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*.

Rodríguez S. B. y Armengol L. (2012). Tecnologías de secuenciación de nueva generación en diagnóstico genético pre- y postnatal. *Elsevier*. 23(2): 56- 66.

Rosales, U. L., Vélis, C. M. E. y Paleta, P. D. (2013). *Vainilla: Flor y esencia de México*. México D.F. Fundación grupo México.

Salazar-Rojas, V. M. (2011). Estrategia de uso y conservación del germoplasma de *Vanilla planifolia* G. Jack en la región Totonacapan Puebla-Veracruz. Tesis de doctorado. Colegio de postgraduados. Puebla.

Salazar-Rojas, V. M., Herrera, C. B. E., Delgado, A. A., Soto, H. M., Castillo, H. F., y Cobos, P. M. (2011). Chemotypical variation in *Vanilla planifolia* Jack. (Orchidaceae) from the Puebla-Veracruz Totonacapan region. *Genetic Resources and Crop Evolution*. 59(5): 875-887.

Scheldeman, X. y Van Zonneveld, M. (2010). Training Manual on Spatial Analysis of Plant Diversity and Distribution. *Bioersivity International*, Rome. 1-179.

Schlüter, P. M., Arenas, M. A. S. y Harris, S. A. (2007). Genetic variation in *Vanilla planifolia* (Orchidaceae). *Economic Botany*. 61: 328.

Serrano, G. R. (2003). Introducción al análisis de datos experimentales: tratamiento de datos en bioensayos. Colección Ciencias Experimentales. Publicaciones de la Universidad Jaume. Castellón de la Plana, Comunidad Valenciana, España. 193.

Singh, N., Choudhury, D. R., Singh, A. K., Kumar, S., Srinivasan, K., Tyagi, R. K., y Singh, R. (2013). Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS One* 8(12).

Soler, M. (2002). *Evolución, la base de la Biología*. Proyecto Sur de Ediciones, S.L. 553.

Soto-Arenas M. A. (1999). Filogeografía y recursos genéticos de las vainillas de México. Project J101. Informe final SNIB-CONABIO resultados: proyecto No. J101. 102 pp.

Soto-Arenas M. A. (2006). La vainilla: Retos y perspectivas de su cultivo. *Biodiversitas*. 66: 1-9.

Soto-Arenas M. A. (2009). Recopilación y análisis de la información existente sobre las especies mexicanas del género *Vanilla*. Instituto Chinoín, A.C. In: "Generación y recopilación de información de las especies de las que México es centro de origen y diversidad genética". SEMARNAT, CONABIO.

Soto-Arenas M. A. y Dressler R. L. (2010). A revisión of the Mexican and Central American species of *Vanilla plumier* ex Miller whit a characterization of their ITS region of the nuclear ribosomal DNA. *Lankesteriana*. 9(3): 285-354.

Steven, J. P., Miroslav, D., Schapire, R. E. [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: [http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/). Accessed on enero-2018.

Templeton, A. R. (2006). *Population Genetics and Microevolutionary Theory*. John Wiley y Sons, Hoboken, N.J.

Thuiller, W., Lavergne, S., Roquet, C., Boulangeat, I. y Araújo, M. B. (2011). Consequences of climate change on the Tree of Life in Europe. *Nature* 448: 550-552

Varela, S., Terribile, L. C., Oliveira, G., Diniz-Filho, J. A. F., González H., y Lima, R. M. S. (2015). ecoClimate vs. Worldclim: variables climáticas SIG para trabajar en biogeografía. *Ecosistemas* 24(3): 88-92.

Villanueva-Viramontes, S. (2017). Estructura Genética De *Vanilla planifolia* Andrews Silvestre En La Península De Yucatán, México: Implicaciones Para La Conservación De La Especie. Tesis de Doctorado. Centro de Investigación Científica de Yucatán, A.C. Yucatán.

Villanueva-Viramontes, S., Hernández, A. M., Fernández, C. G. C., Dorantes, E. A., Dzib, G. R. y Martínez, C. J. (2017). Wild *Vanilla planifolia* and its relatives in the Mexican Yucatan Peninsula: Systematic analyses with ISSR and ITS. *Botanical Sciences*. 95 (2): 169-187.



Wan, J., Wang, C., Han, S., y Yu, J. (2014). Planning the priority protected areas of endangered orchid species in northeastern China. *Biodiversity and Conservation*, 23(6), 1395-1409.

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

Zamora F. A. L., Arévalo G. L., García O. C., Ramírez G. M. R. y Valle G. S. (2016). Calidad de vainilla (*Vanilla planifolia* Jacks ex Andrews) empacada bajo diferentes películas plásticas. *Agroproductividad*. 9 (1): 18-25.

Zheng, X., Levine, D., Shen, J., Gogarten, S., Laurie, C., y Weir, B. (2012). A high-performance computing toolset for relatedness and principal component analysis of snp data. *Bioinformatics*. 28: 3326-3328.

