

Saltillo, Coahuila a 25 de agosto de 2020

Programa de Maestría en Ciencias en Agroplasticultura

Coordinación de Posgrado

P R E S E N T E

Por este conducto nos permitimos informar a esta coordinación que, el documento de tesis preparado por **ANTONNY URIEL VILLEGAS CHAVEZ** titulado "Estudio de la diversidad genética de plantasilvestres de Guayule (*Parthenium argentatum* Gray) mediante marcadores ITS" el cual fue presentado el día de hoy 25 de agosto de 2020, ha sido modificado de acuerdo a las observaciones, comentarios y sugerencias, realizadas por el Comité Evaluador asignado. Por tal motivo, avalamos que el documento corresponde a la versión final del documento de tesis.

Atentamente,

SINODALES



Dra. Ileana Vera Reyes
Presidente



Dra. Mona Kassem
Secretario



Dr. Gerardo Hernandez Vera
Vocal

Vo. Bo. de los Asesores



Dra. Ana Margarita Rodríguez Hernández



Dr. Marco Antonio Castillo Campohermoso

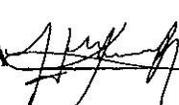


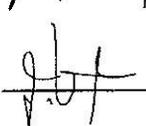
TESIS CON CARACTER ABIERTO

PROGRAMA: MAESTRÍA EN CIENCIAS EN AGROPLASTICULTURA

AUTOR: ANTONNY URIEL VILLEGAS CHAVEZ FIRMA 

TITULO: Estudio de la diversidad genética de plantassilvestres de Guayule (*Parthenium argentatum* Gray) mediante marcadores ITS

ASESORES: Dra. Ana Margarita Rodríguez Hernández FIRMA 

Dr. Marco Antonio Castillo Campohermoso FIRMA 

El Centro de Investigación en Química Aplicada clasifica el presente documento de tesis como ABIERTO.

Un documento clasificado como Abierto se expone en los estantes del Centro de Información para su consulta. Dicho documento no puede ser copiado en ninguna modalidad sin autorización por escrito del Titular del Centro de Información o del Director General del CIQA.

Saltillo, Coahuila, a 25 de Agosto de 2020


Sello de la Institución



Dr. Oliverio Santiago Rodríguez Fernández
Director General del CIQA

CENTRO DE INVESTIGACIÓN EN QUÍMICA APLICADA
Programa de Maestría en Ciencias en Agroplasticultura

TESIS

**Estudio de la diversidad genética de plantas silvestres de Guayule
(*Parthenium argentatum* Gray) mediante marcadores ITS**

Presentada por:

ANTONNY URIEL VILLEGAS CHAVEZ

Para obtener el grado de:

Maestro en Ciencias en Agroplasticultura

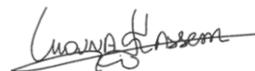
Asesorada por:

Dra. Ana Margarita Rodríguez Hernández
Dr. Marco Antonio Castillo Campohermoso

SINODALES



Dra. Ileana Vera Reyes
Presidente



Dra. Mona Kassem
Secretario



Dr. Gerardo Hernández Vera
Vocal

Saltillo, Coahuila

Agosto, 2020



CENTRO DE INVESTIGACIÓN EN QUÍMICA APLICADA
Programa de Maestría en Ciencias en Agroplasticultura

TESIS

**Estudio de la diversidad genética de plantas silvestres de Guayule
(*Parthenium argentatum* Gray) mediante marcadores ITS**

Presentada por:

ANTONNY URIEL VILLEGAS CHAVEZ

Para obtener el grado de:

Maestro en Ciencias en Agroplasticultura

Asesorado por:

Dra. Ana Margarita Rodríguez Hernández
Dr. Marco Antonio Castillo Campohermoso

Saltillo, Coahuila

Agosto, 2020

Fe de erratas

Por medio de la presente se hace constar que en los documentos donde aparece el título del documento de tesis de ANTONNY URIEL VILLEGAS CHAVEZ debe decir:

- “Estudio de la diversidad genética de plantas silvestres de Guayule (*Parthenium argentatum* Gray) mediante marcadores ITS”, y no
- “Estudio de la diversidad genética de plantassilvestres de Guayule (*Parthenium argentatum* Gray) mediante marcadores ITS”

Atentamente,



Dr. Antonio Cárdenas Flores
Coordinador de Posgrado en Agroplasticultura

Saltillo, Coahuila a 02 de septiembre de 2020



DECLARACIÓN

Declaro que la información contenida en la Parte Experimental así como en la Parte de Resultados y Discusiones de este documento y que forman parte de las actividades de investigación y desarrollo realizadas durante el período que se me asignó para llevar a cabo mi trabajo de tesis, será propiedad del Centro de Investigación en Química Aplicada.

Saltillo, Coahuila a 25 de agosto de 2020



ANTONNY URIEL VILLEGAS CHAVEZ

Nombre y Firma

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por brindarme el apoyo económico durante la maestría; lo que me permitió cumplir con una meta más en el ámbito académico.

Al Centro de Investigación en Química Aplicada (CIQA) que me abrió las puertas y me brindó las herramientas para finalizar mis estudios.

A mis asesores la Dra. Ana Margarita Rodríguez Hernández y Dr. Marco Antonio Castillo Campohermoso por el apoyo y la confianza brindada a lo largo de este proyecto.

A mis sinodales Dra. Ileana Vera Reyes, Dr. Mona Kassem y Dr. Gerardo Hernández Vera por las valiosas aportaciones recibidas para la mejora de este documento.

Al la Universidad de Guadalajara por abrirme sus puertas lo que permitió llevar a cabo gran parte de este proyecto con un gran e invaluable apoyo del Dr. Gerardo.

A mi familia que siempre me alentó y jamás permitió que me rindiera.

A los amigos que conocí durante la maestría que permitieron que este proceso fuera más ameno.

A Naza, gracias por abrirme las puertas de tu casa y corazón, gracias por todo a ti y tu hermosa familia que me abrazo como un integrante más de ella.

A mi madre, gracias a ti por todo, gracia por siempre ser y estar sin importar que, este logro también es tuyo por estar a mi lado y nunca soltarme. TE AMO.

“Nada en la vida es para ser temido, es sólo para ser comprendido. Ahora es el momento de entender más, de modo que podamos temer menos”

Marie Curie

Resumen

El guayule (*Parthenium argentatum*) es una especie perenne nativa del desierto chihuahuense de México, que se utiliza como materia prima para la producción de hule natural (HN), látex y productos bioenergéticos. Sin embargo, la falta de recursos de germoplasma bien caracterizados y con alta productividad (concentración de hule mayor al 17% en peso), aunado a la poca información de la diversidad genética de *P. argentatum* que permitan un avance en la obtención de variedades altamente productivas, se ha convertido en una limitante que impide el desarrollo de la industria del hule de guayule en México. Por lo tanto, en la presente investigación, se estudió la diversidad genética de 160 ejemplares silvestres de *P. argentatum*, pertenecientes a cuarenta poblaciones individuales, de 5 estados (Nuevo León, Coahuila, Zacatecas, Durango y San Luis Potosí) usando el marcador del espacio interno transcrito (ITS). Se amplificó la región ITS1-ITS2 obteniéndose secuencias de una longitud de 649 pb, con una composición nucleotídica de A: 24.1 %; G: 26.5 %; C 25.8 % y T: 23.5 %. Esta región ITS1-ITS2, presentó 152 sitios polifórmicos, 27 haplotipos únicos con una alta diversidad ($H_d: 0.738 \pm 0.013$), una diversidad nucleotídica (P_i) de 0.0027 ± 0.0094 y una distancia genética de 0.0092. La variación genética encontrada dentro de cada población (81.19 %) fue mayor que la encontrada entre las poblaciones de cada estado (16.55 %) y la encontrada entre estados (2.26 %) según mostraron los resultados del análisis de varianza molecular (AMOVA). Este resultado nos indica que el flujo de genes de las poblaciones silvestres dentro del desierto chihuahuense es continuo ya que no hubo una correlación obvia entre la distancia genética y la distancia geográfica. Se encontraron poblaciones importantes a conservar de acuerdo con la información genética silvestre encontrada como lo son las poblaciones SL6 y SL9 en San Luis Potosí, Z5 en Zacatecas y NL2 en el estado de Nuevo León.

Índice de contenido

I. ANTECEDENTES	1
I.1 El Guayule	1
I.2 Importancia económica del hule natural	2
I.3 Industrias co-ubicadas e importancia industrial del hule de guayule	5
I.4 Diversidad genética en la explotación comercial de cultivos	7
I.5. Diversidad genética del guayule	12
I.6. Marcadores moleculares para la diversidad genética	13
I.7 El enfoque filogenético en estudios de diversidad y estructura genética	17
I.8 Análisis bioinformáticos.	18
I.8.1 Alineamiento de secuencias	18
I.8.2 Inferencia filogenética	20
I.8.2.1 Inferencia filogenética basada en máxima verosimilitud (ML)	21
I.8.2.2 Inferencia filogenética basada en métodos bayesianos (IB)	21
I.8.3 Modelos de evolución de ADN	22
II. HIPÓTESIS	25
III. OBJETIVOS	26
III.1 Objetivo general	26
III.2 Objetivos específicos	26
IV. MATERIALES Y MÉTODOS	27
IV.1 Colecta de material vegetal	27
IV.2 Extracción de ADN	28
IV.3 Diseño de cebadores	29
IV.4 Amplificación de las regiones ITS1-ITS2 en ADN ribosómico nuclear	29
IV.5 Purificación y cuantificación de productos de PCR	30
IV.6 Secuenciación de fragmentos amplificados	30
IV.7 Análisis bioinformáticos	31
IV.7.1. Estadísticas descriptivas de las secuencias y estimación de índices de diversidad genética	32
IV.7.2. Reconstrucción filogenética con enfoque bayesiano	32
IV.7.3. Reconstrucción filogenética basada en máxima verosimilitud (ML)	33
IV.7.4. Análisis molecular de varianza (AMOVA)	33
IV.7.5. Construcción de red de haplotipos	33
IV.7.6. Estimación de tiempos de divergencia	34

V. RESULTADOS Y DISCUSIÓN	35
V.1 Colecta de material vegetal	35
V.2 Extracción de ADN	37
V.3 Diseño de cebadores y amplificación de las regiones ITS1-ITS2 en plantas silvestres de guayule	37
V.4. Análisis bioinformáticos	40
V.4.1. Estadísticas descriptivas de las secuencias e índices de diversidad genética	40
V.4.2 Reconstrucción filogenética con inferencia bayesiana (IB) y Máxima Verosimilitud (ML)	46
V.4.3 Análisis AMOVA para plantas silvestres de guayule	51
V.4.3 Red de haplotipos en plantas silvestres de guayule	51
V.4.4. Estimación de tiempos de divergencia	54
VI. CONCLUSIONES	55
Anexos	56
VII. BIBLIOGRAFÍA	73

Índice de figuras

Figura 1. Planta de Guayule (<i>Parthenium argentatum</i>).	1
Figura 2. Área del desierto chihuahuense en México y Estados Unidos.	3
Figura 3. Producción mundial de hule natural.	4
Figura 4. Reemplazo nucleotídico entre bases nitrogenadas en ADN.	10
Figura 5. Maduración de ARNr.	16
Figura 6. Distribución de diámetros medidos en campo de plantas de <i>P. argentatum</i> . .	28
Figura 7. Región de interés en el diseño de cebadores específicos para Asteraceae en ADNr.....	29
Figura 8. Temperaturas máximas promedio para Septiembre 2018.....	36
Figura 9. Temperaturas máximas promedio para Enero 2019.	36
Figura 10. Extracción de ADN genómico total de hojas de <i>P. argentatum</i>	37
Figura 11. Área de diseño de cebadores para amplificar ITS1-ITS2 en la familia Asteraceae.	38
Figura 12. Amplificación de la región ITS1-ITS2 en <i>P. argentatum</i>	39
Figura 13. Purificación de la amplificación de ITS's.	39
Figura 14. Comparativo de distintos índices de diversidad genética entre las regiones ITS1-ITS2 en <i>P. argentatum</i> . Diversidad nucleotídica por sitio (Pi), diversidad nucleotídica (K), diversidad de haplotipos (Hd), sitios variables (S), número de haplotipos (h). ¡Error! Marcador no definido.	
Figura 15. Relación y divergencia entre haplotipos presentes en plantas de guayule silvestre.	46
Figura 16. Reconstrucción filogenética con Inferencia Bayesiana y Máxima verosimilitud de tres especies del género <i>Parthenium</i> ¡Error! Marcador no definido.	
Figura 17. Red de haplotipos con grupos externos (<i>P. incanum</i> y <i>P. rollinsianum</i> , círculos violetas y morados respectivamente). ¡Error! Marcador no definido.	

Índice de tablas

Tabla 1. Marcadores utilizados en genética de poblaciones.	15
Tabla 2. Código de ambigüedades nucleotídicas descrito por la IUPAC.....	31
Tabla 3. Número de poblaciones e individuos de <i>P. argentatum</i> muestreados por estado.....	35
Tabla 4. Estadísticas descriptivas e índices de diversidad genética estimados de manera independiente a partir de cada región ITS y datos concatenados para la especie <i>P. argentatum</i>	42
Tabla 5. Índices descriptivos estimados para poblaciones de guayule silvestres muestreadas para cada estado.	45
Tabla 6. Análisis AMOVA para poblaciones silvestres de guayule.....	51

I. ANTECEDENTES

I.1 El Guayule

El guayule (*Parthenium argentatum*) (Figura 1), es un arbusto leñoso dióico perennifolio, crece formando rodales de forma natural y pertenece a la subtribu *Heliantheae* de la familia *Asteraceae* (Kumar *et al.*, 2009); en promedio tiene una altura de 30 a 100 cm, crece en altitudes que van desde los 600 a 2000 msnm (metros sobre el nivel del mar), se desarrolla con bajas precipitaciones anuales que van de los 250 a los 380 mm, originario de zonas desérticas con climas áridos y semiáridos del norte de México (Chihuahua, Coahuila, Nuevo León, Durango, Zacatecas y San Luis Potosí) y sureste de Estados Unidos de América, en el área del Big Bend (Pecos, Presidio y Brewster). El guayule crece en suelos calcáreos con baja fertilidad y bien drenados, esta planta tolera temperaturas desde -18 hasta 49.5 °C, sin embargo, a temperaturas iguales o menores de 4 °C entra en un estado de inactividad (Veatch-Blohm *et al.*, 2006, Foster *et al.*, 2011, Bekaardt *et al.*, 2010). A partir del año 1852 que el guayule fue descubierto, ha tenido un largo historial de investigación por las inigualables propiedades que éste presenta en la industria (Cruz *et al.*, 2017).



Figura 1. Planta de Guayule (*Parthenium argentatum*).

Esta planta presenta diferentes niveles de ploidía, es decir, sus células presentan diferentes números de juegos de cromosomas, se conoce que hay ejemplares diploides ($2n = 2x = 36$), triploides ($3n = 3x = 54$), y hasta octaploides ($8n = 8x = 144$). Dependiendo del nivel de ploidía, las plantas tienen un método de reproducción diferente, donde,

plantas diploides presentan reproducción sexual mientras que las plantas poliploides realizan una reproducción por apomixis facultativa (reproducción asexual por semilla); sin embargo, se sabe que presentan periodos de recombinación sexual o pueden ser altamente sexuales incluso a niveles altos de ploidia (Sanchez *et al.*, 2014); el conocimiento de esta información es crítica para los avances en el desarrollo de nuevas líneas de germoplasma del guayule, a pesar de que se tiene diferentes germoplasmas disponibles del banco de germoplasma del departamento de agricultura de los Estados Unidos (USDA, por sus siglas en inglés), solo hay dos accesiones diploides y 32 poliploides disponibles (Gore *et al.*, 2011), las cuales presentan una variación genética muy escasa (Cruz *et al.*, 2017, Ilut *et al.*, 2015, Ray *et al.*, 2005).

I.2 Importancia económica del hule natural

En México, el guayule se encuentra de forma natural distribuido en más de 5 millones de hectáreas, específicamente en el desierto chihuahuense (Figura 2), representando más del 90 % del germoplasma silvestre mundial (de Rodríguez *et al.*, 2005, Rasutis *et al.*, 2015, Foster *et al.*, 2011). El guayule es de gran interés industrial ya que, es una fuente renovable de hule natural (HN) (*cis*'1,4-poliisopreno), tiene la capacidad de biosintetizar hule en las células del parénquima, se conocen más de 2,000 especies diferentes de plantas productoras de hule natural, sin embargo, el hule producido por el guayule se denomina hipoalergénico, también presenta otras características como lo es la impermeabilidad viral, fuerza, elasticidad y presenta un alto peso molecular (≥ 1500 kDa), lo cual lo hace útil en aplicaciones médicas tales como producción de condones, guantes quirúrgicos y catéteres en donde se requiere de una alta durabilidad y mínima transmisión de enfermedades (Ray *et al.*, 2005, Bekaardt *et al.*, 2010).

Además de la industria médica, el hule natural es de gran importancia para la sociedad moderna ya que presenta utilidad como materia prima en un amplia gama de industrias como lo es la automotriz, electrónica y Textil (Ilut *et al.*, 2017). El guayule puede considerarse como la mejor fuente sustentable de hule natural por tener un menor impacto ambiental en la cadena de producción, en términos de sustentabilidad el guayule es un cultivo "multiusos" pudiéndose denominar cultivo de "cero desperdicios", adicionalmente a las aplicaciones reportadas para el hule; con las resinas del guayule se

pueden hacer preservantes para maderas, pesticidas y plastificantes; los residuos de la extracción se pueden emplear como biocombustible, o en mezclas con otras fibras para producir papel, entre otras aplicaciones (Rasutis *et al.*, 2015, Sabaini *et al.*, 2018).



Figura 2. Área del desierto chihuahuense en México y Estados Unidos (Guevara 2015).

La producción de HN ha crecido exponencialmente cada año (Figura 3), hasta el 2017 se produjeron mundialmente 14,252,753 toneladas de HN en un total de 11,739,333 hectáreas, teniendo un rendimiento de 1.2141 toneladas por hectárea cultivada donde, el 89 % de esta producción se realiza en Asia, de acuerdo con datos obtenidos de la FAO (<http://www.fao.org/faostat/es/#data/QC>), teniendo así una dependencia a la importación del HN de los países asiáticos productores; en 2018 Tailandia e Indonesia fueron los países con mayores ingresos económicos por la exportación de HN, con \$ 4,475,208,895 y \$ 3,949,214,198 USD respectivamente, los países con mayor demanda de HN fueron Estados Unidos, India y Japón con una inversión de \$ 1,715,131,737, \$ 1,093,479,090 y \$ 1,073,378,487 USD respectivamente, por la importación de 990,302.732, 714,117.093 y 646,064.482 toneladas respectivamente para cada país según estadísticas de la ONU (<https://comtrade.un.org/data/>), para México, en 2018 se tienen datos de importación de HN para 14,134 toneladas con un costo de \$ 75,855,261 USD, para este mismo año se presentó un ingreso de \$ 3,168,662 USD por exportar cerca de 1,322 toneladas. Para Estados Unidos se tienen reportes que, del total de HN importado, el 78 % de este es destinado a la producción de neumáticos (IRSG 2015).

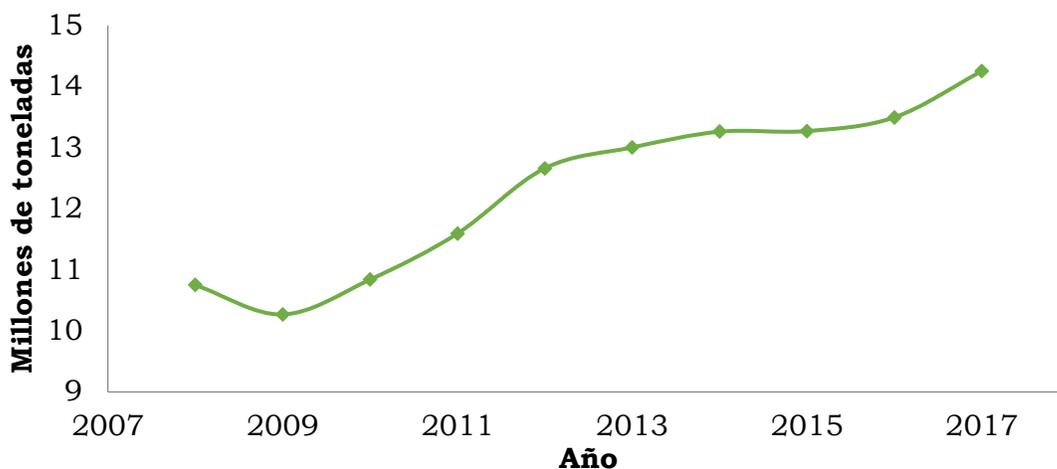


Figura 3. Producción mundial de hule natural (FAO).

En el año 2017 la derrama económica del HN en el mercado de las materias primas fue de 20 mil millones de dólares a nivel mundial. Sin embargo, actualmente debido a una gran incertidumbre causada principalmente por aspectos sociopolíticos y la amenaza de una plaga fúngica, se pone en riesgo el abastecimiento al mercado mundial de hule natural producido 100 % a partir de *H. brasiliensis* (Barrera y Cornish 2016). Con lo anterior la importancia del guayule en el mercado internacional del HN se pone de manifiesto. Sugiriendo una inminente producción industrial del cultivo de guayule y por consiguiente colocando al cultivo de guayule como una alternativa muy prometedora para la producción de HN, donde México tiene una gran área de oportunidad.

Sin embargo, a pesar de todas estas ventajas presentes en el mercado, la producción masiva del HN a partir del guayule sigue siendo limitada principalmente por tres razones, 1) por su bajo contenido de hule en la planta (5 a 10 % de la materia seca) (Veatch *et al.*, 2005, Sabaini *et al.*, 2018); 2) por su formación al interior en las células del parénquima de la corteza y por tanto la necesidad de cortar la planta para su extracción, lo cual encarece su costo (Sabaini *et al.*, 2018) y 3) por el crecimiento de la planta la cual necesita de 2 a 3 años antes de alcanzar su máximo contenido en hule. Estas limitantes nos indican que la explotación del potencial del guayule reside principalmente en poder aumentar su velocidad de crecimiento y/o aumentar la capacidad de la planta en sintetizar

el hule, así como innovar en los métodos de extracción de hule (Rasutis *et al.*, 2015, Ray *et al.*, 2005).

I.3 Industrias co-ubicadas e importancia industrial del hule de guayule.

El guayule tiene la capacidad de ser una alternativa viable en la producción de HN, ya que la calidad que este presenta es comparable con el producido por *Hevea brasiliensis* la cual es la única fuente de HN comercial, con la ventaja de que el HN biosintetizado por guayule no causa alergia al látex tipo I en comparación con el producido por Hevea (Sabaini *et al.*, 2018).

Después de la extracción de HN de la planta de guayule, la biomasa residual se le determina bagazo, la explotación integral del cultivo de guayule incluye el aprovechamiento del bagazo, el cual se ha utilizado para una gran variedad de industrias. Una de estas industrias es la de bioenergía, ya que el bagazo es rico en terpenos contenidos en la resina (16, 24-epoxy-3 α -hydroxylanost-8-ene (argentatina E); 16, 24-epoxi-25-hidroxicicloart-1, 11, 22-trien-3-ona (argentatina F); 16,24-dihydroxycycloart-20, 25-dien-3-one (argentatina G) y 16, 24-dihydroxycycloart-25-en-3-one (argentatina H)) (Maatooq *et al.*, 2002) y adicionalmente, contiene residuos de HN los cuales tienen un alto poder calorífico. La presencia de estos compuestos en el bagazo del guayule le confiere una ventaja en la producción de combustibles químicos con mejor calidad convirtiéndolo en una fuente de alimentación para una bio-refinería de pirólisis co-ubicada para la producción de hidrocarburos renovables (Chundawat *et al.*, 2012, Boateng *et al.*, 2015, Sabaini *et al.*, 2018).

Actualmente el guayule está siendo producido y explotado a escala comercial por la empresa *Yulex Corporation* (<http://yulex.com/>) en EE. UU. También la compañía productora de neumáticos más grande del mundo *Bridgestone Corporation*, a finales del año 2015 anunció su producción de neumáticos para automóviles fabricados con un 100 % de componentes a partir de HN derivados del guayule, fabricados en el Centro Técnico de *Bridgestone* en Japón (<https://noticias.autocosmos.com.mx/2015/10/13/bridgestone-produce-llantas-con-caucho-natural-del-guayule>), por lo que en 2018 creó vínculos con la empresa *Versalis* con el objetivo de complementar su investigación, esta asociación tienen el objetivo de producir neumáticos a partir de HN del cultivo de guayule a partir del

2020 (<https://www.bridgestoneamericas.com/en/newsroom/press-releases/2018/bridgestone-and-versalis-partnering-to-expand-guayule-research-e>).

En 2013 la empresa internacional *Pirelli* firmó un acuerdo con la empresa *Versalis* para un suministro exclusivo de HN de guayule, que se usaría para fabricar neumáticos, actualmente la empresa internacional ya produce neumáticos a partir de HN de guayule, lo describen como un “*neumático de rendimiento ultra-alto*” por lo que lo hace un cultivo alternativo viable al uso de la planta de *H. brasiliensis* y la sustitución de polímeros petroquímicos en sus neumáticos (<https://www.pirelli.com/tyre/ww/en/news/2016/02/03/pirelli-makes-the-first-guayule-based-uhp-tyre/>).

Cooper Tire & Rubber Company y socios del consorcio, incluyendo la Universidad de Clemson, la Universidad de Cornell, *PanAridus* y el Servicio de Investigación Agrícola del Departamento de Agricultura de los Estados Unidos (USDA-ARS por sus siglas en inglés) en agosto del 2017 anunciaron la subvención por cinco años por la cantidad de \$6.9 millones de dólares para la Iniciativa de Investigación y Desarrollo de Biomasa (BRDI, por sus siglas en inglés), este consorcio de empresas destinaron la subvención para “*Asegurar el futuro del caucho natural, una plataforma estadounidense de bioenergía y neumáticos de Guayule*”, *Cooper Tire* es la encargada de producir los neumáticos, actualmente se encuentran haciendo pruebas a los neumáticos ya producidos donde, se encontró que tenían un rendimiento general al menos igual a los neumáticos fabricados con *H. brasiliensis* y caucho sintético, en particular, los neumáticos tuvieron un desempeño significativamente mejor en resistencia a la rodadura, manejo en mojado y frenado en mojado que sus contrapartes convencionales. *Chuck Yurkovich*, Vicepresidente Senior de Investigación y Desarrollo Global de *Cooper* comentó en una entrevista “*En base a nuestros hallazgos, Cooper podría usar caucho de guayule en la producción de neumáticos mañana si hubiera suficiente material disponible para satisfacer nuestras necesidades de producción a un precio competitivo*”, además de los análisis ya mencionados, se realizó un análisis del ciclo de vida (LCA) en la *Universidad de Clemson* para cuantificar el impacto ambiental en la producción donde encontraron que los neumáticos 100 % de HN de guayule presentan entre 6-30 % menos emisiones en 10 categorías diferentes de impacto ambiental y energético en el ciclo de vida en

comparación con un neumático convencional.([http://coopertire.com/News/ Corporate-News-Releases/Cooper-Tire-and-Consortium-Partners-Complete-\\$6-9.aspx](http://coopertire.com/News/Corporate-News-Releases/Cooper-Tire-and-Consortium-Partners-Complete-$6-9.aspx)).

La empresa finlandesa, *Nokian Tyres* (<https://www.nokiantyres.com/>) ha establecido una plantación de guayule en Santa Cruz de la Zarza, Toledo, España. Poniendo en marcha un proyecto centrado en el cultivo de guayule para la fabricación de neumáticos. Los cultivos de guayule en este proyecto son gestionados por agricultores locales, con el fin de conocer su cultivo y su adaptación a las condiciones climáticas y edáficas de la zona. La inversión total en la nueva instalación es de 60 millones de euros, actualmente este proyecto está en desarrollo y tiene un plazo estimado de ejecución de dos años para finalizar en 2020. Dentro de este proyecto se está construyendo un centro de prueba con una extensión de 300 hectáreas donde, se desarrollarán y lanzarán productos nuevos a base de guayule ([www.infotaller.tv/ neumaticos/pruebas-Nokian-Tyres-contar-velocidad_2_1342985689.html](http://www.infotaller.tv/neumaticos/pruebas-Nokian-Tyres-contar-velocidad_2_1342985689.html)).

I.4 Diversidad genética en la explotación comercial de cultivos

La diversidad genética en cultivos es uno de los recursos más importantes para la vida humana y la conservación, estos son esenciales para el desarrollo de la producción agrícola e industrial, de ahí la importancia del conocimiento de la diversidad genética de los cultivos (Sudha *et al.*, 2019).

La diversidad genética es un componente fundamental de la diversidad biológica; el preservar la diversidad genética de las especies de plantas puede afectar significativamente su supervivencia y evolución a largo plazo, por lo tanto, el conocimiento de la diversidad genética, la estructura poblacional de las especies vegetales, entender cómo las fuerzas evolutivas y los sistemas reproductivos actúan sobre variantes alélicas o genotípicas en distintos ambientes es esencial para su manejo (Peng *et al.*, 2018, Golkar y Nourbakhsh, 2019).

Todos los organismos de manera natural se desarrollan en forma de poblaciones, estas tendrán propiedades que trascienden las características de un individuo, el estudio de la genética de poblaciones se dirige a la indagación de la estructura genética de las mismas y los procesos por los que la estructura cambia; al referirse a la estructura genética se

hace referencia a los tipos y las frecuencias de los genes y los genotipos presentes en una población.

La diversidad genética puede ser definida por diversos factores, estos, deberán exponer datos cuantitativos en cuanto a la variación genética dentro de una población, y deben reflejar una relación entre la mutación y la pérdida de variación genética (Carvalho *et al.*, 2019).

Se pueden realizar diferentes análisis para estimar la diversidad genética de una especie, el primer grupo de análisis están enfocados a estados alélicos discretos, donde, los principales parámetros son: la diversidad alélica, riqueza alélica y genotípica, heterocigocidad, diversidad mutacional y tamaño efectivo de la población (θ), diversidad de nucleótidos (P_i) y porcentaje de loci polifórmicos, el segundo grupo de análisis están enfocados a rasgos continuos como lo son el coeficiente de variación genética (CV), varianza genética (V_G) y la heredabilidad. El uso de cada uno de los enfoques y análisis para determinar la diversidad genética dependerá de la naturaleza de los datos y las metodologías utilizadas para dicho estudio (Hughes *et al.*, 2008).

La caracterización y entendimiento de la diversidad genética entre los genotipos y la estructura de la población, así como, las relaciones entre los recursos de germoplasma es un paso clave para conservar las especies de interés, ya que ha sido durante mucho tiempo un objetivo importante para el mejoramiento de los cultivos, debido a la información altamente discriminatoria y la independencia de la etapa de desarrollo y los factores ambientales (Zhang *et al.*, 2014). La correlación de información es la base para elaborar estrategias de conservación e identificar áreas clave de protección, la informática de correlación también proporciona a fitomejoradores una capacidad sustentable para utilizar el germoplasma silvestre de manera efectiva (Wang *et al.*, 2017); ya que se sabe que los niveles la variación genética pueden en muchos casos dar información sobre si las poblaciones en estudio tienen diversidad suficiente para enfrentar riesgos bióticos y abióticos como lo es el constante cambio climático que se tiene en la actualidad; la amenaza que presentan poblaciones silvestres ya sea por el pastoreo o la constante urbanización, por lo que es de gran importancia estudiar y conservar germoplasmas silvestres de plantas de interés comercial como lo es el guayule.

Se ha demostrado que dentro de la agronomía, el tener un cultivo diversificado con respecto a las variedades de una misma especie ha reducido la utilización de agroquímicos como los pesticidas, además que se han encontrado aumentos significativos en el rendimiento de producción (Zhu *et al.*, 2000).

La producción de especies cultivables con potencial de explotación industrial depende principalmente de, la comprensión de la organización y los alcances de las variaciones nucleotídicas en su genoma; este tipo de estudios se realizan principalmente para la comprensión de las relaciones filogenéticas de los diferentes ejemplares y para poder elegir de manera adecuada y sistémica los ejemplares de interés (Agarwal y Khan 2019, Zhai *et al.*, 2018).

La biología de la conservación y campos relacionados con la genética de la conservación, han desarrollado investigaciones sobre posibles efectos ecológicos a corto plazo de la diversidad genética, particularmente en poblaciones pequeñas en peligro de extinción ó de interés industrial. A menudo la investigación se centra en las relaciones entre el tamaño de la población y la diversidad genética, estudios ilustran que los cuellos de botella genéticos (descenso drástico en el número de miembros en algún momento del pasado, llegando a estar a punto de la extinción) y la endogamia (apareamiento de individuos que están estrechamente relacionados genéticamente), pueden alterar la diversidad genética, con consecuencias ecológicas como la reducción de la capacidad de una población para persistir en ambientes cambiantes o estresantes (Hughes *et al.*, 2008); el avance tecnológico en genética molecular ha determinado el surgimiento y el desarrollo de estas disciplinas, considerando estos avances, la ecología molecular ha sido definida actualmente como una ciencia interdisciplinaria en la que las herramientas y métodos de la biología molecular, la genómica y la bioinformática se han unificado (Rodríguez-correa *et al.*, 2017).

El proceso de evolución de la estructura genética de toda población se puede cuantificar mediante diferentes factores básicos como lo son: la mutación, la duplicación de genes, la selección natural, el cambio genético aleatorio, reemplazo de un nucleótido, eliminación de nucleótidos, adición de nucleótidos, e inversión de nucleótidos; los reemplazos nucleotídicos pueden ser divididos en dos clases: transición y transversión

(Figura 4) donde, la transición es el reemplazo de una purina (A: adenina o G: guanina) por otra purina ó de una pirimidina (T: timina o C:citocina) por otra pirimidina y la transversión resulta ser el reemplazo de una purina por una pirimidina, tales procesos permite evaluar los cambios evolutivos de las poblaciones y probar la validez de los supuestos sobre la evolución o la estabilidad de los sistemas genéticos (Leduc y Zhao, 2019).

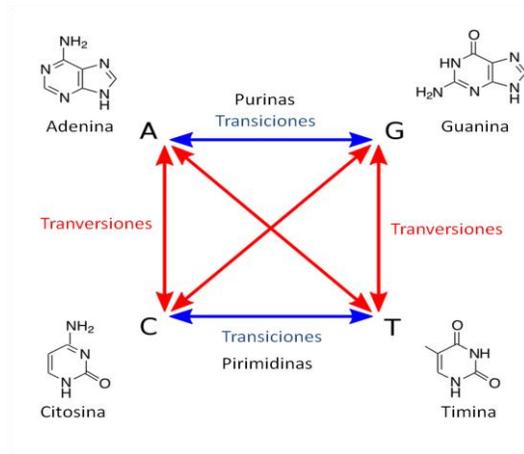


Figura 4. Remplazo nucleotídico entre bases nitrogenadas en ADN (tomado de internet)

La genética de poblaciones explora los niveles de variación genética dentro y entre las poblaciones de interés, además, explica sus patrones en términos de las fuerzas evolutivas; las especies se conforman por un conjunto de poblaciones conectadas mediante un flujo genético; si las poblaciones tienen altos niveles de diversidad genética, implica que son o han sido grandes o que existe elevado flujo genético entre ellas, si al contrario, las poblaciones que forman una especie son muy diferentes entre sí, es posible que sean pequeñas o que hay poco flujo genético entre ellas, entonces, la genética de poblaciones busca entender los procesos evolutivos a una escala ecológica, misma que implica relativamente pocas generaciones, en otras palabras, cambios a corto plazo, y cómo las diferentes fuerzas evolutivas afectan a esta variación (Eguiarte *et al.*, 2013).

En la última década se han incrementado los estudios a nivel molecular de las especies con interés de explotación industrial, algunos de los estudios realizados comúnmente son, estudios filogenéticos, la indagación en la estructura genética de las poblaciones, la identificación de unidades a conservar y la especificación de patrones filogeográficos

mediante acervos genéticos, esto, con el fin de formar grupos genéticos de poblaciones altamente eficientes (Mayonde *et al.*, 2019, Carvalho *et al.*, 2019).

La utilización de los acervos genéticos silvestres ayudan como recurso eficiente en el desarrollo de nuevas líneas de interés, estos, tienen la capacidad de aportar características como lo es la floración tardía, la autofertilidad, la resistencia a la sequía y la salinidad, bajas temperaturas invernales y la resistencia a estrés bióticos y abióticos en general (Zeinalabedini *et al.*, 2019), por lo que el conocimiento de la diversidad genética es imprescindible para adoptar estrategias apropiadas para su mejoramiento genético, restauración, conservación y manejo (Long *et al.*, 2010, Mayonde *et al.*, 2019).

La reciente utilización de datos de biodiversidad (espacial y genética), datos ambientales (con respecto al clima, el suelo y la vegetación), junto con nuevos métodos y herramientas analíticas, ha permitido una capacidad sin precedentes para modelar las distribuciones de especies y ensamblar esos resultados en gran escala (Kosanovic *et al.*, 2018).

En estudios recientes se ha generado evidencia de que el ambiente puede tener un papel importante en la emergencia de nuevos fenotipos así como su evolución, en muchos casos, esto sucede mediante cambios epigenéticos (cambios heredables en el ADN que afectan la expresión génica y el fenotipo) , de manera similar que la variación genética, la variación epigenética (variación de factores no genéticos producidos por cambios en el ambiente celular) impacta a la evolución fenotípica, e incluso fijándose en linajes o especies (Álvarez-Buylla *et al.*, 2017).

La ubicación geográfica y las oscilaciones ambientales pueden ser en gran parte responsables de la configuración de los patrones espaciales de diversidad y divergencia genética entre las especies; el nivel y la distribución de la diversidad de nucleótidos son productos históricos de la evolución a largo plazo de una especie, estos se asocian en gran medida con el potencial evolutivo; además, las causas ecológicas o proximales como el tipo de reproducción y diversas barreras como el aislamiento geográfico y espacio-temporal pueden causar la fragmentación en la distribución de las especies, lo que puede conducir a una reducción en el flujo de genes entre poblaciones aisladas y en la adaptabilidad; este proceso inicia la divergencia alopátrica (especiación causada por

barreras geográficas), y la adaptación local puede, en última instancia, conducir a las poblaciones hacia la especiación (Jia *et al.*, 2018).

La estructura de la comunidad vegetal en un ecosistema está controlada por condiciones ambientales como la topografía, el clima y el suelo; la topografía (elevación, pendiente y exposición) da forma al microclima a escala local, entre los parámetros topográficos, la distribución espacial de la vegetación se ve afectada principalmente por la exposición a la pendiente, en general, las pendientes orientadas al sur tienden a ser más secas y cálidas en comparación con las del norte, ya que reciben mayor radiación solar, lo que resulta en una mayor evapotranspiración y mayores temperaturas máximas diarias, principalmente durante el verano. Esta diferenciación afecta a las comunidades de plantas, especialmente en las áreas desérticas donde la disponibilidad de agua es un factor determinante para el crecimiento de las plantas (Abraham *et al.*, 2018).

1.5. Diversidad genética del guayule

La mejora de guayule para la producción a escala industrial podría acelerarse mediante la reproducción asistida por genómica, pero dicha estrategia está limitada por la escasez de herramientas genómicas disponibles y recursos genéticos bien caracterizados; se ha mejorado y cultivado de forma intermitente como fuente de caucho natural durante el siglo pasado (Ilut *et al.*, 2015), sin embargo, el desarrollo de cultivares con alto rendimiento de caucho ha tenido un éxito parcial, en gran parte debido a una combinación de diversidad genética limitada en la fuente de germoplasma disponible (Thompson y Ray, 1989), ploidía extensa, variación dentro y entre las accesiones de guayule disponibles y un modo complejo de reproducción que se caracteriza por la autoincompatibilidad esporofítica en plantas diploides y apomixis facultativa en plantas poliploides (Cruz *et al.*, 2017).

Ilut *et al.*, (2015) muestran que la variación de la ploidía en diferentes accesiones de guayule está estrechamente relacionada con la divergencia genética entre los niveles de ploidía; esto sugiere que una mayor diferencia en el germoplasma en dichas accesiones es necesario para mantener cultivares genéticamente estables, en contraste con la heterogeneidad genética observada en algunas accesiones cultivadas, donde, distintas accesiones muestran una gran similitud genética; lo que indica que es necesaria una

caracterización genómica adicional del grupo de germoplasma de *Parthenium* silvestre y cultivado existente para construir un panel de diversidad genéticamente representativo.

Ilut *et al.*, (2017), describieron el nivel de ploidia de diversos ejemplares de guayule dentro de la colección de germoplasma de la Unidad Nacional de Recursos de Genética de Plantas de Tierras Áridas (NALPGRU) del USDA-ARS, en donde encontraron que las variaciones genéticas se deben principalmente a las relaciones interespecíficas que presenta el guayule con diferentes especies de *Parthenium* como lo son *Parthenium hysterophorus* y *Parthenium tomentosum*, pero sobre todo con *Parthenium incanum*; la determinación de la diversidad genética descrita se basa en la correlación del nivel de ploidía y lugar o cruza de procedencia de la accesión analizada.

1.6. Marcadores moleculares para la diversidad genética

La utilización de marcadores moleculares para la evaluación de la variación genética en plantas es una herramienta eficaz. Se han desarrollado diversos tipos de marcadores moleculares (Tabla 1), tanto a nivel proteico (isoenzimas) como a nivel del ADN, los cuales han sido efectivos para encontrar la variación genética entre accesiones, caracterizar germoplasmas y en el apoyo al mejoramiento genético en un gran número de plantas de interés medicinal, endémico o de explotación industrial (Carvalho *et al.*, 2019).

Los diferentes marcadores basados en ADN se pueden clasificar en varios grupos, tales como marcadores moleculares basados en hibridación (RFLP), marcadores basados en la reacción en cadena de la polimerasa (PCR) de ADN polimórfico amplificado al azar (RAPD); polimorfismo de longitud del fragmento amplificado (AFLP); repeticiones de secuencia simple (SSR) y de secuencia inter simple (ISSR) y marcadores de secuencias basadas en polimorfismo de un solo nucleótido (SNP). Su objetivo es resaltar la variabilidad del genoma, aunque cada una de ellas combina algunas propiedades deseables y se basa en principios diferentes (Monfared *et al.*, 2018).

Las principales limitaciones de estos métodos son la baja reproducibilidad de RAPD, el alto costo del análisis de AFLP y la necesidad de conocer las secuencias flanqueantes para desarrollar cebadores específicos de especies para el polimorfismo SSR (Ramakrishnan *et al.*, 2016).

Existen otros marcadores para estudiar la diversidad genética y para caracterizar molecularmente a una población como los ITS (espaciador interno transcrito, por sus siglas en inglés) del ADN ribosómico nuclear (ADNr), el cual presenta diversas ventajas como una amplificación fácil mediante cebadores universales, longitud más corta y mayor poder de discriminación a nivel de la especie, por lo que hacen que la secuencia ITS de ADNr sea un gen útil para generar inferencias filogenéticas a nivel de especie en la mayoría de los grupos de plantas (Maina *et al.*, 2019, Sivakumar *et al.*, 2018, Wu *et al.*, 2013, Qian *et al.*, 2009).

Los ITS se refieren al ADN espaciador situado entre la región de la subunidad pequeña y subunidades grandes del ARNr en el cromosoma o en la región transcrita correspondiente en el transcrito del precursor de ARNr; estas regiones son altamente variables y por lo tanto, se usa con frecuencia en estudios de diversidad genética (Hao *et al.*, 2012).

Tabla 1. Marcadores moleculares utilizados en genética de poblaciones (Tomado de Eguiarte *et al.*, 2013).