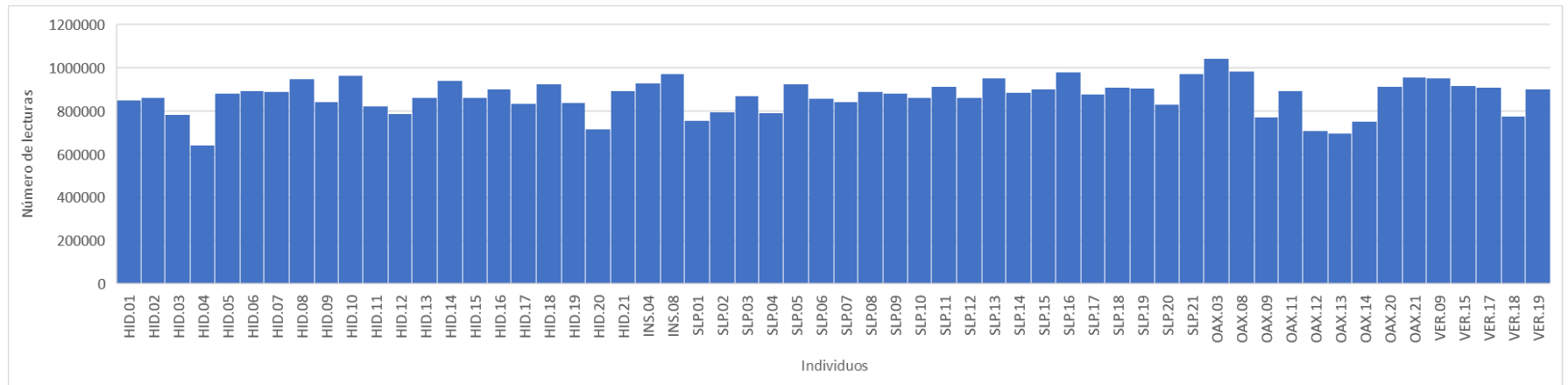
Zhou, H., Alexander, D.H., y Lange, A. (2009). A quasi-Newton method for accelerating the convergence of iterative optimization algorithms. *Statistics and Computing*.

Zotz, G. y G. Schmidt. (2006). Population decline in the epiphytic orchid *Aspasia principissa*. *Biological Conservation*. 129: 82-90.

Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A. y Smith, G.M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer. 478.

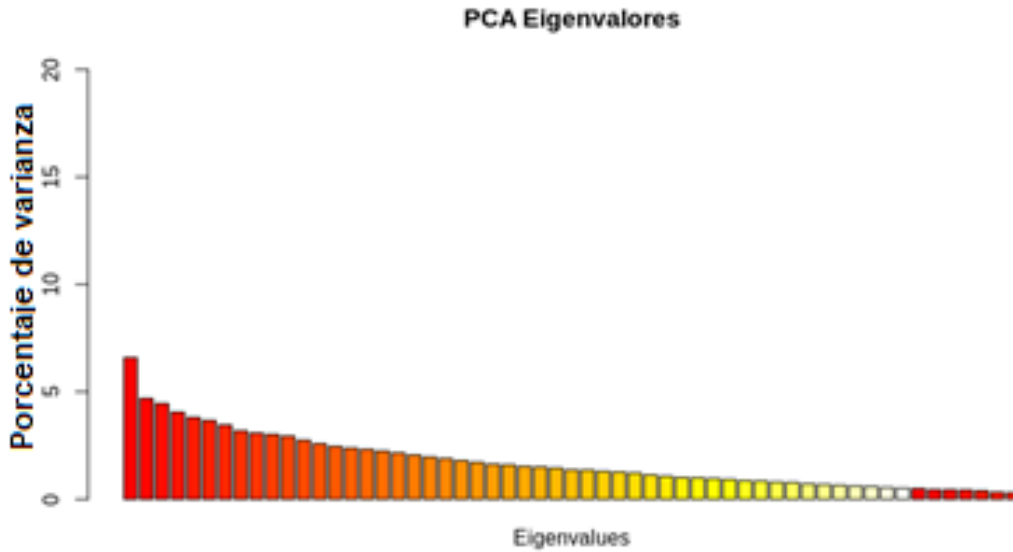
# Anexo 1

Número de lecturas obtenidas por cada individuo secuenciado de *Vanilla planifolia*.



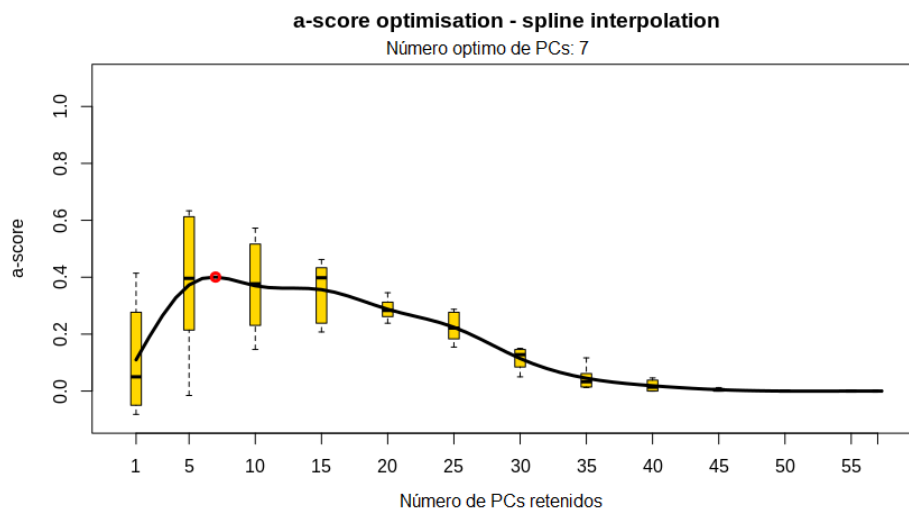
## Anexo II

Figura con el porcentaje de varianza explicada para los componentes generados por el análisis de los 252 SNPs obtenidos del mapeo al genoma de *V. planifolia*.



## Anexo III

Grafico del número óptimo de Componentes Principales (PCs) a utilizar en el Análisis Discriminante de Componentes Principales (DAPC).



## Anexo IV

Resultado del análisis de componentes principales y probabilidad de pertenencia de los individuos al sitio de muestreo. Los 2 ejes corresponden a los dos primeros componentes que explican un 12% de la variación. Los colores corresponden al sitio de muestreo.