Marcador	Características	Ventajas	Desventajas
Variación fenotípica	Variantes observables y cuantificables	Se puede observar en las poblaciones	No representa la variación de todo el genoma
		En muchas ocasiones son adaptativos	Pueden reflejar plasticidad
		Comparable con fósiles	Muchas veces son poligénicos
aloenzimas	Variación en la migración de las enzimas durante la electroforesis, por cambio de aminoácidos	Muchos loci son polimórficos	Poca variación
		Codominantes (identifica homocigotos y heterocigotos)	No detecta mutaciones silenciosas
			Pocos marcadores a lo largo del genoma Difíciles de montar
RFLPs	Variación en los fragmentos generados por enzima por restricción, debido a cambios mutacionales en los sitios de reconocimiento		Información limitada Complicados de montar Dominantes
RAPDs	Variación en la amplificación de fragmentos anónimos, que depende de la presencia diferencial de sitios de unión (mutaciones y cambios estructurales)	Muchos sitios polimórficos a lo largo del genoma	Dominantes
		Alta variación	Cada sitio poco informativo
		Oligonucleótidos universales	Poco reproducible
			Se desconocen los sitios amplificados Muchos artefactos Difíciles de leer
ISSRs	Similar a los RAPDs, pero los oligonucleótidos son secuencias más complejas y largas en tándem (microsatélite)	Muchos sitios polimórficos a lo largo del genoma	Dominantes
		Alta variación	Cada sitio poco informativo
		Más reproducibles que los RAPDs	Se desconocen los sitios amplificados

En los genomas de plantas, similares a otros eucariotas, los genes de ARNr se producen en miles de copias y se organizan en matrices; cada grupo ribosomal eucariota contiene la secuencia externa transcrita 5' (5' ETS), el gen 18S ARNr, el ITS1, el gen 5.8S ARNr, el ITS2, el gen 28S ARNr y finalmente el 3' ETS, durante la maduración del ARNr, se discriminan las regiones ETS como subproductos no funcionales de este proceso, en el cual se degradan rápidamente (Figura 5). Los espaciadores ITS1-ITS2 evolucionan bajo una presión selectiva más baja que los ETS, por lo que se han convertido en uno de los marcadores más populares en estudios filogenéticos, sin embargo, el gen 5.8S muestra una alta presión selectiva y puede usarse para establecer relaciones filogenéticas a nivel inter-específico (QIAN *et al.*, 2009).

El ADNr es útil para estudios filogenéticos a nivel intra-específico ya que muestra varias ventajas sobre otras regiones del genoma; rápida evolución, detección, amplificación y secuenciación fácil, por lo tanto, el análisis del ADNr proporciona un enfoque útil y adecuado para generar marcadores moleculares eficientes para evaluar la organización de la diversidad genética en plantas superiores (Wu *et al.*, 2013, Qian *et al.*, 2009, Maina *et al.*, 2019).

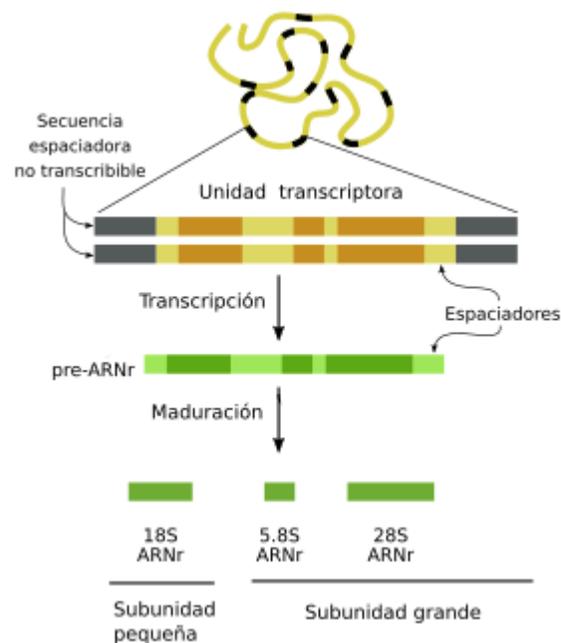


Figura 5. Maduración de ARNr, eliminación de regiones ETS no funcionales (tomada de internet).

Estudios recientes describen a la región del ITS1-ITS2 como un marcador genético y una inminente utilización de esta región del ADNr como un código de barras universal ya que el ITS2 es un posible marcador filogenético general usado ampliamente para reconstrucciones filogenéticas tanto a nivel de género como de especie ya que presenta una alta divergencia inter-específica y se encuentra bien delimitado, tiene un tamaño de pocos cientos de nucleótidos, la comparación de las relaciones de las subespecies con los niveles de orden, la doble verificación de posibles errores de secuencia en alineaciones dirigidas por estructuras secundarias, basado en esto, se propone que la región ITS1-ITS2 sea un “*código de barras de oro*” universal para identificar plantas y hongos (Chen *et al.*, 2010, Lee *et al.*, 2017, Kaushik *et al.*, 2017).

Las regiones ITS1-ITS2 se han utilizado ampliamente como un marcador molecular para discriminar entre diferentes especies en estudios filogenéticos postulándose como uno de los mejores marcadores para la función de código de barras (Kumar *et al.*, 2009), el uso de ADN nuclear tiene ventajas sobre el ADN de plástidos ya que tienen una alta tasa de evolución, la herencia es biparental y contiene distintos genes con longitud adecuada e independientes (Kumari *et al.*, 2018), dentro del ADN nuclear se tienen las secuencias codificantes para el ARN ribosomal el cual, representa aproximadamente el 40 % del total de transcripción dentro de una célula, y, a su vez, este constituye el 80% del ARN celular (Amar *et al.*, 2012). Este marcador molecular se ha utilizado para determinar relaciones intra e inter-específicas en diferentes organismos como hongos (Sims *et al.*, 2019, Scanu *et al.*, 2015, Jagadeesh *et al.*, 2018), animales (De-la-Mora *et al.*, 2018) y distintas plantas (Maina *et al.*, 2019, Kumari *et al.*, 2018, Lee *et al.*, 2017, Kehie *et al.*, 2016, Ghada *et al.*, 2013, Wu *et al.*, 2013, Hao *et al.*, 2012, Kumar *et al.*, 2009).

I.7 El enfoque filogenético en estudios de diversidad y estructura genética

La filogenética es una rama de la biología evolutiva que se ocupa de inferir las relaciones evolutivas entre diferentes grupos de organismos a partir de caracteres homólogos moleculares o morfológicos. La filogenética molecular utiliza matrices de información de moléculas de ADN, generalmente, datos de secuencias. Por lo tanto, la filogenética se basa en el desarrollado de árboles filogenéticos, permitiendo la representación de eventos evolutivos (Wicke y Fischer, 2018).

La evaluación de diversidad genética y la vinculación de los patrones ecológicos espaciales sumados a la información filogenética hacen aún más poderosos los resultados obtenidos, ya que, un conjunto de especies que abarcan nodos filogenéticos más profundos presentan más historia evolutiva y son más diversos que otras áreas con el mismo número de especies conectadas a través de nodos menos profundos, los enfoques filogenéticos informan el patrón y el proceso evolutivo (Allen *et al.*, 2019).

La variación en los contenidos genómicos se puede atribuir a varios eventos moleculares, como la ganancia y/o pérdida de genes, mutaciones sinónimas y puntuales, inserciones, deleciones y deriva genética. La detección de estos eventos es esencial para encontrar relaciones entre diferentes secuencias, genomas y eventualmente organismos. Los árboles filogenéticos podrían establecerse utilizando todos estos eventos para resaltar las relaciones entre las entidades. Estos árboles se generan a partir de los datos moleculares y podrían proporcionar información confiable para inferir la evolución y las relaciones de las especies y familias de genes (Godini y Fallahi, 2019).

I.8 Análisis bioinformáticos.

I.8.1 Alineamiento de secuencias.

El alineamiento múltiple de secuencias (MSA, por sus siglas en inglés) de ADN es un paso crítico para los análisis bioinformáticos ya que desde el alineamiento es que se inician los análisis clásicos de secuencias desarrollados hasta el momento; el alineamiento es el punto de partida de cualquier campo de investigación molecular biológico donde este actúa como una ventana para ver la perspectiva evolutiva, funcional y estructural de las macromoléculas biológicas en un formato conciso (Chowdhury y Garai, 2017).

Alinear es una forma de comparar dos o más secuencias de ADN, para esquematizar las regiones similares y regiones divergentes y establecer la homología de secuencias alineadas, las cuales podrían indicar relaciones funcionales o evolutivas entre los genes. Las secuencias alineadas se escriben con las letras correspondientes a cada nucleótido de acuerdo con el código IUPAC (A; adenina, T; timina, C; citocina, G; guanina); de manera específica, se insertan espacios (gaps) para que las regiones idénticas o similares se alineen. Las representaciones visuales de los alineamientos ilustran la

inserción de “gaps” o huecos que pueden ser interpretados como inserciones o deleciones de nucleótidos llamadas mutaciones “InDel” (Leduc y Zhao, 2019).

Estos alineamientos se utilizan para la generación de árboles filogenéticos, y otros análisis bioinformáticos, y desarrollar modelos de homología sobre estructuras de nucleótidos. Sin embargo, la relevancia biológica de los alineamientos no siempre es clara ya que existe diferentes algoritmos y enfoques para alinear un mismo bloque de secuencias, estas generalmente dan como resultado alineamientos distintos, por lo que es imprescindible realizar estudios más robustos como la filogenia o predicción de estructura de material genético para poder realizar inferencias en la evolución de una población (Chowdhury y Garai, 2017).

En la actualidad se dispone de diferentes algoritmos que permiten realizar el MSA, el enfoque de alineación progresiva reúne todas las secuencias de forma progresiva donde se toma primero en cuenta la mejor alineación por pares del total de secuencias a analizar, este enfoque es utilizado por distintos programas como Multal, Map, Pcma, Multaling, Clustal, T-Coffee, KAlign, entre otros. Un enfoque distinto para realizar alineamiento múltiple de secuencias es el iterativo, este tipo de alineamientos realiza un procesamiento posterior al hacer cambios en la alineación realizada por métodos progresivos, en este enfoque se calculan nuevas matrices de distancia utilizadas en el alineamiento obtenido por alineación progresiva, esto conduce a una segunda ronda de alineación progresiva, los programas que utilizan esta reestimación son Mafft, Muscle, Prime, Prrp y Mummals/Promals (Chowdhury y Garai, 2017). Para datos con una longitud de pb pequeña/media se realizan alineaciones progresivas las cuales son adecuadas y dan un alineamiento funcional para los análisis posteriores, sin embargo siempre se debe tener en cuenta que el elegir un método de alineación sobre todo, puede afectar enormemente la calidad de los resultados y, en consecuencia, la validez de los árboles filogenéticos resultantes (Godini y Fallahi, 2019).

I.8.2 Inferencia filogenética.

La gran cantidad y variedad de información genética generada en los últimos años ha llevado a la necesidad de procesar e interpretar dicha información, en este sentido; la comprensión de dicha información, el modelado de procesos biológicos y los estudios filogenéticos son de gran ayuda (Delgado y Gómez-Vela, 2019).

Después del MSA, se usa esta información para construir un árbol evolutivo, en general, hay dos enfoques principales para reconstruir un árbol; métodos basados en distancia (métodos de matriz de distancia) y basados en caracteres (métodos de búsqueda de árbol). El método de unión de vecinos (NJ, por sus siglas en inglés) es el algoritmo principal en uso para el método basado en la distancia, mientras que, los métodos de máxima parsimonia (MP), máxima verosimilitud (ML, por sus siglas en inglés) e inferencia bayesiana (IB) se clasifican como métodos basados en caracteres (Godini y Fallahi, 2019).

Dentro de los métodos basados en caracteres, los métodos ML e IB son métodos probabilísticos basados en modelos de evolución molecular, donde se toma en cuenta conocimiento *a priori* acerca de los caracteres, especialmente cuando son caracteres moleculares, frecuentemente secuencias de nucleótidos de ADN (Peña, 2011).

Los modelos evolutivos describen la probabilidad relativa de cambio de un nucleótido a otro; estos modelos nos informan sobre todos los eventos de mutación que han ocurrido a través del tiempo; los modelos de sustitución de nucleótidos son muy importantes en los métodos probabilísticos de inferencia filogenética (ML e IB) (Godini y Fallahi, 2019).

Los estudios filogenéticos se realizan con la finalidad de obtener un árbol filogenético, el cual, muestra la historia evolutiva del grupo taxonómico en estudio, es decir, un árbol filogenético que sea reflejo del proceso de evolución, donde las entidades biológicas son el resultado de “descendencia con modificación” de especies ancestrales (Peña, 2011).

Habiendo numerosas modificaciones en un organismo a lo largo del tiempo, estos cambios nunca se consideran como evolución hasta que se heredan a través de material genético y en generaciones consecutivas. En filogenética, un grupo que incluye una

especie ancestral, donde apareció una modificación conservada a través de las siguientes generaciones se denomina clado o grupo monofilético (Godini y Fallahi, 2019).

Para evaluar la construcción del árbol filogenético se deben tomar en cuenta cuatro criterios; consistencia, eficiencia, robustez y velocidad computacional. La consistencia hace referencia a la estimación de un método, esta debe ser de la misma manera para datos infinitos (los métodos como la distancia y la ML son consistentes, mientras que la MP no lo es); la eficiencia se considera de acuerdo con la probabilidad de obtener el árbol correcto (la ML e IB son métodos con alta eficiencia); la robustez significa que el método responda correctamente a suposiciones críticas; finalmente, la velocidad computacional depende del algoritmo de los métodos, la capacidad de procesamiento del equipo de cómputo y el tamaño de los datos.

1.8.2.1 Inferencia filogenética basada en máxima verosimilitud (ML).

Este método es uno de los métodos más comunes para inferencias filogenéticas, el principio de este método es evaluar la probabilidad de un árbol hipotético para predecir las relaciones de las entidades de interés, es decir, el objetivo de este método es obtener el árbol que tiene la probabilidad más alta de producir las secuencias observadas dado un modelo evolutivo.

Para iniciar un análisis de ML, es necesario realizar búsquedas heurísticas del árbol con la mayor verosimilitud partiendo de un árbol inicial generado por el método de máxima parsimonia (método empleado por el programa RAxML); la validez de los clados inferidos se evalúa mediante un análisis de bootstrap.

1.8.2.2 Inferencia filogenética basada en métodos bayesianos (IB).

Este método usa todos los sitios contenidos en las matrices de secuencias alineadas de los MSA, basado en las similitudes y diferencias entre las secuencias, se construyen varios árboles de prueba con una topología distinta, este método incluye variables aleatorias con distribuciones de probabilidad definidas *a priori*, aquí, el algoritmo de cadena de Markov Monte Carlo (MCMC) muestrea valores de parámetros en el espacio multidimensional, los cuales son la base para el cálculo de las probabilidades posteriores de cada clado. La probabilidad posterior del árbol está definida como la probabilidad del árbol o hipótesis filogenética condicionada a los datos. (Godini y Fallahi, 2019).

La interpretación fácil de resultados estadísticamente consistentes y eficientes son ventajas de este método. Algunas de las desventajas de la IB son su demanda computacional y tendencia para sobreestimar los valores de soporte (probabilidades posteriores) de los nodos.

Dado que el cálculo de la probabilidad posterior del árbol implica integrar todas las combinaciones posibles de valores de parámetros (longitudes de ramas, topología del árbol y parámetros del modelo evolutivo) y dichas combinaciones representan un número infinito de hipótesis imposible de integrar analíticamente, la IB emplea Cadenas de Markov Monte Carlo (MCMC, por sus siglas en inglés). Dichas cadenas son algoritmos estocásticos que muestrean aleatoriamente las distribuciones de probabilidad de los parámetros para aproximar la probabilidad posterior del árbol filogenético.

I.8.3 Modelos de evolución de ADN

Comprender los procesos que producen variaciones genéticas a lo largo de la evolución de las poblaciones naturales es fundamental en la evolución molecular y la biología estructural; los modelos de evolución se utilizan ampliamente para representar la aproximación más probable de la ruta evolutiva de los genes y las proteínas (Zamudio *et al.*, 2019).

Para poder determinar las relaciones filogenéticas moleculares es necesario construir un modelo matemático que describa la evolución en el tiempo de las secuencias, este puede ser construido empíricamente, usando propiedades calculadas a través de comparaciones de secuencias observadas, o paramétricamente, usando propiedades químicas o biológicas de ADN y aminoácidos. Estos modelos permiten estimar la distancia genética real o esperada entre dos secuencias homólogas, medido por el número esperado de sustituciones de nucleótidos por sitio que se han producido en los linajes evolutivos entre ellos y su ancestro común más reciente; estas distancias pueden representarse como longitudes de rama en un árbol filogenético, las secuencias existentes forman las puntas del árbol, mientras que las secuencias ancestrales forman los nodos internos y generalmente no se conocen (Liò y Goldman, 1998, Sullivan y Joyce, 2005).

Mientras que los métodos de MP asumen un modelo implícito de evolución (número mínimo de sustituciones a lo largo de la filogenia), los métodos basados en distancias (p. ej. NJ), y los probabilísticos (ML e IB) requieren de un modelo explícito de evolución; los métodos de distancia reducen la información contenida en la matriz de secuencias alineadas a una matriz de distancias genéticas empleando un modelo evolutivo; en cambio, los métodos de ML y Bayesianos pueden estimar el valor de cada uno de los parámetros del modelo, dada una topología y la matriz de datos.

La implementación de los modelos evolutivos comenzó en 1965, cuando se propuso la teoría del reloj molecular, esta asevera que la tasa de evolución molecular es aproximadamente constante en el tiempo para todas las proteínas en todos los linajes (Zuckerandl y Pauling, 1965). De acuerdo con esta teoría, cualquier divergencia entre genes, proteínas o linajes se puede fechar simplemente midiendo el número de cambios entre secuencias; sin embargo, la hipótesis del reloj molecular resulta ser demasiado simple. Actualmente se sabe que las tasas de mutación varían entre genomas y dentro de ellos, y se ven afectadas por diversos factores, como la posición cromosómica, el contenido de G + C, las bases vecinas más cercanas y diferente eficiencia de los sistemas de reparación entre las cadenas de ADN durante la replicación y transcripción (Liò y Goldman, 1998).

En la actualidad existe una gran variedad de modelos evolutivos con diferentes niveles de complejidad, el más simple es el propuesto por Jukes y Cantor (1969), el cual asume que todas las sustituciones de nucleótidos ocurren a la misma tasa de cambio, y todos los nucleótidos tienen la misma probabilidad de ser reemplazados por cualquier otro.

Siendo este el primer modelo de sustitución nucleotídica, se encontraron parámetros no considerados por lo que se han desarrollado distintos modelos y cada uno de ellos presentan diferencia en el número de variables a considerar.

De acuerdo con lo antes descrito, en este trabajo se estudió la diversidad genética de plantas silvestres de guayule analizando las regiones ITS1-ITS2 del ADNr, realizando estudios filogenéticos y correlacionando la diversidad genética encontrada con las condiciones geográficas en las que se obtuvo el material vegetal. Esto, en un futuro, encaminado a una explotación agroindustrial integral-sustentable en donde se tomen en

cuenta las diferencias genéticas encontradas y poder implementar un plan de mejoramiento genético e identificar áreas importantes a conservar, la generación de conocimiento sobre la genética de esta planta, evitando el desgaste genético en las posibles plantaciones comerciales, así como aporta datos y responde a interrogantes surgidas en investigaciones previas (Ilut *et al.*, 2015).

II. HIPÓTESIS

El uso de marcadores ITS en plantas silvestres de guayule nativas de sitios geográficos diferentes nos dará información sobre su diversidad genética dentro y entre poblaciones colectadas, permitiendo identificar áreas importantes a conservar, así como información sobre su posible evolución y estructura de la población.

III. OBJETIVOS

III.1 Objetivo general

Estudiar la diversidad genética de plantas silvestres de guayule distribuidas en cinco estados del Norte de México, mediante secuenciación de fragmentos ITS1-ITS2 para encontrar áreas importantes a conservar.

III.2 Objetivos específicos

III.2.1 Identificar poblaciones de guayule silvestre distribuidas en los cinco estados objeto de estudio para colecta de material vegetal.

III.2.2 Diseñar cebadores específicos para la familia *Asteraceae* que amplifiquen regiones ITS1-ITS2 para evaluar su efectividad como marcadores genéticos en guayule.

III.2.3 Evaluar la diversidad genética dentro y entre las poblaciones de guayule mediante el uso de programas bioinformáticos específicos.

III.2.4 Describir la estructura genética dentro de la especie del guayule mediante correlaciones genéticas y geográficas.

IV. MATERIALES Y MÉTODOS

IV.1 Colecta de material vegetal

Se ubicaron geográficamente los sitios de colecta con base en: (a) los sitios reportados en la literatura, (b) la experiencia previa del grupo de Biociencias y Agrotecnología del CIQA y (c) comunicaciones personales con los habitantes y lugareños de cada una de las zonas de colecta. Cada población se delimitó por una distancia mínima de 400 m entre ellas, esto, con la finalidad de maximizar la representatividad de la diversidad genética de las plantas colectadas (Emami *et al.*, 2018).

Se colectaron *in situ* hojas jóvenes y sanas, cada población estuvo conformada de 3 a 10 plantas silvestres, esto en relación con la extensión territorial (tamaño de rodal) de la misma, existiendo una distancia mínima de 12 m entre plantas colectadas de acuerdo con lo reportado por Emami *et al.*, (2018). De cada muestra vegetal colectada se documentaron características fenotípicas particulares tales como: forma, tamaño y color de la hoja, presencia y frecuencia de dentado, altura, ancho y diámetro de tallo de cada planta (Figura 6), ya que estos parámetros confieren uniformidad al momento de evaluar rendimiento de germoplasma (Coffelt y Johnson 2011). Las hojas colectadas se congelaron con hielo seco (CO₂ en estado sólido) al momento del corte y transportadas sin perder la congelación al laboratorio de Agrobiotecnología en las instalaciones del CIQA para posteriormente ser almacenadas a -80 °C hasta el momento de su procesamiento.

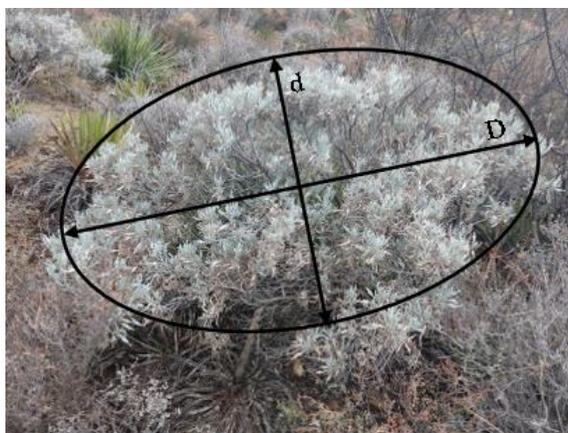


Figura 6. Ilustración de diámetros medidos en campo de plantas silvestres de *P. argentatum*, siendo “D” el eje mayor y “d” el eje semi mayor.

IV.2 Extracción de ADN

Se realizó la extracción de ADN total de las muestras colectadas, empleando el kit “DNeasy Plant Mini kit” (Qiagen, Países Bajos) siguiendo las instrucciones del fabricante; de forma resumida, el protocolo consiste en los siguientes pasos: se realizó ruptura mecánica con nitrógeno líquido en un tubo para centrifuga de 1.5 ml, seguida de una lisis celular. Con la finalidad de eliminar el ARN presente en la muestra se utilizó ARNasa; los precipitados se eliminaron en un solo paso mediante centrifugación utilizando la columna QIAshredder® (provista en el kit), que es una unidad de filtración y homogeneización.

La solución amortiguadora de unión y el etanol se agregaron al lisado limpio para promover la unión del ADN a la membrana de una segunda columna DNeasy® (provista por el kit). El ADN se unió a la membrana, mientras que los contaminantes, como las proteínas y los polisacáridos, se eliminaron de manera eficiente mediante dos pasos de lavado con etanol al 100 %. Finalmente, el ADN se diluyó en un volumen de 30 μ L de agua libre de nucleasas; posterior a la extracción se confirmó calidad e integridad del ADN mediante electroforesis en agarosa al 1 % aplicando 90 volts de corriente eléctrica por 1 hora, se realizó la tinción de cada gel con bromuro de etidio (100 μ L/L) por 15 minutos y se capturó imagen mediante el fotodocumentador Gel Doc™ EZ Documentation System (Biorad, EE. UU.).

IV.3 Diseño de cebadores

La región ITS1-ITS2 del ADN ribosómico nuclear, se amplificó mediante PCR usando cebadores específicos. El diseño de cebadores específicos se realizó mediante el alineamiento de 64 secuencias obtenidas de GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), la selección de las secuencias utilizadas se basó en la presencia completa de la región ITS1-ITS2. Todas las secuencias pertenecen a plantas terrestres de la familia *Asteraceae*, esto, con el fin de que los cebadores diseñados fueran específicos para esta familia, las secuencias se alinearon empleando el algoritmo de alineamiento ClustalX versión 2.1 (Allen *et al.*, 2019), implementado en el programa BioEdit versión 7.0.5 (Hall 1999).

Los cebadores se diseñaron en las regiones altamente conservadas 18S y 26S (Figura 7), el cebador de inicio se diseñó en la región codificante para la subunidad 18S y el cebador final se diseñó en la región codificante para la subunidad 26S (identificadas con una flecha).

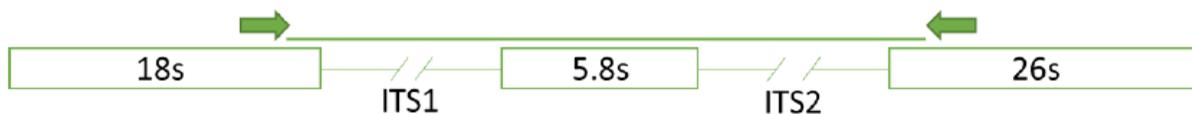


Figura 7. Región de interés en el diseño de cebadores específicos para Asteraceae en ADNr.

Los cebadores diseñados en este proyecto (AGB-ITS-F, AGB-ITS-R), se compararon con cebadores universales ITS-U1-F e ITS-U4-R anteriormente reportados para la región ITS1-ITS2, siguiendo las condiciones de amplificación reportadas (Cheng *et al.*, 2016).

IV.4 Amplificación de las regiones ITS1-ITS2 en ADN ribosómico nuclear

Las reacciones en cadena de la polimerasa (PCR), se realizaron en un volumen total de 25 μ L conteniendo 2.5 μ L de solución amortiguadora High Fidelity 10x, 1 μ L de $MgSO_4$ 50 mM, 0.5 μ L de mix dNTP's 10 mM, 0.5 μ L de cada primer 10 mM, 0.1 μ L de Platinum Taq High Fidelity (Thermo Fisher Scientific. EE. UU.), y 1 μ L de ADN total vegetal, esto, de acuerdo con las condiciones de uso de la enzima. La programación del perfil térmico

de PCR fue de la siguiente manera: 1 ciclo a 94 °C por 2 min; seguido de 25 ciclos a 94 °C por 15 s; 55 °C por 30 s; 68 °C por 1 min, y finalmente 1 ciclo adicional a 68 °C por 7 min. Se evaluó la cantidad y calidad de los fragmentos de ADN amplificados mediante electroforesis en geles de agarosa al 1%, donde la banda esperada de aproximadamente 700 pb se corroboró mediante el uso de un marcador de peso molecular (ADN Ladder 100 pb, Promega, EE. UU.). La imagen del gel se capturó con un sistema de fotodocumentación Gel Doc™ EZ Documentatio System (Biorad, EE. UU.), previamente teñido con bromuro de etidio.

IV.5 Purificación y cuantificación de productos de PCR

Se realizó la purificación del fragmento amplificado mediante el kit GenElute™ PCR Clean-Up Kit (Sigma-Aldrich, EE. UU.), siguiendo las especificaciones del fabricante; brevemente se describe el protocolo a continuación. Se preparó la columna GenElute plasmid mini spin™ para filtrar por centrifugación la muestra de PCR a la cual se le había adicionado previamente la solución de unión, la cual promueve la unión del ADN a la columna, se realizó lavados con solución de etanol y finalmente se separó el ADN de la columna con 30 µL de agua estéril libre de nucleasas.

Teniendo las muestras purificadas se realizó la cuantificación de ácidos nucleicos en el laboratorio de Biología Molecular de la Universidad Autónoma de Coahuila campus Saltillo. La cuantificación consistió en tomar 2 µL de la muestra y colocarlos en una microplaca del espectrofotómetro Epoch (BioTek, EE.UU.), para la cuantificación se utilizó agua libre de nucleasas como blanco, y se realizaron las lecturas para cada muestra bajo el principio de espectrofotometría.

IV.6 Secuenciación de fragmentos amplificados

Los fragmentos amplificados fueron enviados a secuenciación como servicio externo en el Laboratorio Nacional de Genómica para la Biodiversidad (Langebio), el cual se encuentra ubicado en Irapuato, Guanajuato, México. Esta secuenciación se realizó en ambos sentidos de la cadena de ADN, con los mismos cebadores utilizados en la amplificación del fragmento. Se empleó la técnica de Sanger y tecnología capilar con los sistemas ABI 3730 XL (Applied Biosystem, California, USA).

IV.7 Análisis bioinformáticos

Se realizaron ensamblajes para las secuencias correspondientes a cada muestra corroborando con los cromatogramas y alineamientos entre las mismas utilizando los programas ClustalX versión 2.1, y BioEdit versión 7.0.5. Se realizó un alineamiento entre las secuencias de ambos sentidos para cada una de las muestras estudiadas, con la finalidad de corroborar la correcta secuenciación, además, en este paso se ubicaron y editaron manualmente los sitios con ambigüedades.

Las secuencias utilizadas para los análisis estadísticos se llamaron de acuerdo al código de ambigüedades nucleotídicas descrito por la IUPAC (Stireman *et al.*, 2018).

Tabla 2. Código de ambigüedades nucleotídicas descrito por la IUPAC.

Código de ambigüedades nucleotídicas IUPAC		
Símbolo	Significado	Ácido nucleico
A	A	Adenina
C	C	Citocina
G	G	Guanina
T	T	Timina
U	U	Uracilo
M	A o C	
R	A o G	
W	A o T	
S	C o G	
Y	C o T	
K	G o T	
V	A o C o G	
H	A o C o T	
D	A o G o T	
B	C o G o T	
X	G o A o T o C	

IV.7.1. Estadísticas descriptivas de las secuencias y estimación de índices de diversidad genética.

Después del alineamiento, las secuencias fueron analizadas con el programa DnaSP versión 6.12.03 (Rozas, 2009) para estimar los índices de polimorfismos y la demografía histórica.

Se realizó una reconstrucción de haplotipos con el software DnaSP por el método Phase, esto, para inferir las fases gaméticas, la cual se utilizó para poder hacer las inferencias de los índices de diversidad genética en datos concatenados presentes en las secuencias analizadas.

Dentro de los datos descriptivos se calculó el porcentaje de sitios informativos, contenido de G+C, sitios variables (S), total de mutaciones (Eta), diversidad nucleotídica por sitio entre dos secuencias (Pi), Diferencias nucleotídicas (k), theta por sitio y por secuencia (θ), número de haplotipos (h), diversidad de haplotipos (Hd), además pruebas de neutralidad clásicas como la D de Tajima, D y F de Fu y Li para inferir si las regiones ITS1 e ITS2 evolucionan de manera aleatoria (neutral), entre las plantas silvestres de guayule en la región ITS1-ITS2 secuenciadas (Losos, 2014).

IV.7.2. Reconstrucción filogenética con enfoque bayesiano

Se realizó la estimación del modelo evolutivo que mejor explicara las secuencias analizadas por regiones y concatenadas, mediante el programa jModeltest versión 2.1.10.(Posada y Crandall 1998); para seleccionar el modelo evolutivo se usaron los siguientes criterios: prueba de razones de verosimilitud, prueba jerárquica de razón de verosimilitud (LRTs, hLRTs, respectivamente) y criterios de información (Criterio de Información de Akaike; AIC y Criterio de Información Bayesiano; BIC); para ello se calcularon las verosimilitudes de cada criterio para los 203 posibles modelos evolutivos (Sullivan y Joyce 2005).

La reconstrucción del árbol filogenético se realizó con el programa MrBayes versión 3.2.7(Huelsenbeck y Ronquist, 2001), empleando el juego de 160 secuencias alineadas incluyendo las regiones 18S, ITS1, 5.8S y ITS2 con un total de 645 pb. Se utilizó el modelo evolutivo que mejor se ajustó a los datos, el archivo de entrada al programa se editó en formato Nexus, este análisis se llevó a cabo mediante dos corridas independientes de

manera simultánea para cada una de las regiones analizadas (18S, ITS1, 5.8S y ITS2). La configuración de cada corrida independiente fue: 8 cadenas de Markov Monte Carlo para 10,000,000 de generaciones, muestreando cada 4,000 generaciones, teniendo un número final de 2,500 árboles muestreados, se utilizó una temperatura de cadena de 0.05 con 1 cadenas en frío y 7 cadenas calientes, Se descartó el 25 % de las muestras resultantes, a esto se le llama “burn-in”. Al finalizar el análisis se obtuvo un árbol consenso del 50%, es decir, incluyó clados o grupos monofiléticos que estuvieron presentes en al menos el 50% de los árboles muestreados.

IV.7.3. Reconstrucción filogenética basada en máxima verosimilitud (ML)

Para el análisis de ML se utilizó el programa RAxML versión 8.2.10 (Stamatakis *et al.*, 2008), el archivo de entrada se editó a formato Phylyp, se realizó análisis para datos concatenados (160 taxa con 645 pb) y cuatro particiones (18S, ITS1, 5.8S y ITS2), utilizando el modelo evolutivo GTR, siendo este el único modelo en el programa. Se realizaron 100 búsquedas de árbol con la máxima verosimilitud partiendo de un árbol basado en máxima parsimonia, se utilizó la configuración estándar (raxmlHPC-SSE3 -s inputfile.phy -noutput -m GTR -#100 -p 12345), el soporte de los clados se obtuvo a partir de 100 réplicas de un análisis de bootstrap estándar.

IV.7.4. Análisis molecular de varianza (AMOVA)

Para evaluar la estructura genética de las poblaciones estudiadas, se realizó un AMOVA empleando el programa Arlequin versión 3.5.2.2 (Excoffier *et al.*, 1992), en donde se especificaron los 27 haplotipos presentes en guayule así como sus frecuencias en las poblaciones, estos datos se obtuvieron en el programa DnaSP (ver apartado V.4.1. Estadísticas **descriptivas de las secuencias e índices de diversidad genética** Cada población (rodal) se tomó como la unidad básica, después, se ordenó por grupos independientes por cada estado, donde cada estado contaba con la información de haplotipos y frecuencias para cada una de sus poblaciones.

IV.7.5. Construcción de red de haplotipos

Para la construcción de la red de haplotipos se utilizó el programa Network versión 5.0.1.1 (Bandelt *et al.*, 1999), y el algoritmo median joining (MJ, por sus siglas en inglés), modificando el peso de sitios hipervariables (511 y 62 peso =5), al cual, se le realizó una

corrección de parsimonia, para generar una red con el menor número de homoplasias (ambigüedades en correlación de haplotipos).

IV.7.6. Estimación de tiempos de divergencia

Para la estimación de tiempos de divergencia se utilizó el paquete de programas BEAST versión 2.4.7 (Drummond y Rambaut, 2007), se realizó una búsqueda en GenBank de secuencias de la región estudiada para especies de los géneros *Carramboa* y *Layia*, las cuales pertenece a la tribu *Millerieae* y *Madieae* respectivamente, la primer tribu es genéticamente cercana a la tribu *Heliantheae*, a la cual pertenece el guayule, la segunda se utiliza como grupo externo en el análisis, se utilizaron las secuencias AF465848.1 y AF465849.1 para *Carramboa* y DQ188082.1, DQ188079.1, DQ188077.1 y DQ188075.1 para *Layia*. Después de adicionar las secuencias externas se alinearon con el algoritmo MUSCLE y se preparó el archivo en el programa BEAUTi para configurar el análisis; se utilizó el modelo de reloj estricto para la estimación de los tiempos de divergencia, se adicionaron dos nodos de calibración en el análisis, el primero corresponde a la tribu *Heliantheae*, que incluye todas las muestras analizadas; se agregó información bajo un modelo coalescente, la edad promedio para la tribu de 12.88 millones de años con un intervalo del 95% de densidad posterior más alta (HPD) de 6.30-19.50 con una desviación estándar de 4. El segundo nodo incluyo todas las secuencias pertenecientes a la tribu *Millerieae* donde se utilizó un modelo log normal, para el parámetro offset se utilizó un valor de 3 millones de años, una desviación estándar de 0.4 y un promedio de 0.3; tanto el nodo 1 como el dos se tratan de calibraciones secundarias basadas en los resultados de Panero y Crozier, (2016). Se realizó el análisis empleando 10,000,000 de generaciones del algoritmo MCMC, muestreando cada 2,000 generaciones, para obtener un total de 5,000 muestras de árboles y parámetros. El árbol resultante se visualizó mediante el programa FigTree (Rambaut, 2012), el programa Tracer (Rambaut, 2012) se utilizó para ver gráficamente los valores de los parámetros con sus estadísticas descriptivas y evaluar la convergencia y estacionariedad de las cadenas de Markov.

V. RESULTADOS Y DISCUSIÓN

V.1 Colecta de material vegetal

Se realizó colecta de material vegetal de *P. argentatum* en 40 poblaciones distribuidas en los estados de Durango, San Luis Potosí, Zacatecas, Nuevo León y Coahuila. Las poblaciones se codificaron de la siguiente manera: D1 a D13 para las poblaciones ubicadas en Durango, SL1 a SL12 para poblaciones de San Luis Potosí, Z1 a Z5 para poblaciones de Zacatecas, NL1 a NL5 para poblaciones de Nuevo León y C1 a C5 para poblaciones ubicadas en el estado de Coahuila (Tabla 3). Se obtuvieron en total 176 muestras de tejido foliar, una muestra por planta, para extracción de ADN y amplificación de las regiones genómicas nucleares ITS1-ITS2.

Tabla 3. Número de poblaciones e individuos de *P. argentatum* muestreados por estado.

Estado	No. de poblaciones muestreadas	No. de individuos muestreados
Durango	13	59
San Luis Potosí	12	56
Zacatecas	5	20
Nuevo León	5	20
Coahuila	5	21
Total de muestras colectadas		176

Adicional a la colecta de guayule se colectaron tres muestras de especies de *Parthenium* cercanas a guayule (*P. incanum* y *P. Rollinsianum*) en los estados de Zacatecas y San Luis Potosí respectivamente, siguiendo la metodología de colecta de guayule, se ubicó una población para cada especie.

La colecta de muestras para los estados de Durango y San Luis Potosí se realizó en el mes de septiembre del 2018; las temperaturas promedio máximas fueron de 28 y 33 °C, respectivamente (Figura 8). Para los estados de Zacatecas, Nuevo León y Coahuila, la colecta se realizó en el mes de enero del 2019; las temperaturas máximas promedio fueron 23, 22 y 21 °C, respectivamente (Figura 9). Todas las temperaturas promedio se obtuvieron de la base de datos de CONAGUA (<https://smn.cna.gob.mx/es/climatologia/temperaturas-y-lluvias/resumenes-mensuales-de-temperaturas-y-lluvias>).

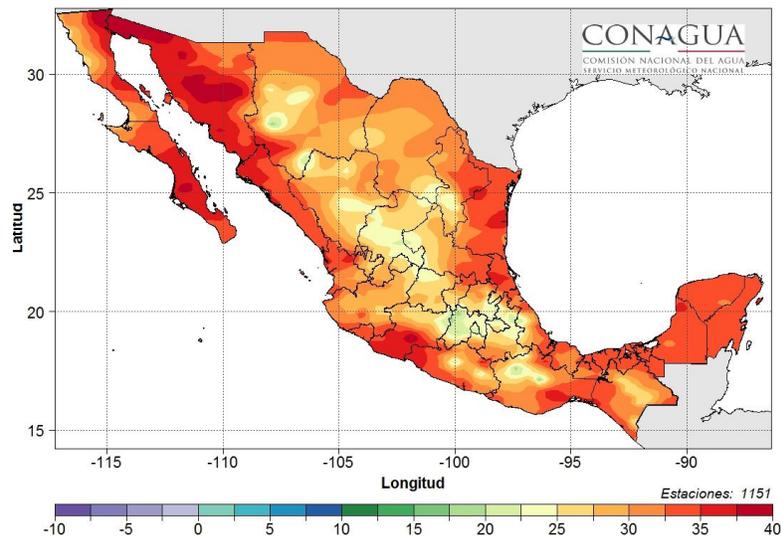


Figura 8. Temperaturas máximas promedio para septiembre 2018 (CONAGUA).

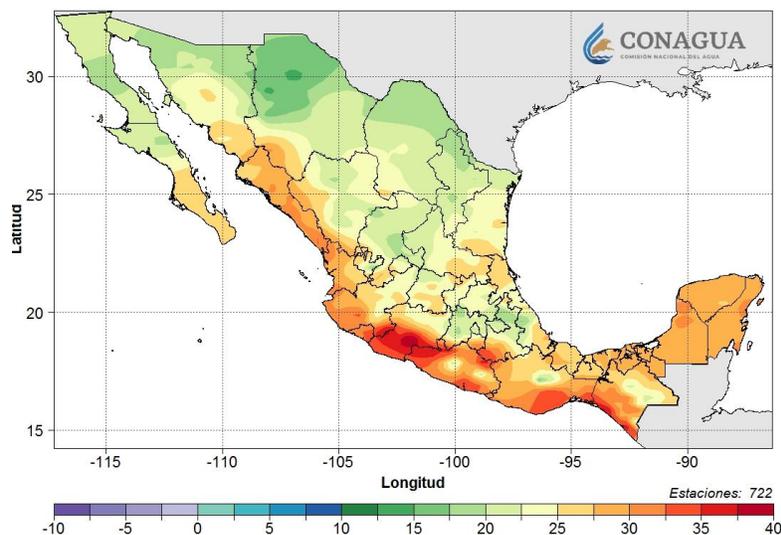


Figura 9. Temperaturas máximas promedio para enero 2019 (CONAGUA).

V.2 Extracción de ADN

Se obtuvo ADN genómico total de buena calidad de las 176 muestras procesadas (Figura 10). Debido a las condiciones particulares de cada muestra de tejido foliar, en algunos casos se observó un poco de variación en la cantidad y calidad del ADN extraído.

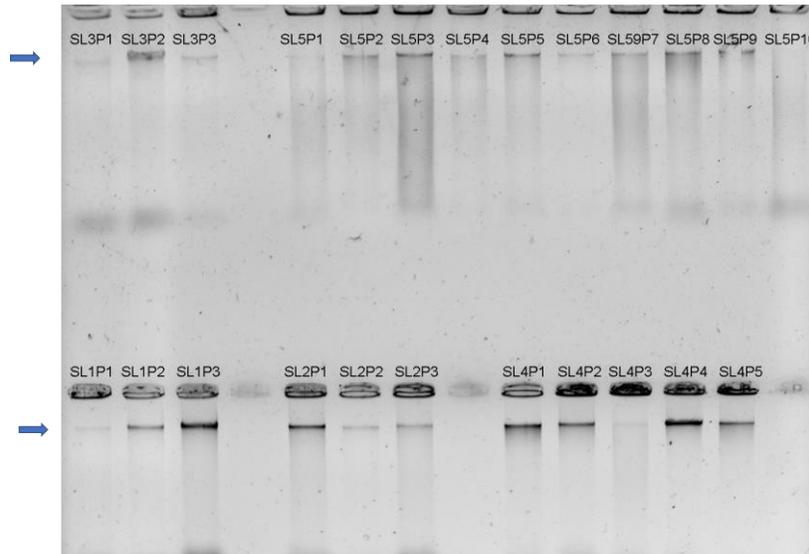


Figura 10. Extracción de ADN genómico total de hojas de *P. argentatum*. ADN extraído de muestras de tejido foliar de plantas con distintas condiciones de campo en un gel de agarosa al 1 %.

V.3 Diseño de cebadores y amplificación de las regiones ITS1-ITS2 en plantas silvestres de guayule.

En este trabajo se estudiaron 160 de las 176 plantas silvestres de guayule, ya que, por cuestiones de logística no se pudo realizar el envío del total de muestras para su secuenciación. Se determinó si el marcador ITS es capaz de discriminar a nivel de especie, posteriormente, se estudió la diversidad genética y se describió la estructura de las poblaciones estudiadas. Hasta el momento no se encontraron reportes de estudios de este tipo para plantas silvestres de guayule, siendo este trabajo, el primer estudio donde se utiliza la región ITS1-ITS2 para la inferencia de dicha diversidad.

Cheng *et al.*, (2016), reportaron cebadores con universalidad y especificidad mejoradas para la amplificación de la región ITS1-ITS2, para ser usados en estudios filogenéticos y como código de barras del reino *Plantae*. Dichos cebadores universales, fueron

diseñados con secuencias de hongos y plantas de diferentes especies y niveles taxonómicos (Cheng *et al.*, 2016). En el presente estudio, se decidió diseñar un par de cebadores utilizando secuencias ITS únicamente de plantas terrestres de la familia *Asteraceae* a la cual pertenece la planta de guayule, estos cebadores además de ser diseñados para dicha familia también minimizan la posibilidad de amplificar secuencias ITS de hongos ya que solo se utilizaron secuencias de plantas para su diseño, omitiendo secuencias fúngicas. Para corroborar que los cebadores amplificaron correctamente la región de interés, se compararon las secuencias resultantes del estudio con la única secuencia de la misma región reportada en GenBank (KY397499.1) para guayule, por Jayasena *et al.*, (2017), en donde se confirmó que se trataba de la misma región amplificada exitosamente.

El diseño de los cebadores fue dirigido para que en un solo paso de amplificación se pudieran obtener tanto el ITS1 como el ITS2 (Figura 11). El amplicón fue de aproximadamente 700 pares de bases (pb), el diseño tuvo como resultado el cebador de inicio; *AGB-ITS-F* 5'- GTA GGT GAA CCT GCG GAA GG -3' y un cebador final *AGB-ITS-R* 5'- TTG TAG TCC CGC CTG ACC TG -3' los cuales se sintetizaron por medio de servicio externo en la empresa Integrated DNA Technologies (IDT), (<https://www.idtdna.com/pages>).



Figura 11. Área de diseño de cebadores para amplificar ITS1-ITS2 en la familia *Asteraceae*, A) *AGB-ITS-F*; B) *AGB-ITS-R*.

Se encontró que los cebadores AGB-ITS-F y AGB-ITS-R presentan mejor rendimiento en la reacción de PCR en comparación con cebadores universales reportados anteriormente, esto comprobado de manera cualitativa por la intensidad de la banda observada en un gel de agarosa al 1 %.

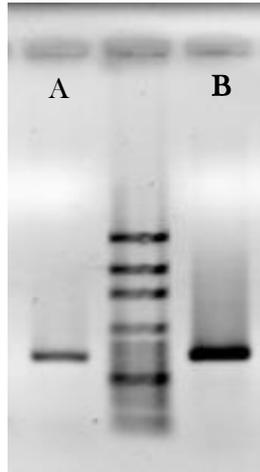


Figura 12. Amplificación de la región ITS1-ITS2 en *P. argentatum* con los cebadores universales ITS-U1-F e ITS-U4-R (A), y AGB-ITS-F y AGB-ITS-R (B).

Después de la amplificación de las muestras se realizó una purificación donde se obtuvieron las secuencias libres de inhibidores de la reacción Sanger (Figura 13), las cuales, se enviaron al laboratorio LANGEBIO en promedio 200 ng del amplicón liofilizado cumpliendo con las condiciones requeridas para esta metodología de secuenciación.

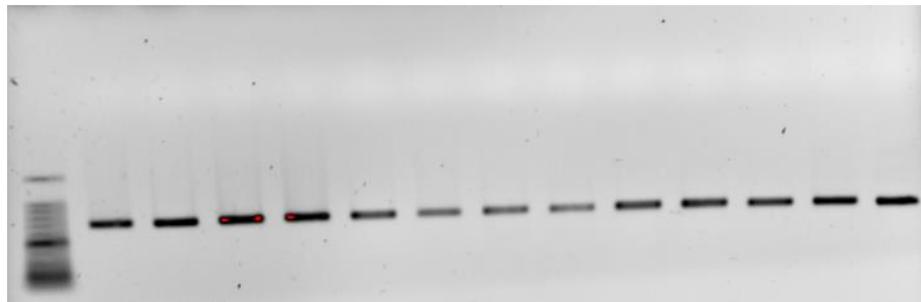


Figura 13. Purificación de la amplificación de ITS1-ITS2 con un tamaño aproximado a 700 pb.

V.4. Análisis bioinformáticos

V.4.1. Estadísticas descriptivas de las secuencias e índices de diversidad genética

En el programa BioEdit se realizó el alineamiento de las secuencias de guayule silvestre con las cinco regiones concatenadas (18S, ITS1, 5.8S, ITS2 y 26S) junto con las 6 secuencias de *P. incanum* y *P. rollinsianum* (Sims *et al.*, 2019), usando el método ClustalX con algunas correcciones manuales. El alineamiento dio como resultado una matriz de 645 pb, si las secuencias presentaban datos faltantes al inicio o al final, se representaron con signos de interrogación en los alineamientos utilizados para los programas.

La matriz de alineamiento de 649 pb corresponde a la longitud de datos concatenados, donde las bases 1 - 13 correspondieron a la región codificante 18S, las bases 14 - 274 al ITS1, las bases 275 - 437 a la región codificante 5.8S y finalmente las bases 438 - 645 al ITS2. Las frecuencias nucleotídicas observadas variaron entre todas las secuencias analizadas: 25.23 % a 23.59 % para A, de 23.91 % a 21.76 % para T, de 26.83 % a 25.99 % para G y de 26.56 % a 25.55 % para C, presentando un promedio de 24.10 %, 26.55 %, 25.81 %, 23.54 % para A, G, C y T respectivamente, y un porcentaje de GC de 52.36 % y AT de 47.64%, las frecuencias nucleotídicas para cada una de las muestras estudiadas se presentan en el Anexo 2.

Para la familia *Asteraceae*, se ha reportado una longitud de esta región que oscila entre las 650 y 750 pb, y frecuencias nucleotídicas de A: 25 %, T: 24 %, C: 26 %, y G: 25 % con porcentaje de GC: 51 % y AT: 49 % (Amar *et al.*, 2012). Las regiones ITS1-ITS2 excluyendo la región 5.8S, suman 470 pb lo cual es similar a lo reportado para otras especies de la familia *Asteraceae* como lo es el caso del género *Montanoa* que representa una longitud para esta región de 480 pb (Plovanich y Panero, 2004).

Del total de secuencias analizadas se encontraron 30 haplotipos diferentes, de los cuales 27 solo se presentan en *P. argentatum*, 2 haplotipos para *P. rollinsianum* y 1 haplotipo para *P. incanum* (Anexo 3. Haplotipos presentes en plantas de guayule silvestre), se calculó el índice de inserciones y deleciones (InDel), donde se encontró un total de 152 sitios con InDel corroborando la variación de la región estudiada.

Además del análisis de datos concatenados, se realizó una comparación entre las regiones ITS1 y la región ITS2 (Tabla 4), de acuerdo a los resultados de esta, se encontró que el ITS2 es más informativo que el ITS1, coincidiendo con los resultados de Lee *et al.*, (2017) y Chen *et al.*, (2010) quienes también encontraron más información discriminatoria en la región ITS2 de diferentes tés de Taiwán y plantas medicinales respectivamente, concluyendo la importancia de estos marcadores moleculares, sin embargo, para realizar los análisis consenso de la diversidad genética dentro de la especie se utilizaron datos concatenados, ya que, la inclusión de ambos ITS y la región codificante 5.8S nos brinda más información, esto de acuerdo a la mayoría de los estudios encontrados donde se estudia esta región (Maina *et al.*, 2019, Kumari *et al.*, 2018, Lee *et al.*, 2017, Kehie *et al.*, 2016, Ghada *et al.*, 2013, Wu *et al.*, 2013, Hao *et al.*, 2012, Kumar *et al.*, 2009).

De acuerdo con los resultados de los análisis descriptivos para cada población se encontró que, las poblaciones del estado de Zacatecas representan un gran valor informativo ya que presenta un bajo número de muestras y contiene información valiosa como lo son los haplotipos únicos encontrados en el estado, esto, en comparación con San Luis Potosí que es de los estados más informativos en cuanto a diversidad genética, pero con mayor número de muestras, por lo que el estado de Zacatecas representa un área de oportunidad para estudios posteriores.

Se realizaron pruebas de neutralidad a datos concatenados y a las regiones ITS1 e ITS2 por separado, donde se encontró que, el ITS1 no presenta una variación nucleotídica bajo presión de selección, es decir, los polimorfismos podrían ser resultado de una evolución neutral, para datos concatenados y para la secuencia del ITS2 presentaron significancia estadística, lo cual indica que estas regiones evolucionaron bajo una presión de selección.

La región ITS1 comprendió una longitud de 262 pb, mientras que la región ITS2 fue de 208 pb, sin embargo, en esta última región se encontraron un mayor número de mutaciones y sitios informativos. En general, el ITS2 presenta más sitios polifórmicos, (**¡Error! No se encuentra el origen de la referencia.**).

Tabla 4. Estadísticas descriptivas e índices de diversidad genética estimados de manera independiente a partir de cada región ITS y datos concatenados para la especie *P. argentatum*.

Datos de polimorfismos	ITS1	ITS2	Datos concatenados
Número de secuencias.	320	320	320
Longitud de alineamiento.	262	208	645
Porcentaje C+G	0.522	0.534	0.533
Sitios variables (S).	11	19	23
Total de mutaciones (Eta).	12	21	26
Diversidad nucleotídica (Pi) ± SD.	0.0024 ± 0.0001	0.0037 ± 0.0002	0.0026 ± 0.0001
Diferencias nucleotídicas (k).	0.64107	0.77645	1.31822
θ por secuencia.	1.89156	3.31023	4.09898
θ por sitio.	0.00733	0.01607	0.00831
Número de haplotipos (h).	13	21	27
Diversidad de haplotipos (hd) ± SD.	0.539 ± 0.014	0.571 ± 0.022	0.738 ± 0.013
SD: Desviación estándar			

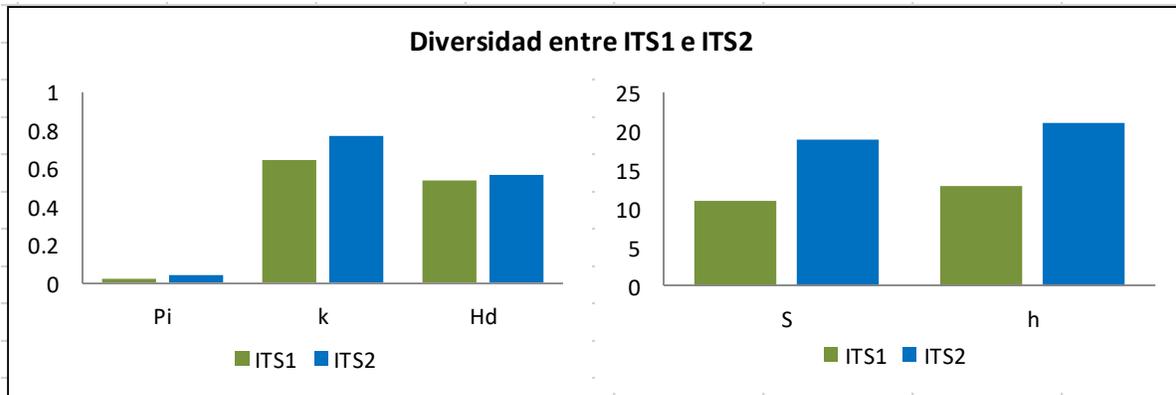


Figura 14. Comparativo de distintos índices de diversidad genética entre las regiones ITS1 e ITS2 en *P. argentatum*. Diversidad nucleotídica por sitio (Pi), diversidad nucleotídica (K), diversidad de haplotipos (Hd), sitios variables (S), número de haplotipos (h).

La diversidad de nucleótidos y haplotipos observadas en guayule presenta algunas diferencias con respecto a datos reportados para otras especies de la familia *Asteraceae* y plantas desérticas, para guayule se obtuvo un Hd: 0.738 ± 0.013 y Pi: 0.00267 ± 0.00014 . Comparado con lo reportado en otros estudios podemos considerar una alta diversidad genética para las poblaciones de guayule estudiadas. Para otras plantas desérticas como la jojoba (*Simmondsia chinensis*), se reportó una diversidad genética de Hd= 0.45, este valor es considerado como alto por los autores (Kumar *et al.*, 2019).

Con respecto a las distancias genéticas dentro de la especie, se encontró que existe poca variación genética (máxima variación 0.0092), dentro de las poblaciones (máxima variación 0.0039), al igual que entre los estados estudiados (máxima variación 0.0011).

Los resultados de la reconstrucción de haplotipos dieron un total de 352 haplotipos presentes en las secuencias, de los cuales 30 haplotipos son únicos. Estos haplotipos únicos se compararon genéticamente entre ellos, resultando de interés los haplotipos 23, 24, 25, 26 y 27 que se encuentran sólo en el estado de Zacatecas en las poblaciones Z1, Z3, Z3, Z5 y Z5 respectivamente, lo cual, hace referencia a que en estas poblaciones se tienen haplotipos únicos y con un alto valor genético a conservar.

En la estimación de estadística descriptiva para cada estado se encontró que los estados de San Luis Potosí y Zacatecas presentaron el mayor número de haplotipos y secuencias con InDel (Tabla 5).

El total de haplotipos encontrados en la reconstrucción, se utilizaron para la inferencia de la red de haplotipos, donde se puede observar 3 haplotipos probablemente ancestrales, ya que, presentan mayor frecuencia; el número de individuos que tenían presencia de cada haplotipo varió desde 107 hasta 1, donde los más frecuentes A, B y C tienen una frecuencia de 107, 88 y 57 respectivamente (**¡Error! No se encuentra el origen de la referencia.**).

Los haplotipos más frecuentes se encuentran en los 5 estados muestreados en el estudio, por lo que se puede entender que no existe una barrera física o geográfica para compartir esta información entre las poblaciones de la especie y presentar un flujo de genes continuo (Fehlberg y Ranker 2009); entonces, estos haplotipos han estado presentes a lo largo de la existencia de esta especie, a partir de estos se ramifican en diversos haplotipos menos frecuentes, entonces se infiere que cada haplotipo menos frecuente es una mutación del haplotipo más frecuente.

Tabla 5. Índices descriptivos estimados para poblaciones de guayule silvestres muestreadas para cada estado.

Índices descriptivos		Durango	San Luis Potosí	Nuevo León	Coahuila	Zacatecas
Haplotipos		10	15	6	4	12
Transiciones		6	9	4	1	8
Tranversiones		3	6	2	1	6
Sustituciones nucleotídicas por sitio		9	15	6	2	14
Indels		1	34	12	1	12
Sitios polimórficos		9	47	18	3	26
Composición nucleotídica	% C	25.70	25.73	25.76	25.7	25.74
	% T	23.76	23.74	23.69	23.81	23.66
	% A	23.98	24.01	24	23.94	24.02
	% G	26.56	26.53	26.55	26.55	26.58

Las secuencias únicas de diversidad de haplotipos para datos concatenados (Tabla 4), se tomaron como referencia para estimar las similitudes y divergencias entre los haplotipos encontrados.

En la Figura 14 podemos observar que, dentro de los haplotipos encontrados en guayule silvestre, se presentan dos grandes grupos, los cuales a su vez se ramifican de acuerdo con la divergencia genética que existe entre ellos. En el grupo I podemos encontrar a la gran mayoría de haplotipos presentes (17 haplotipos), donde se ramifican de acuerdo con la divergencia genética entre uno y otro, observándose que dicha divergencia es pequeña (longitud de rama), mientras que el grupo II incluye un menor número de haplotipos (10) relacionados entre ellos, pero con una mayor divergencia genética. Dentro de este segundo, se destaca el haplotipo 15, el cual presenta la rama más larga, es decir, el haplotipo más divergente dentro de este grupo, este haplotipo único se encuentra en la población 8 del estado de San Luis Potosí, de acuerdo con esto, es de gran importancia la conservación de este tipo de información única silvestre en plantas de guayule.

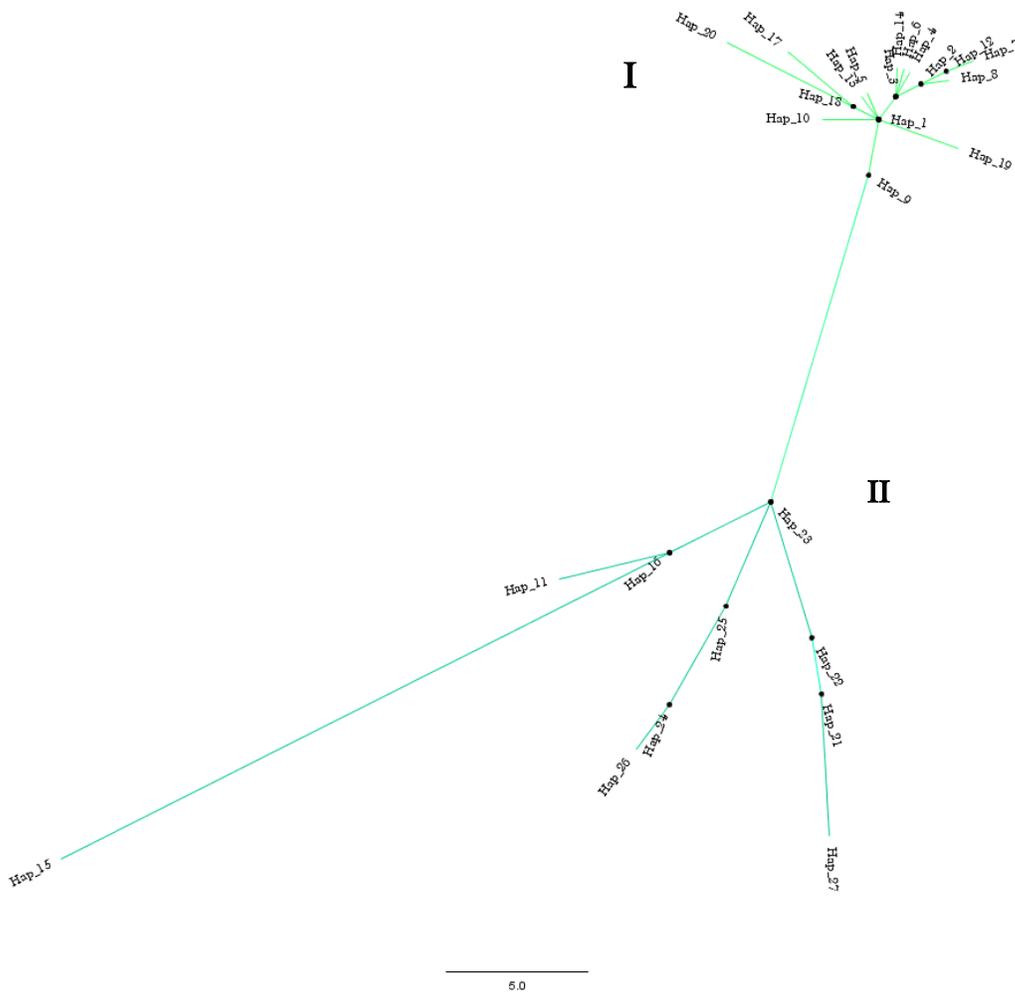


Figura 14. Relación y divergencia entre haplotipos presentes en plantas de guayule silvestre usando el programa Network.

V.4.2 Reconstrucción filogenética con inferencia bayesiana (IB) y Máxima Verosimilitud (ML).

Para realizar la reconstrucción filogenética se determinó que el modelo de sustitución nucleotídica que mejor describe a los datos fue el modelo de Jukes y Cantor (JC69); de acuerdo con los resultados obtenidos en jModelTest; dicho modelo asume la misma frecuencia para los cuatro nucleótidos y una sola tasa de sustitución nucleotídica (Jukes y Cantor, 1969).

Tanto en el árbol de ML y de IB se pudo establecer claramente la monofila de las especies estudiadas con altos valores de soporte (Probabilidad posterior; PP y Valores de

bootstrap; VB) para cada clado, lo cual deja en claro la eficiencia de los ITS como marcadores moleculares capaces de diferenciar taxones a nivel de especie dentro de un mismo género.

En el árbol de ML se agruparon todas las muestras de guayule silvestre en un clado sin ninguna diferenciación, por lo que en este análisis se tendrían que adicionar una(s) secuencia(s) o agregar más regiones a las utilizadas para poder confirmar la estructura dada en el análisis con IB; también podría explicarse este resultado por la naturaleza del método, el cual evaluó la probabilidad de las secuencias dadas de guayule silvestre encontradas con base a la búsqueda de 100 árboles con la mayor verosimilitud no fueron suficientes para encontrar diferencias. En este ámbito, la IB ofreció resultados más robustos y consistentes, en este análisis se calcula la probabilidad posterior de un árbol con base a los datos obtenidos de las plantas silvestres de guayule.

Para las muestras de guayule no se presenta alguna estructura o diferenciación con respecto a la ubicación geográfica de estas ya que, en ambos árboles se encuentran todas las muestras en un solo clado. Se realizó la prueba de mantel en donde el coeficiente de correlación mostró un valor positivo débil ($r = 0.1590$ $P < 0.01$), esto indica que existe un leve aislamiento por distancia geográfica pero este no representa un factor determinante en la estructura genética encontrada en las muestras, también se puede entender este resultado con la prueba de AMOVA donde la mayor variación genética se encuentra dentro de las poblaciones (Ver apartado V.4.3.), en otras palabras, se entiende que las poblaciones más cercanas entre si presentan una mayor interacción genética que las poblaciones más lejanas, por lo que esta interacción genética marca la gran diversidad genética encontrada (Li *et al.*, 2013), esto concuerda con los resultados de Cruz *et al.*, (2017), en donde por el genotipado de plantas de guayule del banco de germoplasma de los Estados Unidos, (USDA), encontraron que los individuos analizados no presentaban un alta diferenciación genética, por lo que el estudio y conservación del germoplasma silvestre de guayule es de inminente importancia por la información que esta aporta a la caracterización de la planta.

En este estudio se reveló una alta diversidad genética en plantas silvestres de guayule, a menudo dentro de las poblaciones existe una fuerte asociación entre la diversidad y el

modo de reproducción. Estas, dan lugar a la generación de hipótesis sobre cómo surge y se mantiene la diversidad en las poblaciones naturales (Kolář *et al.*, 2017) como lo es la hibridación de plantas para enriquecerlas genéticamente.

Para la construcción de árboles con IB se utilizó la simulación MCMC, después de aplicar un “burn-in” de 25%, se obtuvieron 1,875 muestras de árboles y valores de parámetros. Se obtuvo una desviación estándar de las frecuencias de biparticiones menor a 0.01 por lo cual se asume que las cadenas de Markov convergieron; asimismo, se generó un árbol consenso cuyos clados o grupos inferidos estuvieron presentes en al menos el 50% del total de los árboles muestreados.

Dado que los resultados del análisis de IB y ML son muy similares, se reportan ambos resultados en el árbol obtenido en el análisis de IB (**¡Error! No se encuentra el origen de la referencia.**), en el Anexo 4 se presenta el árbol del análisis de ML. Los resultados del análisis de ML se reportaron en el árbol de IB bajo cada rama en color verde para cada clado.

Los análisis infirieron tres clados, los cuales corresponden a las 3 especies analizadas; el clado I corresponde a la especie *P. rollinsianum*, el cual se utilizó como grupo externo para ambos análisis, donde se puede observar que el enraizado del árbol ocurre de manera correcta; el clado II corresponde a la especie *P. incanum* y el clado III a *P. argentatum*.

Dentro del clado II encontramos a los tres ejemplares colectados de la especie *P. incanum* (M1. M2 y M3), con el máximo valor para PP de 1 (azul), y un buen VB de 98 (verde).

Para el clado III tenemos todos los ejemplares de guayule, con una PP de 0.979 (azul) y un VB de 100 (verde), para el análisis con IB encontramos diferentes longitudes de ramas, estas, hacen referencia a las divergencias genéticas encontradas en estos ejemplares silvestres de guayule para este análisis, de los cuales destacan D11P3, SL6P1, SL9P1, NL2P4, Z5P3 y SL6P3, siendo esta última la secuencia más divergente (de acuerdo a la longitud de la rama), dentro del clado III se genera un pequeño clado que agrupa a los ejemplares C1P2, C3P3 y C3P4, aunque con valores de soporte bajos (PP de 0.676 y VB

de 55, datos no mostrados), podrían representar una variación genética a considerar, por lo que se recomienda realizar las análisis con estos ejemplares y verificar los resultados obtenidos.

El clado III, para el análisis ML fue encontrado en un clado, es decir, que mediante este análisis no se pudieron detectar la variación genética intraespecífica de guayule, en comparación con la IB, en ambos árboles se presentan los nodos de las muestras C1P2, C3P3 y C3P4, sin embargo, los valores (No mostrados) de soporte son insuficientes para poder asegurar estas diferencias genéticas.

En ambos árboles se pudo corroborar fuertemente la monofila para cada especie analizada incluyendo a cada especie en un clado bien soportado tanto en los valores de PP en el análisis de IB y los VB para el análisis ML.

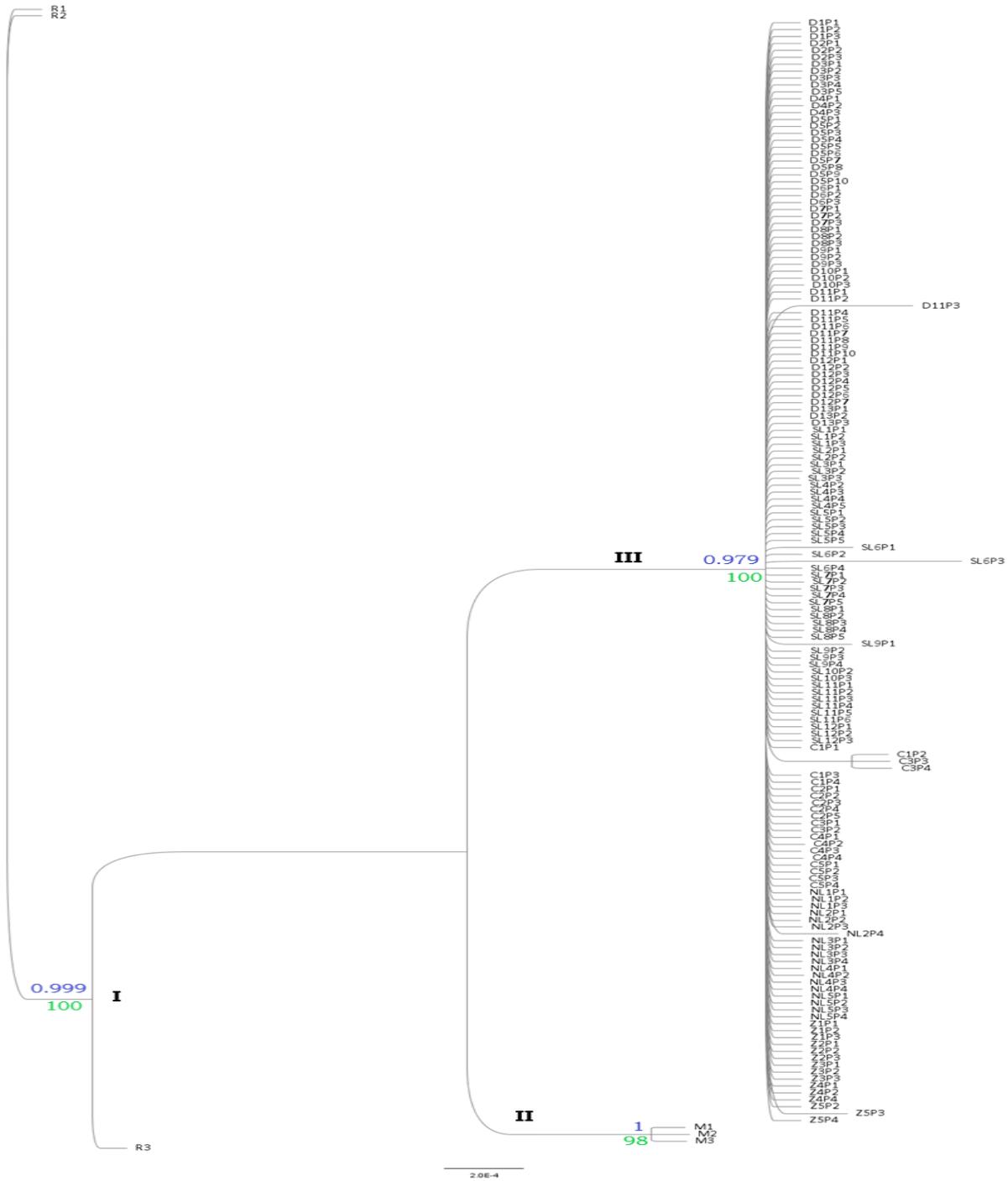


Figura 16. Reconstrucción filogenética con Inferencia Bayesiana y Máxima verosimilitud (ver Anexo 4), de tres especies del género *Parthenium* a partir de 160 secuencias de las regiones genómicas 18S, ITS1, 5.8S, ITS2 (645 bp). Valores de probabilidad posterior (IB) y bootstrap (ML) se muestran por encima y debajo de las ramas, respectivamente.

V.4.3 Análisis AMOVA para plantas silvestres de guayule.

Como resultado del análisis molecular de varianza (AMOVA) (Tabla 6), se puede observar que la mayor divergencia genética (81.19 %), se encuentra dentro de las poblaciones silvestres de guayule, mientras que la menor divergencia genética (2.26 %) está presente entre los estados analizados en este estudio.

Tabla 6. Análisis AMOVA para poblaciones silvestres de guayule.

Fuente de variación	Grados de libertad	Suma de cuadrados	Varianza	Porcentaje de variación (%)
Entre estados	4	9.749	0.0156	2.26
Entre poblaciones dentro de estados	35	51.247	0.1142	16.55
Dentro de las poblaciones (rodal)	280	156.822	0.56008	81.19
Total	319	217.818	0.68988	

Estos resultados se podrían entender por los hábitos de reproducción de *P. argentatum*, ya que, esta planta se reproduce mediante apomixis facultativa, es decir, presenta una reproducción asexual mediante semillas, sin embargo, este tipo de reproducción permite eventos de reproducción sexual esporádica, lo cual facilita al intercambio genético dentro de una misma población por la cercanía de los ejemplares y los intercambios genéticos entre ejemplares a largas distancias (poblaciones diferentes) son menos probables a suceder (Cruz *et al.*, 2017).

V.4.3 Red de haplotipos en plantas silvestres de guayule.

Se realizó la estimación de red de haplotipos (**¡Error! No se encuentra el origen de la referencia.**), es decir, la frecuencia de cada uno de los haplotipos en los estados estudiados, que, a su vez nos permite inferir información histórica sobre las poblaciones, se utilizaron las secuencias de la especie *P. rollinsianum* y *P. incanum* como grupo

externo, en donde, podemos observar la divergencia entre haplotipos y en casos específicos las similitudes entre los haplotipos presentes en las distintas especies.

Cada gráfico circular representa a un haplotipo, en este se representan con un color diferente a cada estado de acuerdo con su frecuencia (**¡Error! No se encuentra el origen de la referencia.**). Podemos observar que son tres haplotipos los que ocurren con mayor frecuencia (A, B y C) en las poblaciones muestreadas; cada uno de estos haplotipos presenta un número diferente de ramificaciones conectando otros haplotipos menos frecuentes, estas se pueden entender como una divergencia y/o mutación del haplotipo más frecuente. Cada punto rojo representa a un ancestro hipotético inferido por el programa.

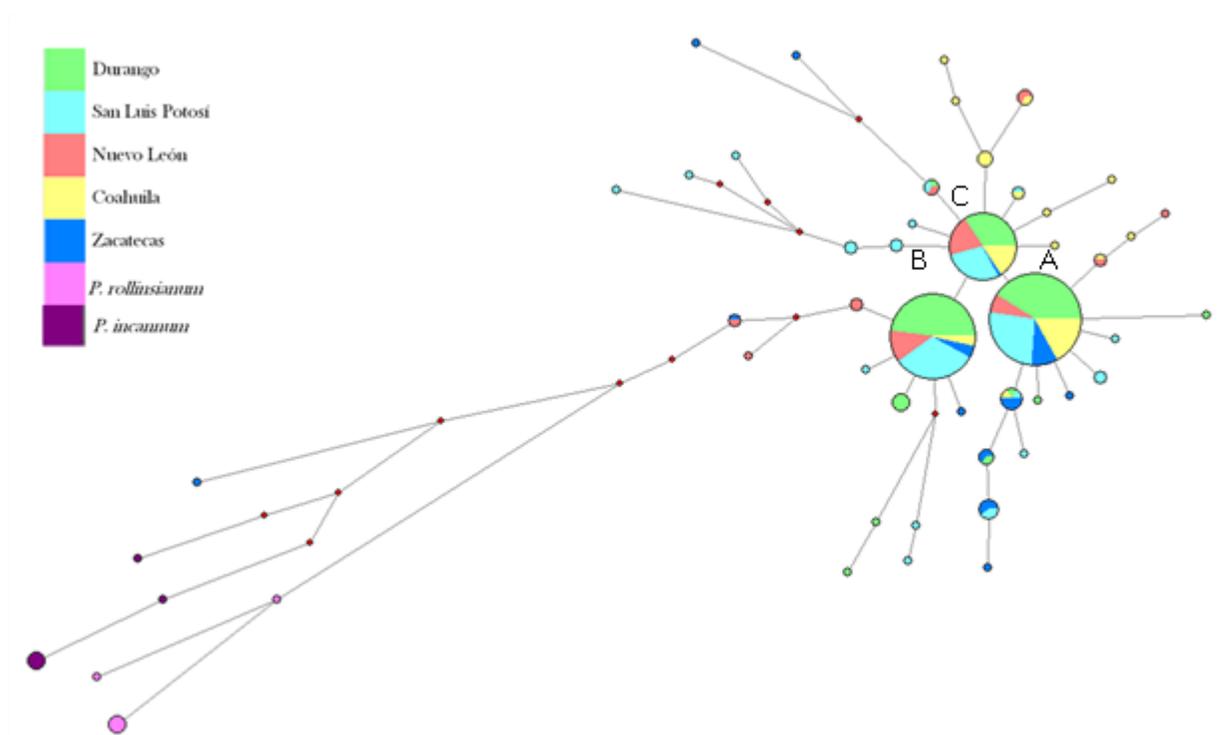


Figura 17. Red de haplotipos inferida a partir del programa Network utilizando el algoritmo MJ, con grupos externos (*P. incanum* y *P. rollinsianum*, círculos violetas y morados respectivamente), donde, A, B y C representan los haplotipos más frecuentes.

Según los resultados de la red de haplotipos el total de 160 individuos analizados presentan un flujo de información genética (vea **¡Error! No se encuentra el origen de la referencia.**) ya sea en eventos recientes o históricos, lo cual se puede entender cómo, no tener una barrera física o geográfica para dicho flujo, esto también se confirmó con la prueba de mantel realizada en el programa Arlequín el cual no encontró correlación significativa entre las distancias genéticas por pares (modelo JC), y las distancias geográficas para cada población; estos resultados se respaldan con los resultados del AMOVA, que indican que existe una mayor diversidad genética dentro de las poblaciones (81.19 %) comparado la variación encontrada entre los estados estudiados (2.26 %).

De acuerdo a la diversificación o mutación de los haplotipos más ancestrales se podría entender que, dentro de las poblaciones se tiene una diversificación la cual podría ser causada por la adaptación de las poblaciones al medio en donde se desarrollan comenzando como mutaciones epigenéticas y fueron heredadas a nuevas generaciones hasta formar parte del genoma de las especies silvestres muestreadas (Fehlberg y Ranker, 2009), dicha divergencia se puede observar a nivel de población, pero, dentro del conjunto de poblaciones no representa una diferencia genética significativa esto corroborado con el AMOVA (Tabla 6).

Por la ubicación en la red, se puede inferir que el haplotipo para la muestra Z5P3 presenta mayor similitud con la especie *P. incanum*, esto se puede observar también en los árboles filogenéticos de IB y ML, se podría entender como la razón del alargamiento en su rama en el árbol construido con IB, de acuerdo con su hábito reproductivo, se sabe que plantas de guayule son capaces de reproducirse con plantas genéticamente similares como lo es *P. incanum* (Sanchez *et al.*, 2014).

Ya que sabemos que el guayule tiene una reproducción por apomixis facultativa con eventos esporádicos de recombinación genética sexual, también se sabe que las especies de *P. argentatum* y *P. incanum* son muy cercanas genéticamente y de manera natural se puede dar intercambio genético por sus hábitos de reproducción, este intercambio de información genética podría ser la razón de la cercanía de la muestra Z5P3 con la especie *P. incanum*; se podría relacionar con su hábito de reproducción, ya

que se sabe que en una planta apomítica, sin importar su ploidía, genera embriones viables con ploidía variable (Cruz *et al.*, 2017).

V.4.4. Estimación de tiempos de divergencia.

La divergencia de las especies *P. argentatum*, *P. incanum*, *P. rollinsianum* y *H. annuus*, se estima que ocurrió aproximadamente hace 15.62 millones de años (Ma) con un intervalo del 95 % de densidad posterior más alta (95 % HPD) de 10.71 - 20.87 Ma, cabe mencionar que la convergencia y estacionariedad de las cadenas de Markov se corroboró graficando los valores de los parámetros en el programa Tracer.

VI. CONCLUSIONES

Según los valores de bootstrap y la probabilidad posterior obtenidos en la inferencia filogenética bajo los métodos de máxima verosimilitud e inferencia bayesiana se comprobó la eficacia de los marcadores ITS para discriminar a nivel de especie, además de ser útiles en encontrar diferencias a nivel intraespecífico, por lo que podemos proponer a la región ITS1-ITS2 como un nuevo código de barras universal para la identificación de una gama más amplia de taxones de plantas.

De acuerdo con los datos encontrados, las poblaciones silvestres de guayule no presentan barreras físicas o geográficas para un flujo genético, con base en esto, la población de guayule presente en el desierto chihuahuense es grande y no ha sufrido cuellos de botella genéticos, el germoplasma muestreado es de gran valor para conservar por la diversidad genética contenida.

Se encontró que el guayule presenta una gran diversidad genética, además, se identificaron haplotipos ancestrales presentes en los cinco estados muestreados; también se encontraron haplotipos únicos en distintas plantas, esta información genética es invaluable y de gran importancia para su resguardo.

Dentro de los estados de Zacatecas y San Luis Potosí presentan los datos más relevantes de diversidad genética, proponiendo las poblaciones SL6, SL9 y Z5, ya que son las más informativas y presentan la mayor divergencia genética en comparación al resto de muestras analizadas.

Anexos

Anexo 1. Secuencias utilizadas para diseño de cebadores para la familia Asteraceae

Código GenBank	Especie	Código GenBank	Especie
>KJ510658.1	<i>Alloispermum scabrifolium</i>	>KY397492.1	<i>Ligularia przewalskii</i>
>KX671845.1	<i>Ambrosia trifida</i>	>DQ101217.1	<i>Lindheimera texana</i>
>KY397478.1	<i>Anacyclus depressus</i>	>FJ697005.1	<i>Melampodium leucanthum</i>
>DQ444723.1	<i>Arctotis aspera</i>	>EF108398.1	<i>Melampodium paludosum</i>
>KY397479.1	<i>Arctotis gumbletonii</i>	>KY397494.1	<i>Olearia lyrata</i>
>DQ444729.1	<i>Arctotis venusta</i>	>KP718466.1	<i>Oncosiphon africanus</i>
>FM177852.1	<i>Arnica chamissonis</i>	>KY397495.1	<i>Osteospermum ecklonis</i>
>HM032736.1	<i>Arnica montana</i>	>KY397496.1	<i>Osteospermum jucundum</i>
>AF374892.1	<i>Bahia absinthifolia</i>	>KY397497.1	<i>Osteospermum sinuatum</i>
>AF465849.1	<i>Carramboa trujillensis</i>	>KY397498.1	<i>Othonna arborescens</i>
>KY397482.1	<i>Centaurea cyanoides</i>	>KY397499.1	<i>Parthenium argentatum</i>
>HQ407430.1	<i>Centaurea montana</i>	>KX671862.1	<i>Parthenium auriculatum</i>
>KT249764.1	<i>Centaurea phrygia</i>	>KY968852.1	<i>Parthenium hysterophorus</i>
>AB032053.1	<i>Chromolaena odorata</i>	>KX671863.1	<i>Parthenium integrifolium</i>
>AY538652.1	<i>Echinops sicutus</i>	>KJ510657.1	<i>Sabazia sarmentosa</i>
>AY196731.1	<i>Engelmannia pinnatifida</i>	>KY397501.1	<i>Sanvitalia procumbens</i>
>AF465855.1	<i>Espeletia schultzei</i>	>KY397502.1	<i>Schoenia filifolia</i>
>AF374871.1	<i>Eutetras palmeri</i>	>MF290124.1	<i>Senecio magnificus</i>
>FN823080.1	<i>Tanacetum parthenium</i>	>KY397505.1	<i>Senecio pinnatifolius var. Maritimus</i>
>FJ696962.1	<i>Galinsoga parviflora</i>	>KY397506.1	<i>Senecio pinnatifolius var. pinnatifolius</i>
>GU818550.1	<i>Galinsoga quadriradiata</i>	>MF290162.1	<i>Senecio pinnatifolius</i>
>KY397486.1	<i>Gazania splendens</i>	>AB457301.1	<i>Stevia rebaudian</i>
>KY397487.1	<i>Guizotia abyssinica</i>	>KY397507.1	<i>Streptoglossa macrocephala</i>
>KF767534.1	<i>Helianthus annuus</i>	>KY397508.1	<i>Tagetes tenuifolia</i>
>JX121518.1	<i>Helianthus maximiliani</i>	>KY397500.1	<i>Tanacetum coccineum</i>
>KX671859.1	<i>Heliopsis helianthoides</i>	>EF577320.1	<i>Tanacetum parthenium</i>
>KY397490.1	<i>Hymenopappus tenuifolius</i>	>KY397509.1	<i>Tetragonotheca texana</i>
>KY231643.1	<i>Ichthyothere scandens</i>	>KY397510.1	<i>Waitzia nitida</i>
>KY696303.1	<i>Inula ensifolia</i>	>FJ980320.1	<i>Xanthium sibiricum</i>
>LT855156.1	<i>Jurinea heterophylla</i>	>KY397488.1	<i>Xerochrysum bracteatum</i>
>DQ188075.1	<i>Layia heterotricha</i>	>KY397489.1	<i>Xerochrysum macranthum</i>
>HQ416383.1	<i>Liatris spicata</i>	>KY397511.1	<i>Xerochrysum viscosum</i>

Anexo 2. Frecuencia nucleotídica por población.

Frecuencia nucleotídica				
Planta	% A	% G	% C	% T
D1P1	23.9130%	26.5528%	25.6211%	23.9130%
D1P2	24.0683%	26.5528%	25.7764%	23.6025%
D1P3	23.9875%	26.6355%	25.7009%	23.6760%
D2P1	23.9502%	26.5941%	25.8165%	23.6392%
D2P2	23.9502%	26.5941%	25.8165%	23.6392%
D2P3	23.9875%	26.6355%	25.7009%	23.6760%
D3P1	23.9130%	26.5528%	25.6211%	23.9130%
D3P2	23.9130%	26.5528%	25.6211%	23.9130%
D3P3	23.9502%	26.5941%	25.6610%	23.7947%
D3P4	23.9130%	26.5528%	25.6211%	23.9130%
D3P5	23.9875%	26.6355%	25.7009%	23.6760%
D4P1	23.9130%	26.5528%	25.6211%	23.9130%
D4P2	23.9130%	26.5528%	25.6211%	23.9130%
D4P3	23.9130%	26.5528%	25.6211%	23.9130%
D5P1	23.9130%	26.5528%	25.6211%	23.9130%
D5P2	23.9502%	26.5941%	25.8165%	23.6392%
D5P3	23.9875%	26.6355%	25.7009%	23.6760%
D5P4	23.9875%	26.6355%	25.7009%	23.6760%
D5P5	24.0683%	26.5528%	25.7764%	23.6025%
D5P6	23.9502%	26.5941%	25.8165%	23.6392%
D5P7	23.9875%	26.6355%	25.8567%	23.5202%
D5P8	23.9502%	26.5941%	25.6610%	23.7947%
D5P9	24.0250%	26.6771%	25.5850%	23.7129%
D5P10	24.0250%	26.6771%	25.7410%	23.5569%
D6P1	23.9130%	26.5528%	25.6211%	23.9130%
D6P2	24.2089%	26.4241%	25.7911%	23.5759%
D6P3	23.9875%	26.6355%	25.8567%	23.5202%

D7P1	23.9502%	26.5941%	25.8165%	23.6392%
D7P2	23.9130%	26.5528%	25.6211%	23.9130%
D7P3	23.9502%	26.5941%	25.8165%	23.6392%
D8P1	23.9502%	26.5941%	25.8165%	23.6392%
D8P2	23.9502%	26.5941%	25.8165%	23.6392%
D8P3	23.9502%	26.5941%	25.8165%	23.6392%
D9P1	24.0683%	26.5528%	25.7764%	23.6025%
D9P2	23.9130%	26.5528%	25.6211%	23.9130%
D9P3	24.0683%	26.5528%	25.7764%	23.6025%
D10P1	23.9130%	26.5528%	25.6211%	23.9130%
D10P2	24.0683%	26.5528%	25.7764%	23.6025%
D10P3	24.0683%	26.5528%	25.7764%	23.6025%
D11P1	24.1433%	26.6355%	25.7009%	23.5202%
D11P2	23.9502%	26.5941%	25.8165%	23.6392%
D11P3	24.1433%	26.6355%	25.5452%	23.6760%
D11P4	24.1058%	26.5941%	25.6610%	23.6392%
D11P5	24.1058%	26.5941%	25.6610%	23.6392%
D11P6	23.9875%	26.4798%	25.8567%	23.6760%
D11P7	24.1058%	26.5941%	25.6610%	23.6392%
D11P8	23.9502%	26.5941%	25.8165%	23.6392%
D11P9	23.9502%	26.5941%	25.8165%	23.6392%
D11P10	23.9130%	26.5528%	25.7764%	23.7578%
D12P1	23.9130%	26.5528%	25.6211%	23.9130%
D12P2	23.9130%	26.5528%	25.6211%	23.9130%
D12P3	23.9130%	26.5528%	25.6211%	23.9130%
D12P4	24.0683%	26.5528%	25.7764%	23.6025%
D12P5	23.9130%	26.5528%	25.6211%	23.9130%
D12P6	23.9130%	26.5528%	25.6211%	23.9130%
D12P7	23.9502%	26.5941%	25.6610%	23.7947%
D13P1	24.0625%	26.4063%	25.7813%	23.7500%

D13P2	23.9502%	26.5941%	25.8165%	23.6392%
D13P3	23.9875%	26.6355%	25.7009%	23.6760%
SL1P1	23.9502%	26.5941%	25.8165%	23.6392%
SL1P2	23.9130%	26.5528%	25.6211%	23.9130%
SL1P3	23.9502%	26.5941%	25.8165%	23.6392%
SL2P1	23.9502%	26.5941%	25.6610%	23.7947%
SL2P2	24.0683%	26.5528%	25.7764%	23.6025%
SL3P1	24.4019%	25.9968%	26.1563%	23.4450%
SL3P2	23.8690%	26.6771%	25.7410%	23.7129%
SL3P3	23.8318%	26.6355%	25.8567%	23.6760%
SL4P2	24.0683%	26.5528%	25.7764%	23.6025%
SL4P3	24.0683%	26.5528%	25.7764%	23.6025%
SL4P4	23.9502%	26.5941%	25.6610%	23.7947%
SL4P5	23.9130%	26.5528%	25.6211%	23.9130%
SL5P1	24.0683%	26.5528%	25.7764%	23.6025%
SL5P2	24.0683%	26.5528%	25.7764%	23.6025%
SL5P3	24.1058%	26.4386%	25.8165%	23.6392%
SL5P4	24.1058%	26.5941%	25.6610%	23.6392%
SL5P5	23.9502%	26.5941%	25.6610%	23.7947%
SL6P1	23.9875%	26.6355%	25.7009%	23.6760%
SL6P2	23.9130%	26.5528%	25.7764%	23.7578%
SL6P3	23.7947%	26.4386%	25.9720%	23.7947%
SL6P4	24.0683%	26.5528%	25.7764%	23.6025%
SL7P1	23.9502%	26.5941%	25.8165%	23.6392%
SL7P2	23.9502%	26.5941%	25.6610%	23.7947%
SL7P3	23.9130%	26.5528%	25.6211%	23.9130%
SL7P4	24.4409%	26.5176%	25.5591%	23.4824%
SL7P5	23.9130%	26.5528%	25.6211%	23.9130%
SL8P1	24.1776%	26.8092%	25.9868%	23.0263%
SL8P2	23.9130%	26.5528%	25.6211%	23.9130%

SL8P3	23.9130%	26.5528%	25.6211%	23.9130%
SL8P4	23.9502%	26.5941%	25.8165%	23.6392%
SL8P5	23.9130%	26.5528%	25.6211%	23.9130%
SL9P1	23.7578%	26.7081%	25.6211%	23.9130%
SL9P2	23.9130%	26.5528%	25.6211%	23.9130%
SL9P3	23.9130%	26.5528%	25.6211%	23.9130%
SL9P4	23.9130%	26.5528%	25.6211%	23.9130%
SL10P2	24.2472%	26.4659%	25.9905%	23.2964%
SL10P3	23.9130%	26.5528%	25.6211%	23.9130%
SL11P1	24.2857%	26.3492%	26.0317%	23.3333%
SL11P2	24.0625%	26.4063%	25.6250%	23.9063%
SL11P3	23.9875%	26.4798%	25.8567%	23.6760%
SL11P4	23.9502%	26.4386%	25.6610%	23.9502%
SL11P5	23.9063%	26.4063%	25.7813%	23.9063%
SL11P6	24.1758%	26.3736%	25.7457%	23.7049%
SL12P1	23.9502%	26.5941%	25.8165%	23.6392%
SL12P2	23.9502%	26.5941%	25.8165%	23.6392%
SL12P3	23.9502%	26.5941%	25.8165%	23.6392%
C1P1	23.5938%	26.7188%	25.9375%	23.7500%
C1P2	23.8318%	26.6355%	25.8567%	23.6760%
C1P3	23.9875%	26.6355%	25.7009%	23.6760%
C1P4	23.9502%	26.5941%	25.8165%	23.6392%
C2P1	23.9502%	26.5941%	25.6610%	23.7947%
C2P2	23.9130%	26.5528%	25.6211%	23.9130%
C2P3	23.9130%	26.5528%	25.6211%	23.9130%
C2P4	24.3671%	26.4241%	25.7911%	23.4177%
C2P5	23.9130%	26.5528%	25.6211%	23.9130%
C3P1	23.9130%	26.5528%	25.6211%	23.9130%
C3P2	25.1376%	26.2385%	26.4220%	22.2018%
C3P3	24.8175%	26.8248%	26.4599%	21.8978%

C3P4	25.0000%	26.8248%	26.2774%	21.8978%
C4P1	25.0000%	26.6423%	26.2774%	22.0803%
C4P2	25.0916%	26.5568%	26.5568%	21.7949%
C4P3	24.2089%	26.4241%	25.9494%	23.4177%
C4P4	24.2472%	26.4659%	25.8320%	23.4548%
C5P1	24.2472%	26.4659%	25.8320%	23.4548%
C5P2	24.2472%	26.4659%	25.8320%	23.4548%
C5P3	24.2472%	26.4659%	25.8320%	23.4548%
C5P4	24.2472%	26.4659%	25.8320%	23.4548%
NL1P1	24.2857%	26.5079%	25.8730%	23.3333%
NL1P2	24.2857%	26.5079%	25.8730%	23.3333%
NL1P3	24.2472%	26.4659%	25.8320%	23.4548%
NL2P1	24.2089%	26.4241%	25.9494%	23.4177%
NL2P2	24.2857%	26.5079%	25.8730%	23.3333%
NL2P3	25.2285%	26.6910%	26.3254%	21.7550%
NL2P4	24.2089%	26.5823%	25.9494%	23.2595%
NL3P1	24.2857%	26.5079%	25.8730%	23.3333%
NL3P2	24.3671%	26.4241%	25.9494%	23.2595%
NL3P3	24.2857%	26.5079%	25.8730%	23.3333%
NL3P4	24.2472%	26.4659%	25.8320%	23.4548%
NL4P1	23.6301%	26.5411%	26.0274%	23.8014%
NL4P2	24.2089%	26.4241%	25.9494%	23.4177%
NL4P3	24.2089%	26.4241%	25.9494%	23.4177%
NL4P4	24.3671%	26.4241%	25.9494%	23.2595%
NL5P1	24.3631%	26.5924%	25.9554%	23.0892%
NL5P2	24.3243%	26.5501%	25.9141%	23.2114%
NL5P3	24.3631%	26.5924%	26.1146%	22.9299%
NL5P4	24.2857%	26.5079%	26.0317%	23.1746%
Z1P1	25.0000%	26.6423%	26.2774%	22.0803%
Z1P2	24.2089%	26.4241%	25.7911%	23.5759%

Z1P3	24.2857%	26.5079%	25.7143%	23.4921%
Z2P1	24.2089%	26.4241%	25.7911%	23.5759%
Z2P2	24.2472%	26.4659%	25.8320%	23.4548%
Z2P3	24.2089%	26.4241%	25.7911%	23.5759%
Z3P1	24.2472%	26.4659%	25.9905%	23.2964%
Z3P2	24.0829%	26.6348%	25.9968%	23.2855%
Z3P3	24.2857%	26.5079%	25.8730%	23.3333%
Z4P1	24.4057%	26.4659%	25.8320%	23.2964%
Z4P2	24.4019%	26.6348%	25.8373%	23.1260%
Z4P4	24.2472%	26.4659%	25.8320%	23.4548%
Z5P2	24.2424%	26.6348%	25.8373%	23.2855%
Z5P3	24.5192%	26.6026%	26.1218%	22.7564%
Z5P4	24.2472%	26.4659%	25.9905%	23.2964%
R1	24.2236%	26.2422%	25.7764%	23.7578%
R2	24.2236%	26.2422%	25.7764%	23.7578%
R3	24.2613%	26.4386%	25.5054%	23.7947%
M1	24.0506%	26.7405%	26.5823%	22.6266%
M2	24.0506%	26.7405%	26.5823%	22.6266%
M3	24.0506%	26.7405%	26.5823%	22.6266%

Anexo 3. Haplotipos presentes en plantas de guayule silvestre

Ext_Id Frecuencia.

Hap_1 114

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACTGA-
GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTTCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTCGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CTAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_2 98

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTTCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTCGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_3 66

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTTCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTCGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CTAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_4 **3**

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTGACGTGTGTTTAT
GGCTGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CTAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_5 **1**

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACTGA-
GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CGAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_6 **1**

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CTAGGTGTGGTTGGCCTAAATAGGAGTCTCGCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_7 **2**

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACTGA-

GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTGCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCGTGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_8 4

GAAGGATCATTGTGCGAACCCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTGCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTGCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_9 7

GAAGGATCATTGTGCGAACCCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACTGA-
GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTGCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTC-

Hap_10 1

GAAGGATCATTGTGCGAACCCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACTGA-
GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTGCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC

CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAAAACGTAGCAAAATGCGATACTTGG
TGTGAATTGCAAAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTCGCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CTAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_11 1

GAAGGATCATTGTCTGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTTCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAACAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTCGCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAAATTGGTCTCCCATGCC
AAAGTGTGGTTGGCCTAAATAGGAATCTCCCCAAAAGGGACGCACGTCTAGTGGTG
GTTGATAACACAGTCGTCTCATGTCTGCGTTTTTCATTCTTGGGTGTGGATGCTCTT
AAAATACCCC-----

Hap_12 4

GAAGGATCATTGTCTGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTTCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTCGCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCGTGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_13 1

GAAGGATCATTGTCTGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACTGA-
GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTTCGACGTGTGTTAAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA

GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CTAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTGCGTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_14 1

GAAGGATCATTGTCTGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTGCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAATAAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAATAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CTAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTGCGTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_15 1

GAAGGATCATTGTCTGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTGCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAATAAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAATAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CTAGGTGTGGATGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTGCGTTCATTCTTGGGT-----

Hap_16 2

GAAGGATCATTGTCTGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTGCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAATAAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAATAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAAATTGGTCTCCCATGCC
TAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGGT

GGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTCT
TAAATACCCC-----

Hap_17 1

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACTGA-
GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTTCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAATAAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAACAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTCGCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAAATTGGTCTCCCATGCC
AAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGGT
GGTTGATAAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTCT
TAAATACCCCGACGTGTTGTCT

Hap_18 2

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACTGA-
GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTTCGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAATAAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTCGCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAAATTGGTCTCCCATGCC
TAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGGT
GGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTCT
TAAATACCCCGACGTGTTGTCT

Hap_19 1

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACTGA-
GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTTCGACGTGTGTTTAT
GGCCCCCATAACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAATAAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTCGCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCCGACGTGTTGTCT

Hap_20 1

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACTGA-
GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTGACGTGTGTTTAT
GGCCCCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAACAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAAATTGGTCTCCCATGCC
CAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGGT
GGTTGATAACACAGTCGTCTCATGTGCGTTTTTCATTCTTGGGTGTGGATGCTCT
TAAAATACCCCGACGTGTTGT--

Hap_21 2

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTGACGTGTGTTTAT
GGCTGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCCGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGGCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCC-----

Hap_22 1

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGACCGAAGCTTATGTTTCGAGCCTCGTGAGGCCTTGTGACGTGTGTTTAT
GGCCGCACCATACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCCGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGGCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCC-----

Hap_23 1

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACTGA-

GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTGCGACGTGTGTTTAT
GGCCGCACCATAACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCATGCC
AAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCC-----

Hap_24 1

GAAGGATCATTGTGCGAACCCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGACGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTGCGACGTGTGTTTAT
GGCTGCACCATAACATATGGGGCATCATGGATACATGTCGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCGTGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCC-----

Hap_25 1

GAAGGATCATTGTGCGAACCCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTGCGACGTGTGTTTAT
GGCCGCACCATAACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTGCCCCCAC
CAAGCATTCTTCTAGGGACGTTTCTGGTTGGGGCGGAGATTGGTCTCCCGTGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCC-----

Hap_26 1

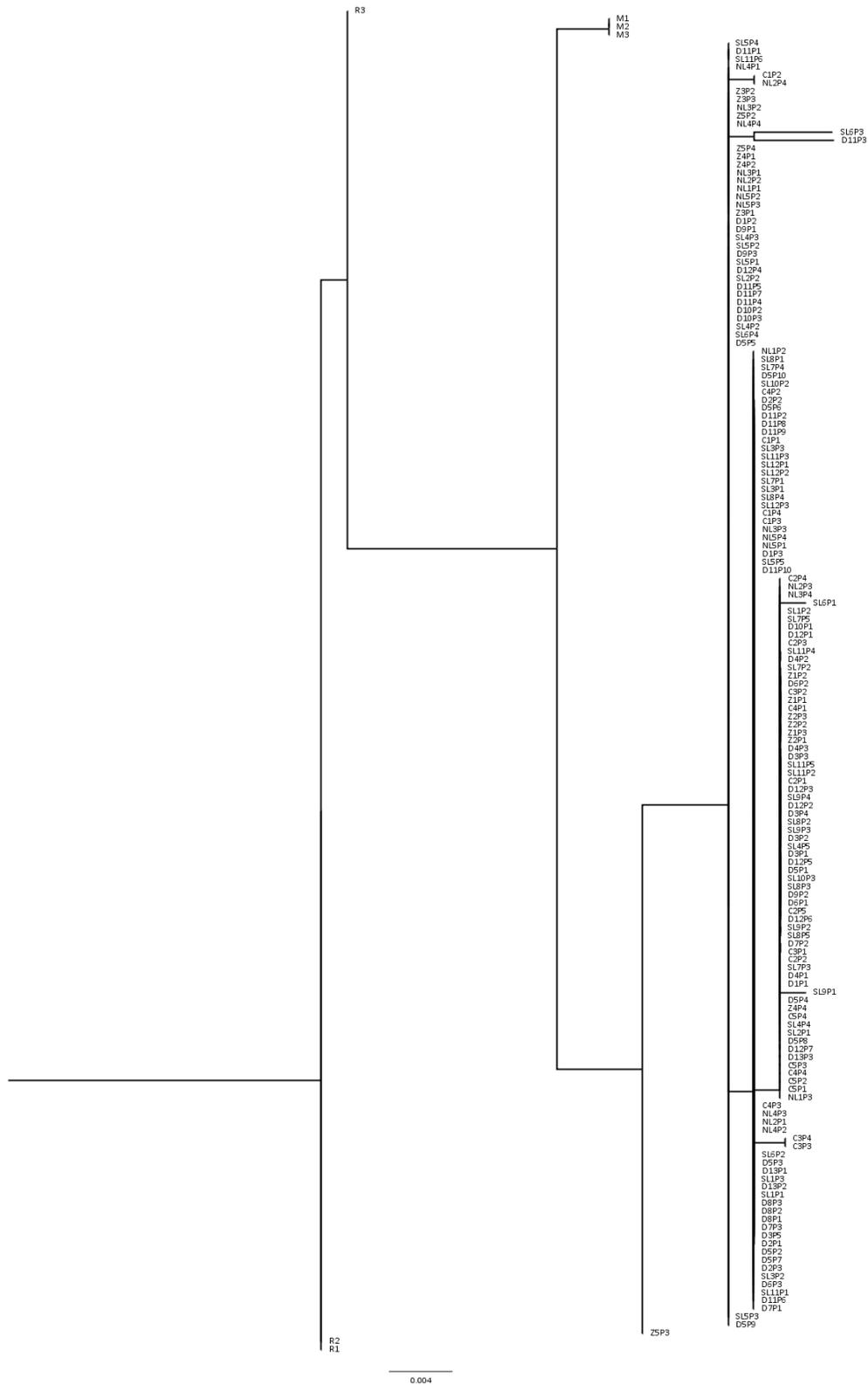
GAAGGATCATTGTGCGAACCCCTGCATAGCAGAACGACCCGTGAACAAGTAAAA
ACACCTGACCGA-
GCGAGGACCGACGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTGCGACGTGTGTTTAT
GGCTGCACCATAACATATGGGGCATCATGGATACATGTTGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAACCTAACTTAAAGGGCCCGTGCTATTGCGCCC

CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCTGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTCGCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCTCCCGTGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACGTCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCC-----

Hap_27 1

GAAGGATCATTGTCGAACCCTGCATAGCAGAACGACCCGCGAACAAGTAAA
AACACCTGACCGA-
GCGAGGATCGAAGCTTATGTTTTCGAGCCTCGTGAGGCCTTGTGACGTGTGTTTAT
GGCTGCACCATACATATGGGGCATCATGGATACATGTCGACACAATAACAAACCCC
CGGCACCGTACGTGCCAAGGAAAATAAACTTAAAGGGCCCGTGCTATTGCGCCC
CGTTCGCGGTGTGCGCATTGTACGTGGCAACTTTGCAAACCTTAAAACGACTCTCGG
CAACGGATATCTCGGCTCACGCATCGATGAAGAACGTAGCAAAATGCGATACTTGG
TGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTTGAACGCAAGTTGCGCCCGAA
GCCATCCGGTTGAGGGCACGTCTGCCTGGGCGTCACGCATCACGTCGCCCCAC
CAAGCATTCTTCTAGGGACGTCTCTGGTTGGGGCGGAGATTGGTCGCCCGTGCC
CAAGGTGTGGTTGGCCTAAATAGGAGTCTCCCCAAAAGGGACGCACTGCTAGTGG
TGGTTGATAACACAGTCGTCTCATGTCGTGCGTTTTTCATTCTTGGGTGTGGATGCTC
TTAAAATACCCC-----

Anexo 4. Árbol filogenético inferido con el método ML



VII. BIBLIOGRAFÍA

Abraham, EM; Aftzalanidou, A; Ganopoulos, I; Osathanunkul, M; Xanthopoulou, A; Avramidou, E; Sarrou, E; Aravanopoulos, F; Madesis, P. 2018. Genetic diversity of *Thymus sibthorpii* Bentham in mountainous natural grasslands of Northern Greece as related to local factors and plant community structure. *Industrial Crops and Products* 111:651-659.

Agarwal, S; Khan, S. 2019. Genetic diversity of 18 male and 18 female accessions of *Jajoba* [*Simmondsia chinensis* (link) Schneider] using EST-SSRs. *Meta Gene* 19:134-141.

Allen, JM; Germain-Aubrey, CC; Barve, N; Neubig, KM; Majure, LC; Laffan, SW; Mishler, BD; Owens, HL; Smith, SA; Whitten, WM; Abbott, JR; Soltis, DE; Guralnick, R; Soltis, PS. 2019. Spatial Phylogenetics of Florida Vascular Plants: The Effects of Calibration and Uncertainty on Diversity Estimates. *iScience* 11:57-70.

Álvarez-Buylla, ER; Garay-Arroyo, A; García-Ponce de León, B; De la paz Sanchez, M; González-Ortega, E; Dávila-Velderrain, J; Martínez-García, JC; Piñeyron-Nelson, A. 2017. *La Ecología Evolutiva del Desarrollo en México*. 88:14-26.

Amar, MH; Hassan, AHM; Sherbeny, Eam El; . 2012. Assessment of genetic diversity in some wild plants of asteraceae family by ribosomal dna sequence. *Egyptian Journal of Genetics And Cytology* 41:195-208.

Bandelt, HJ; Forster, P; Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16(1):37-48.

Barrera, CS; Cornish, K. 2016. High performance waste-derived filler/carbon black reinforced guayule natural rubber composites. *Industrial Crops and Products* 86:132-142.

Bekaardt, CR; Coffelt, TA; Fenwick, JR; Wiesner, LE. 2010. Environmental , irrigation and fertilization impacts on the seed quality of guayule (*Parthenium argentatum* Gray). *Industrial Crops & Products* 31(3):427-436.

Boateng, AA; Mullen, CA; Elkasabi, Y; McMahan, CM. 2015. Guayule (*Parthenium argentatum*) pyrolysis biorefining: Production of hydrocarbon compatible bio-oils from guayule bagasse via tail-gas reactive pyrolysis. *Fuel* 158:948-956.

Carvalho, YGS; Vitorino, LC; Souza, UJB de; Bessa, LA; Carvalho, YGS; Vitorino, LC; Souza, UJB de; Bessa, LA. 2019. Recent Trends in Research on the Genetic Diversity of

Plants: Implications for Conservation. *Diversity* 11(4):62.

Chen, S; Yao, H; Han, J; Liu, C; Song, J; Shi, L; Zhu, Y; Ma, X; Gao, T; Pang, X; Luo, K; Li, Y; Li, X; Jia, X; Lin, Y; Leon, C. 2010. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. *PLoS ONE* 5(1):e8613.

Cheng, T; Xu, C; Lei, L; Li, C; Zhang, Y; Zhou, S. 2016. Barcoding the kingdom Plantae: New PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources* 16(1):138-149.

Chowdhury, B; Garai, G. 2017. A review on multiple sequence alignment from the perspective of genetic algorithm. *Genomics* 109(5-6):419-431.

Chundawat, SPS; Chang, L; Gunawan, C; Balan, V; McMahan, C; Dale, BE. 2012. Guayule as a feedstock for lignocellulosic biorefineries using ammonia fiber expansion (AFEX) pretreatment. *Industrial Crops and Products* 37(1):486-492.

Coffelt, T; Johnson, L. 2011. A set of descriptors for evaluating guayule germplasm. *Industrial Crops and Products* 34(2):1252-1255.

Cruz, VM V.; Lynch, AL; Ray, DT; Niaura, WS; Purdy, PH; Dierig, DA. 2017. Analysis of mode of reproduction of guayule (*Parthenium argentatum* A. Gray) using flow cytometry and identification of polyhaploids for breeding. *Industrial Crops and Products* 107:618-623.

De-la-Mora, M; Piñero, D; Oyama, K; Farrell, B; Magallón, S; Núñez-Farfán, J. 2018. Evolution of *Trichobaris* (Curculionidae) in relation to host plants: Geometric morphometrics, phylogeny and phylogeography. *Molecular Phylogenetics and Evolution* 124:37-49.

Delgado, FM; Gómez-Vela, F. 2019. Computational methods for Gene Regulatory Networks reconstruction and analysis: A review. *Artificial Intelligence in Medicine* 95:133-145.

Drummond, AJ; Rambaut, A. 2007. BEAST : Bayesian evolutionary analysis by sampling trees. 8:1-8.

Eguiarte, LE; Aguirre-Liguori, JA; Jardón-Barbolla, L; Aguirre-Planter, E; 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* 16(1):42-56.

Emami, A; Shabaniyan, N; Rahmani, M-S; Khadivi, A; Mohammad-Panah, N. 2018. Genetic characterization of the *Crataegus* genus: Implications for in situ conservation. *Scientia Horticulturae* 231:56-65.

Excoffier, L; Smouse, PE; Quattro, JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131(2):479-491. Disponible en <https://pubmed.ncbi.nlm.nih.gov/1644282>.

Fehlberg, SD; Ranker, TA. 2009. Evolutionary history and phylogeography of *Encelia farinosa* (Asteraceae) from the Sonoran, Mojave, and Peninsular Deserts. *Molecular Phylogenetics and Evolution* 50(2):326-335.

Foster, MA; Coffelt, TA; Petty, AK. 2011. Guayule production on the southern high plains. *Industrial Crops and Products* 34(3):1418-1422.

Ghada, B; Ahmed, BA; Messaoud, M; Amel, S-H. 2013. Genetic diversity and molecular evolution of the internal transcribed spacer (ITSs) of nuclear ribosomal DNA in the Tunisian fig cultivars (*Ficus carica* L.; Moracea). *Biochemical Systematics and Ecology* 48:20-33.

Godini, R; Fallahi, H. 2019. A brief overview of the concepts, methods and computational tools used in phylogenetic tree construction and gene prediction. *Meta Gene* 21:100586.

Golkar, P; Nourbakhsh, V. 2019. Analysis of genetic diversity and population structure in *Nigella sativa* L. using agronomic traits and molecular markers (SRAP and SCoT). *Industrial Crops and Products* 130:170-178.

Gore, MA; Coyle, G; Friebe, B; Coffelt, TA; Salvucci, ME. 2011. Complex ploidy level variation in guayule breeding programs. *Crop Science* 51(1):210-216.

Guevara, X. 2015. The industrialization of Salins Peñon Blanco case lagoon Santa María. *Gremium* 2:34-49. Consultado 16 ene. 2019. Disponible en <https://www.printfriendly.com/p/g/ZvCrF6>.

Hall, TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp* 41:95-98.

Hao, H-C; Chang, J-Y; Chung, F-Y. 2012. Internal transcribed spacer sequence analysis of *Angelica* from different habitats. *Genomic Medicine, Biomarkers, and Health Sciences* 4(1-2):43-44.

Huelsenbeck, JP; Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees . *Bioinformatics* 17(8):754-755.

Hughes, AR; Inouye, BD; Johnson, MTJ; Underwood, N; Vellend, M. 2008. Ecological consequences of genetic diversity. *Ecology Letters* 11(6):609-623.

Ilut, DC; Sanchez, PL; Coffelt, TA; Dyer, JM; Jenks, MA; Gore, MA. 2017. A century of guayule: Comprehensive genetic characterization of the US national guayule (*Parthenium argentatum* A. Gray) germplasm collection. *Industrial Crops and Products* 109:300-309.

Ilut, DC; Sanchez, PL; Costich, DE; Friebe, B; Coffelt, TA; Dyer, JM; Jenks, MA; Gore, MA. 2015. Genomic diversity and phylogenetic relationships in the genus *Parthenium* (Asteraceae). *Industrial Crops and Products* 76:920-929.

IRSG. 2015. No Title (en línea, sitio web). Disponible en <http://www.rubberstudy.com/default.aspx>.

Jagadeesh, D; Prasanna Kumar, MK; Chandrakanth, R; Devaki, NS. 2018. Molecular diversity of internal transcribed spacer among the monoconidial isolates of *Magnaporthe oryzae* isolated from rice in Southern Karnataka, India. *Journal of Genetic Engineering and Biotechnology* 16(2):631-638.

Jayasena, AS; Fisher, MF; Panero, JL; Secco, D; Bernath-Levin, K; Berkowitz, O; Taylor, NL; Schilling, EE; Whelan, J; Mylne, JS. 2017. Stepwise Evolution of a Buried Inhibitor Peptide over 45 My. *Molecular Biology and Evolution* 34(6):1505-1516.

Jia, Y; Zhu, J; Wu, Y; Fan, W-B; Zhao, G-F; Li, Z-H. 2018. Effects of Geological and Environmental Events on the Diversity and Genetic Divergence of Four Closely Related Pines: *Pinus koraiensis*, *P. armandii*, *P. griffithii*, and *P. pumila*. *Frontiers in Plant Science* 9:1264.

Jukes, TH; Cantor, CR. 1969. Evolution of Protein Molecules. *Mammalian Protein Metabolism: Volume III* :21-132.

Kaushik, AC; Pal, A; Kumar, A; Dwivedi, VD; Bharadwaj, S; Pandey, A; Mishra, SK; Sahi, S. 2017. Internal transcribed spacer sequence database of plant fungal pathogens: PFP-ITSS database. *Informatics in Medicine Unlocked* 7:34-38.

Kehie, M; Kumaria, S; Devi, KS; Tandon, P. 2016. Genetic diversity and molecular evolution of Naga King Chili inferred from internal transcribed spacer sequence of nuclear ribosomal DNA. *Meta Gene* 7:56-63.

Kolář, F; Čertner, M; Suda, J; Schönswetter, P; Husband, BC. 2017. Mixed-Ploidy Species: Progress and Opportunities in Polyploid Research. s.l., Elsevier Ltd, vol.22. p. 1041-1055.

Kosanic, A; Anderson, K; Harrison, S; Turkington, T; Bennie, J. 2018. Changes in the geographical distribution of plant species and climatic variables on the West Cornwall peninsula (South West UK). PLOS ONE 13(2):e0191021.

Kumar, J; Heikrujam, M; Sharma, K; Agrawal, V. 2019. SRAP and SSR marker-assisted genetic diversity, population structure analysis and sex identification in Jojoba (*Simmondsia chinensis*). Industrial Crops and Products 133:118-132.

Kumar, S; Hahn, FM; McMahan, CM; Cornish, K; Whalen, MC. 2009. Comparative analysis of the complete sequence of the plastid genome of *Parthenium argentatum* and identification of DNA barcodes to differentiate *Parthenium* species and lines. BMC Plant Biology .

Kumar, S; Hahn, FM; McMahan, CM; Cornish, K; Whalen, MC. 2009. Comparative analysis of the complete sequence of the plastid genome of *Parthenium argentatum* and identification of DNA barcodes to differentiate *Parthenium* species and lines. BMC Plant Biology 9:1-12.

Kumari, S; Kanth, BK; Jeon, Y; Jang, J-Y; Kim, H-S; Lee, G-J. 2018. Internal transcribed spacer-based CAPS marker development for *Lilium hansonii* identification from wild *Lilium* native to Korea. Scientia Horticulturae 236:52-59.

Leduc, D; Zhao, ZQ. 2019. Phylogenetic position of the parasitic nematode *Trophomera* (Nematoda, Benthimermithidae): A molecular analysis. Molecular Phylogenetics and Evolution 132:177-182.

Lee, S-C; Wang, C-H; Yen, C-E; Chang, C. 2017. DNA barcode and identification of the varieties and provenances of Taiwan's domestic and imported made teas using ribosomal internal transcribed spacer 2 sequences. Journal of Food and Drug Analysis 25(2):260-274.

_____. 2017. DNA barcode and identification of the varieties and provenances of Taiwan's domestic and imported made teas using ribosomal internal transcribed spacer 2 sequences. Journal of Food and Drug Analysis 25(2):260-274.

Li, X; Zhang, S; Yang, Z; Song, K; Yi, T. 2013. Conservation genetics and population

diversity of *Erigeron breviscapus* (Asteraceae), an important Chinese herb. *Biochemical Systematics and Ecology* 49:156-166.

Liò, P; Goldman, N. 1998. Models of molecular evolution and phylogeny. *Genome research* 8(12):1233-44.

Long, C; Jiang, Z; Dao, Z. 2010. Genetic diversity of *Ottelia acuminata* (Hydrocharitaceae) from the Eastern Himalayas, revealed by ISSR markers. *Botanica Orientalis-Journal of Plant Science* 7:56-63.

Losos, JB. 2014. *The Princeton guide to evolution*. 1 ed. Baum, DA; Futuyma, DJ; Hoekstra, HE; Lenski, RE; Moore, AJ; Peichel, CL; Schluter, D; Whitlock, MJ (eds.). New Jersey, Princeton University Press, vol.51. 877 p.

Maatooq, GT; El Gamal, AAH; Furbacher, TR; Cornuelle, TL; Hoffmann, JJ. 2002. Triterpenoids from *Parthenium argentatum* x *P. tomentosum*. *Phytochemistry* 60(8):755-760.

Maina, N; Baraket, G; Salhi-Hannachi, A; Sakka, H. 2019. Sequence analysis and molecular evolution of Tunisian date palm cultivars (*Phoenix dactylifera* L.) based on the internal transcribed spacers (ITSs) region of the nuclear ribosomal DNA. *Scientia Horticulturae* 247:373-379.

Mayonde, S; Cron, GV; Glennon, KL; Byrne, MJ. 2019. Genetic diversity assessment of *Tamarix* in South Africa – Biocontrol and conservation implications. *South African Journal of Botany* 121:54-62.

Monfared, MA; Samsampour, D; Sharifi-Sirchi, GR; Sadeghi, F. 2018. Assessment of genetic diversity in *Salvadora persica* L. based on inter simple sequence repeat (ISSR) genetic marker. *Journal of Genetic Engineering and Biotechnology* 16(2):661-667.

Panero, JL; Crozier, BS. 2016. Macroevolutionary dynamics in the early diversification of Asteraceae. *Molecular Phylogenetics And Evolution* 99:116-132.

Peña, C. 2011. Métodos de inferencia filogenética. *Revista peruana de biología* 18(182):265-267.

Peng, YQ; Fan, LL; Mao, FY; Zhao, YS; Xu, R; Yin, YJ; Chen, X; Wan, DG; Zhang, XH. 2018. Genetic diversity and population structure of a protected species: *Polygala tenuifolia* Willd. *Comptes Rendus Biologies* 341(3):152-159.

Plovanich, AE; Panero, JL. 2004. A phylogeny of the ITS and ETS for *Montanoa*

(Asteraceae: Heliantheae). *Molecular Phylogenetics and Evolution* 31(3):815-821.

Posada, D; Crandall, KA. 1998. BIOINFORMATICS APPLICATIONS NOTE MODELTEST : testing the model of DNA substitution. *Molecular Biology and Evolution* 14(9):817-818.

Qian, J; Sun, Y; Duan, Y-H. 2009. Internal Transcribed Spacer Region of rDNA in Common Wheat and Its Genome Origins. *Acta Agronomica Sinica* 35(6):1021-1029.

QIAN, J; SUN, Y; DUAN, Y-H. 2009. Internal Transcribed Spacer Region of rDNA in Common Wheat and Its Genome Origins. *Acta Agronomica Sinica* 35(6):1021-1029.

Ramakrishnan, M; Ceasar, SA; Duraipandiyan, V; Al-Dhabi, NA; Ignacimuthu, S. 2016. Using molecular markers to assess the genetic diversity and population structure of finger millet (*Eleusine coracana* (L.) Gaertn.) from various geographical regions. *Genetic Resources and Crop Evolution* 63(2):361-376.

Rambaut, A. 2012. FigTree (en línea, sitio web). Consultado 20 feb. 2020. Disponible en <http://tree.bio.ed.ac.uk/software/figtree/>.

_____. 2012. Tracer (en línea, sitio web). Consultado 20 feb. 2020. Disponible en <http://tree.bio.ed.ac.uk/software/tracer/>.

Rasutis, D; Soratana, K; McMahan, C; Landis, AE. 2015. A sustainability review of domestic rubber from the guayule plant. *Industrial Crops and Products* 70:383-394.

Ray, DT; Coffelt, TA; Dierig, DA. 2005. Breeding guayule for commercial production. *Industrial Crops and Products* 22(1):15-25.

_____. 2005. Breeding guayule for commercial production. *Industrial Crops and Products* 22(1):15-25.

Rodríguez-correa, H; González-rodríguez, A; Oyama, K. 2017. Perspectivas de la Ecología Molecular en un país megadiverso. *Enfermería Universitaria* 88:3-13.

de Rodriguez, DJ; Angulo-Sánchez, JL; Rodríguez-García, R; Díaz-Solis, H. 2005. Recent data on guayule from natural stands in Mexico. *Industrial Crops and Products* 22(1):87-93.

Rozas, J. 2009. *Bioinformatics for DNA Sequence Analysis*. s.l., s.e., vol.537.

Sabaini, PS; Boateng, AA; Schaffer, M; Mullen, CA; Elkasabi, Y; McMahan, CM; Macken, N. 2018. Techno-economic analysis of guayule (*Parthenium argentatum*) pyrolysis biorefining: Production of biofuels from guayule bagasse via tail-gas reactive pyrolysis. *Industrial Crops and Products* 112:82-89.

_____. 2018. Techno-economic analysis of guayule (*Parthenium argentatum*) pyrolysis biorefining: Production of biofuels from guayule bagasse via tail-gas reactive pyrolysis. *Industrial Crops and Products* 112:82-89.

Sanchez, PL; Costich, DE; Friebe, B; Coffelt, TA; Jenks, MA; Gore, MA. 2014. Genome size variation in guayule and mariola: Fundamental descriptors for polyploid plant taxa. *Industrial Crops and Products* 54:1-5.

Scanu, B; Linaldeddu, BT; Deidda, A; Jung, T. 2015. Diversity of Phytophthora Species from Declining Mediterranean Maquis Vegetation, including Two New Species, *Phytophthora crassamura* and *P. ornamentata* sp. nov. *PLOS ONE* 10(12):e0143234.

Sims, LL; Chee, C; Bourret, T; Hunter, S; Garbelotto, M. 2019. Genetic and phenotypic variation of *Phytophthora crassamura* isolates from California nurseries and restoration sites. *Fungal Biology* 123(2):159-169.

Sivakumar, G; Swami, SK; Nagarajan, G; Mehta, SC; Tuteja, FC; Ashraf, M; Patil, NV. 2018. Molecular characterization of *Hyalomma dromedarii* from North Western Region of India based on the gene sequences encoding Calreticulin and Internally Transcribed Spacer Region 2. *Gene Reports* 10:141-148.

Stamatakis, A; Hoover, P; Rougemont, J. 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology* 57(5):758-771.

Stireman, JO; Cerretti, P; O'Hara, JE; Blaschke, JD; Moulton, JK. 2018. Molecular phylogeny and evolution of world Tachinidae (Diptera). *Molecular Phylogenetics and Evolution* .

Sudha, GS; Ramesh, P; Sekhar, AC; Krishna, TS; Bramhachari, PV; Riazunnisa, K. 2019. Genetic diversity analysis of selected Onion (*Allium cepa* L.) germplasm using specific RAPD and ISSR polymorphism markers. *Biocatalysis and Agricultural Biotechnology* 17:110-118.

Sullivan, J; Joyce, P. 2005. Model Selection in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 36(1):445-466.

Thompson, AE; Ray, DT. 1989. Breeding guayule. *Plant Breed. Rev* 6:93-165.

Veatch-Blohm, ME; Ray, DT; McCloskey, WB. 2006. Water-stress-induced changes in resin and rubber concentration and distribution in greenhouse-grown guayule. *Agronomy Journal* 98(3):766-773.

Veatch, ME; Ray, DT; Mau, CJD; Cornish, K. 2005. Growth, rubber, and resin evaluation of two-year-old transgenic guayule. *Industrial Crops and Products* 22(1):65-74.

Wang, Y; Ghouri, F; Shahid, MQ; Naeem, M; Baloch, FS. 2017. The genetic diversity and population structure of wild soybean evaluated by chloroplast and nuclear gene sequences. *Biochemical Systematics and Ecology* 71:170-178.

Wicke, K; Fischer, M. 2018. Phylogenetic diversity and biodiversity indices on phylogenetic networks. *Mathematical Biosciences* 298:80-90.

Wu, C-T; Hsieh, C-C; Lin, W-C; Tang, C-Y; Yang, C-H; Huang, Y-C; Ko, Y-J. 2013. Internal transcribed spacer sequence-based identification and phylogenetic relationship of I-Tiao-Gung originating from *Flemingia* and *Glycine* (Leguminosae) in Taiwan. *Journal of Food and Drug Analysis* 21(4):356-362.

Zamudio, GS; Prosdocimi, F; de Farias, ST; José, M V. 2019. A neutral evolution test derived from a theoretical amino acid substitution model. *Journal of Theoretical Biology* 467:31-38.

Zeinalabedini, M; Majidian, P; Ashori, R; Gholaminejad, A; Ebrahimi, MA; Martinez-Gomez, P. 2019. Integration of molecular and geographical data analysis of Iranian *Prunus scoparia* populations in order to assess genetic diversity and conservation planning. *Scientia Horticulturae* 247:49-57.

Zhai, SH; Yin, GS; Yang, XH. 2018. Population Genetics of the Endangered and Wild Edible Plant *Ottelia acuminata* in Southwestern China Using Novel SSR Markers. *Biochemical Genetics* 56(3):235-254.

Zhang, Q-P; Liu, D-C; Liu, S; Liu, N; Wei, X; Zhang, A-M; Liu, W-S. 2014. Genetic diversity and relationships of common apricot (*Prunus armeniaca* L.) in China based on simple sequence repeat (SSR) markers. *Genetic Resources and Crop Evolution* 61(2):357-368.

Zhu, Y; Chen, H; Fan, J; Wang, Y; Li, Y; Chen, J; Fan, J; Yang, S; Hu, L; Leung, H; Mew, TW; Teng, PS; Wang, Z; Mundt, CC. 2000. Genetic diversity and disease control in rice. *Nature* 406(6797):718-722.

Zuckerandl, E; Pauling, L. 1965. Evolutionary Divergence and Convergence in Proteins. *Evolving Genes and Proteins* :97-166.